Мнистарство просвете науке и технолошког развоја
Матични научни одбор за физику
Датум: 19.07.2022. године
Београд

Институт за физику у Београду
Др Марија Митровић Данкулов, виши научни сарадник

Прегревица 118
11080 Београд

Поштована др Митровић Данкулов,

Матични научни одбор за физику је на својој седници од 15. јула 2022. године разматрао Ваш захтев за признавање и категорисање следеће публикације:

Поглавље 1, Social Networks Theory, Marija Mitrović Dankulov, María del Mar Alonso-Almeida, Fariya Sharmeen, Agnieszka Lukasiewicz, објављено у монографији Digital Social Networks and Travel Behaviour in Urban Environments, DOI: 10.4324/9780429488719, Routledge (Taylor&Fransis Group), 2019.

Одбор је једногласно донео одлуку да се наведена публикација категорише као истакнута монографија међународног значаја, а Ваш допринос као монографска студија у оквиру ове монографије, која се на основу Прилога 1 Правилника о стицању истраживачких и научних звања категорише као публикација категорије **М13** - Монографска студија/поглавље у књизи М11 или рад у тематском зборнику водећег међународног значаја.

С поштовањем,

Председник Матичног научног одбора за физику

др Антун Балаж, научни саветник

Мнистарство просвете науке и технолошког развоја
Матични научни одбор за физику
Датум: 19.07.2022. године
Београд

Институт за физику у Београду
Др Марија Митровић Данкулов, виши научни сарадник

Прегревица 118
11080 Београд

Поштована др Митровић Данкулов,

Матични научни одбор за физику је на својој седници од 15. јула 2022. године разматрао Ваш захтев за признавање и категорисање следеће публикације:

Поглавље 3, Classifying Networks with dk-Series, Marija Mitrović Dankulov, Guido Caldarelli, Santo Fortunato and Dmitri Krioukov, објављено у монографији Multiplex and Multilayer Networks, DOI:10.1093/oso/9780198809456.001.0001, Oxford University Press, 2018.

Одбор је једногласно донео одлуку да се наведена публикација категорише као монографија међународног значаја, а Ваш допринос као монографска студија у оквиру ове монографије, која се на основу Прилога 1 Правилника о стицању истраживачких и научних звања категорише као публикација категорије **М14** - Монографска студија/поглавље у књизи М12 или рад у тематском зборнику међународног значаја.

С поштовањем,

Председник Матичног научног одбора за физику

др Антун Балаж, научни саветник

Мнистарство просвете науке и технолошког развоја
Матични научни одбор за физику
Датум: 19.07.2022. године
Београд

Институт за физику у Београду
Др Марија Митровић Данкулов, виши научни сарадник

Прегревица 118
11080 Београд

Поштована др Митровић Данкулов,

Матични научни одбор за физику је на својој седници од 15. јула 2022. године разматрао Ваш захтев за признавање и категорисање следеће публикације:

Поглавље 2, The Structure and Dyinamics of Meetup Sodial Networks, Marija Mitrović Dankulov and Jelena Smiljanić, објављено у монографији Scientific Computing: Studies and Application, Editor Caj Erling, ISBN: 978-1-53612-564-1, Nova Science Pub Inc, 2017.

Одбор је једногласно донео одлуку да се наведена публикација категорише као истакнута монографија међународног значаја, а Ваш допринос као монографска студија у оквиру ове монографије, која се на основу Прилога 1 Правилника о стицању истраживачких и научних звања категорише као публикација категорије **M13** - Монографска студија/поглавље у књизи M11 или рад у тематском зборнику водећег међународног значаја.

С поштовањем,

Председник Матичног научног одбора за физику

др Антун Балаж, научни саветник

COMPUTER SCIENCE, TECHNOLOGY AND APPLICATIONS

Additional books in this series can be found on Nova's website under the Series tab.

Additional e-books in this series can be found on Nova's website under the eBooks tab.

SCIENTIFIC COMPUTING

STUDIES AND APPLICATIONS

CAJ ERLING
EDITOR

nova
science publishers
New York

NOTICE TO THE READER

CONTENTS

Chapter 2

THE STRUCTURE AND DYNAMICS OF MEETUP SOCIAL NETWORKS

Marija Mitrović Dankulov* and Jelena Smiljanić
Scientific Computing Laboratory,
Center for the Study of Complex Systems,
Institute of Physics Belgrade,
University of Belgrade, Belgrade, Serbia

Abstract

Computational social science is an emerging interdisciplinary field whose main focus is on investigating human and social dynamics using computational modeling and data analytics. Collective behavior in social systems is of special interest due to its complex nature which demands the use of computational techniques for uncovering and understanding the mechanisms that underlie its emergence. Here we demonstrate how computational techniques can be used to quantify the structure and dynamics of Meetup social groups. Dynamics and structure of any Meetup social group strongly depend on the association of its members with the community. We map the data about the group members and their participation in the group events onto bipartite networks and use tools of complex network theory to analyze their structure and structure of their monopartite projections. We explore how the structure of social networks of these groups evolve with time and identify the key topological features that influence collective dynamics and success/survival of the Meetup social group.

*Corresponding Author Email: mitrovic@ipb.ac.rs.

1. Introduction

The abundance of social data is the main driving force behind the expansion of a new interdisciplinary fields commonly known as *computational social science* [1, 2]. Researchers from different fileds of science including physics, computer science, mathematics, economics, and sociology, make the best use of data, computational techniques and resources to provide us with quantitative description and understanding of complex social systems. Computational methods from statistical mechanics and complex network theory have proven to be a useful tool for quantitative description and uncovering the mechanisms that underlie the collective behavior in social systems [3, 4]. These techniques have been udsed to explain the emergence of various collective social phenomena, including segregation [5], cooperation [6], collective emotions [7–10], popularity [11, 12], etc.

One manifestation of collective social behavior is a social group, informal community of individuals that emerges through repeated interactions among them. Although evolution and organization of social groups have been subject of interest in sociology and related scientific disciplines for more than a century [13–16], the development of information communication technologies and new quantitative techniques have provided new insight into the structure and dynamics of various social groups. The availability of data enabled the study of human behavior and the emergence of social groups in online environment on a scale that was not possible just two decades ago [8, 17–22]. However, the dynamics and structure of event-driven social groups are still insufficiently explored, mainly due to the lack of data. These groups have a vital place in society since they are a significant part of human social life [23–25]. The dynamics of these groups is very localized in time and space. Furthermore, their members meet and build social interactions at group events, which makes them very different from other offline and online social groups. Recent research on several types of these groups [26, 27], has shown that universal features characterize their dynamics and that structure of social relations among their members strongly influences member's activity in the group and thus her association with the community.

In this chapter, we demonstrate how computational techniques from statistical physics and complex network theory can be used to explore in details the structure and dynamics of event-driven social groups. As a case study, we use four different Meetup groups, communities of individuals formed online that get engaged into joint offline activities. Meetup members are organized into thematic social groups, which differ in physical location, type of activity and size. Thus, the analysis of these four groups allows us to detect and describe their universal structural and dynamical features. In particular, we show how the data about members activity in a Meetup group can be mapped onto a bipartite graph, and how this network can be used to obtain the group's social relations. The topological structure of a bipartite network and its monopartite projection reveals important details about social dynamics in these groups. We show that four considered groups have very similar structural properties, which further highlights the fact that universal principles govern dynamics of these groups. These universal principles can be uncovered through quantitative analysis of time series related to individual and collective activity. Our analysis shows that correlations between the activity of individual members are the principal cause of specific structural and dynamical patterns, which are characteristic for Meetup, and presumably other event-driven, social groups.

In Section 2 we describe in detail the four considered Meetup social groups and the data used for empirical analysis. The detailed description of structural properties of bipartite and monopartite weighted social networks is given in Section 3, while the analysis of dynamics of these groups is given in Section 4.

2. Data

Meetup [28] *is an online social networking site that facilitates offline group meetings in various locations around the world.* It enables its members to find and join social groups by their interests and get involved into activities of these groups. Therefore, Meetup groups bring together people with common interests who are usually not acquainted outside of the group's social circle. This fact allows us to study the evolution of group's structure and dynamics from its beginnings, as well as their mutual influence, using computational techniques from complex network theory and statistical physics.

Depending on the interest of its members, a group can be assigned to one of 33 categories, which include careers, hobbies, socializing, health, politics, books, etc. Besides topic and location, these groups also differ in size, event
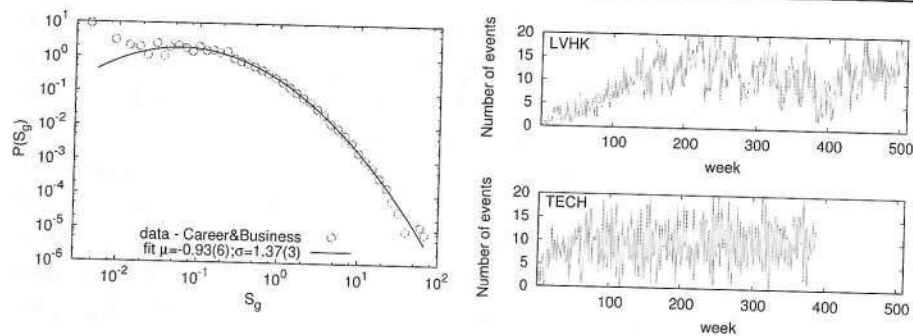
Figure 1. (left) Probability density distribution of group sizes S_g belonging to Meetup category *Career&Business*, which exhibits log-normal behaviour $f(S_g) = \frac{1}{S_g \sigma_g \sqrt{2\pi}} \exp(-\frac{(S_g - \mu_g)^2}{2\sigma_g^2})$. Here S_g is group size rescaled with average group size of category *Career&Business*. (right) Time series of number of events organised per week for two out of four groups we considered in this chapter.

dynamics, and organizational structure. Figure 1(left) shows the probability density distribution of sizes of groups belonging to category *Career&Business*. This distribution exhibits a log normal behavior, which suggests that some multiplicative mechanisms underlie the growth of Meetup social groups. Time series of the number of organized events per week for two different groups, shown in Fig. 1(right), clearly demonstrate the difference in event dynamics: the number of events organized on a weekly basis in two different groups fluctuates differently. Despite their differences, Meetup groups have one thing in common, the way their members prepare and organize their events. Before each event, all members of the group receive an invitation to participate in an event to which they can reply with yes or no. These recordings of event attendance can be then used to analyze members event participation patterns [27] and structure of social network formed between the members of the group.

In this work, we will demonstrate the power of computational techniques by analyzing the data from four different Meetup groups, where each one of them has more than 3000 members and organized events, see Table 1. To obtain more general conclusions, we have selected groups from different categories and with a different type of activities. Specifically, our first group *geamclt* (GEAM) brings together individuals, *foodie thrill-seekers*, who are mostly interested in

trying new exotic foods and drinks. Thus, the activities of this group include collective visits to restaurants and bars. The second group (LVHK) consists of hikers who seek excitement and like to socialize through physical activities, while members of our third (PGHF) group like to attend free, or almost free, social events. Our fourth group (TECH) gathers technology professionals interested in networking, entrepreneurship, environmental sustainability, and professional development. We used the Meetup public API to collect the data. For all four groups we compiled the list of all events organized by the group since its beginning, and for each event, we collected the list of members who attended that event. Every member of the Meetup community has a unique identifiers which allows us to follow her activity through time. The detailed description of the data is given in Table 1. To protect the privacy of Meetup users we have conducted a full anonymisation of the data. The anonymised data used for the analysis presented in this work can be found at [29].

Table 1. Description of four Meetup groups studied in this chapter. N_M is total number of group members, while N_E is total number of organized events

Meetup group	Acronym	Category	N_M	N_E
geamclt	GEAM	Food & Drink	5377	3986
pittsburgh-free	PGHF	Socializing	4995	4617
techlifecolumbus	TECH	Tech	3217	3162
VegasHikers	LVHK	Outdoors & Adventure	6061	5096

3.　Structure of Meetup Social Groups

The data from event-driven social groups are typically mapped onto monopartite weighted graph [30–32]. In this framework, each member is represented by a network node, while the link weight between two individuals is either raw [33] or normalized [30, 31] number of common events that they have attended together. The problem with this approach is that repeated co-occurrence of two group members at group's events does not necessarily imply social relationship between them. For instance, according to this approach two very active individuals, where by very active we mean persons who attended a significant number

of group's events, will have a high probability to meet on one or even several different events. In the monopartite network model, this manifests as a social relationship between these two individuals, although their relationship can be just a consequence of chance.
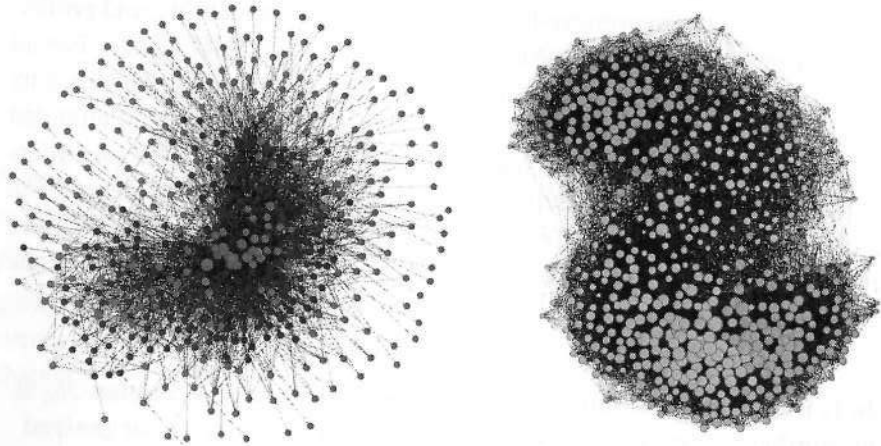

Figure 2. (left) A part of Meetup bipartite network of GEAM community: light gray circles represent nodes from Members partition while nodes from Events partition are represented with white circles. (right) Weighted network of significant social relations between the 10% most active members of LVHK group. Nodes of the same shade of gray belong to the same community.

In our previous work [27], we have shown that relevant social relations in event-driven social group can be obtained by mapping the data onto bipartite network and then use this network to filter out redundant links. First we map the data onto a network with two partitions *Members* (M) and *Events* (E). In bipartite networks, by definition, only links between nodes belonging to two different partitions are allowed. Here, the link between nodes i^M and j^E denotes presence of member i^M at event j^E. This way we obtain binary bipartite networks of size $N_M + N_E$, which can be represented as a block diagonal matrix B, where $B_{i^M j^E} = 1$ if there is a link between member i^M and event j^E, otherwise $B_{i^M j^E} = 0$. Example of bipartite networks of GEAM Meetup group is shown in Fig. 2(left). Bipartite networks created in this way contain all the necessary information about event-driven activity in Meetup groups. For instance, degree

of a node q_i^M in Members partition, defined as

$$q_{i^M}^M = \sum_{j^E=1}^{N_E} B_{i^M j^E} , \tag{1}$$

is equal to the number of events that member represented by that node has attended, while the degree of a node belonging to Events partition is equal to the size of corresponding event,

$$q_{j^E}^E = \sum_{i^M=1}^{N_M} B_{i^M j^E} . \tag{2}$$

In order to filter out redundant connections we use technique based on the configuration model of bipartite random graphs (CMBR) [34, 35]. In this filtering technique, CMBR is used for estimation of probability that two members have occurred together on w events by chance, i.e. link π-value. If this chance is low, below some threshold θ, we can claim at confidence level $1 - \theta$ that two members have correlated participation in group events, i.e. social relation between them is significant (real) and its strength is w. Before calculating this probability we first need to determine the probability $p_{i^M j^E}$ that member i^M has attended some event j^E by chance, i.e. a probability that there is a link between member i^M and event j^E in bipartite random network. For this we use graphs generated with CMBR and maximal-entropy approach for calculating the link probability in ensemble of networks [36–38]. CMBR graph has the same degree distribution for both partitions as the real network but all other properties, for instance degree-degree correlations, are randomized. Specifically, if we use G to denote the ensemble of bipartite random graphs with given degree distributions, then probability to have a specific realization of graph G from this ensemble is given by

$$P(G) = \frac{1}{Z} e^{-\sum_{i^M j^E} \alpha_{i^M} q_{i^M}^M + \beta_{j^E} q_{j^E}^E} , \tag{3}$$

where α_{i^M} and β_{j^E} are Lagrangian multipliers and Z is partition function. The partition function for the ensemble of CMBR graphs can be easily calculated (see [27, 35]) and it is equal to

$$Z = \sum_G e^{-\sum_{i^M j^E} (\alpha_{i^M} + \beta_{j^E}) B_{i^M j^E}} = \prod_{i^M j^M} (1 + e^{-(\alpha_{i^M} + \beta_{j^E})}) . \tag{4}$$

The values of Lagrangian multipliers can be calculated using the following equations

$$q_{i^M}^M = -\frac{\partial Z}{\partial \alpha_{i^M}} = \sum_{j^E=1}^{N_E} \frac{e^{-\alpha_{i^M}-\beta_{j^E}}}{1+e^{-\alpha_{i^M}-\beta_{j^E}}}, \tag{5}$$

and

$$q_{j^E}^E = -\frac{\partial Z}{\partial \beta_{j^E}} = \sum_{i^M=1}^{N_M} \frac{e^{-\alpha_{i^M}-\beta_{j^E}}}{1+e^{-\alpha_{i^M}-\beta_{j^E}}}. \tag{6}$$

If we define a coupling parameter $\lambda_{i^M j^E} = \alpha_{i^M} + \beta_{j^E}$ then the link probability between member i^M and event j^E is equal to

$$p_{i^M j^E} = \frac{\partial Z}{\partial \lambda_{i^M j^E}} = \frac{e^{\lambda_{i^M j^E}}}{1+e^{-\lambda_{i^M j^E}}} = \frac{e^{-\alpha_{i^M}+\beta_{j^E}}}{1+e^{-\alpha_{i^M}+\beta_{j^E}}}. \tag{7}$$

The links in CMBR graph are uncorrelated, thus the probability that two members i_1^M and i_2^M have attended the same event j^E is simply a product of $p_{i_1^M j^E}$ and $p_{i_2^M j^E}$. The probability that two members have co-occurred in w events is given by Poisson binomial distribution

$$P_{i_1^M i_2^M}(w) = \sum_{E_w} \prod_{j^E \in E_w} (p_{i_1^M j^E} p_{i_2^M j^E}) \prod_{j^E \in E_w} (1 - p_{i_1^M j^E} p_{i_2^M j^E}), \tag{8}$$

where E_w is the subset of w events that can be chosen from given E events [36]. Finally, π-value for a link between members i_1^M and i_2^M is a probability that these two members have attended together at least $w_{i_1^M i_1^M}$ events

$$\pi(w_{i_1^M i_2^M}) = \sum_{w \geq w_{i_1^M i_2^M}} P_{i_1^M i_2^M}(w). \tag{9}$$

If this π-value for a pair of members is high we can assume that their activity in group events is not correlated and that $w_{i_1^M i_2^M}$ co-occurrences are a result of chance. On the other hand, if the π-value for some pair of members is low, i.e. smaller than θ, we can assume at the level of confidence $1 - \theta$ that activity of these members is correlated to a certain extent. This means that the link of weight w between these members can not be created using CBMR graphs. The π-values of weighted links between members of one Meetup group can be then used to filter out the redundant links and create weighted network of significant social relations.
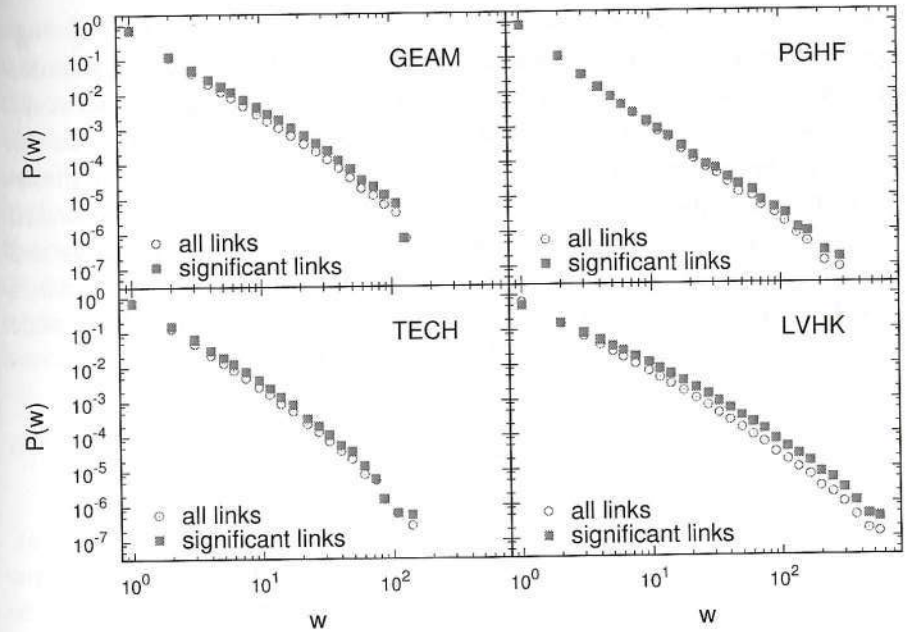
Figure 3. Probability distribution of link weights in a weighted network before and after filtering, for four considered Meetup groups. [Figure right reprinted with permissions from [27]]

We map the data from one Meetup group onto bipartite networks and then apply the described filtering technique to obtain the weighted network of significant social relations between its members at the 95% confidence level ($\theta = 0.05$). This weighted network can be represented with weighted matrix W, where $W_{i_1^M i_2^M}$ denotes the strength of the social relation between members i_1^M and i_2^M. The probability distribution of the number of common events, given in Fig. 3, has power-law behavior for all four networks. Fat-tail distribution of commons is typically found in techno-social systems characterized by high heterogeneity of members activity [7, 17, 27, 33]. The weighted network obtained from a projection of bipartite network onto one of the partitions is usually very dense, i.e. has a huge number of connections per node. We decrease the network density using the described filtering technique. We see from Fig. 3 that although this method puts more emphasis on the links with higher weight, the overall distribution of link weights is preserved.

In our previous work [27], we have examined in detail how the topological properties of member's ego network evolve with the number of attended events. So far, the structure of bipartite and weighted monopartite networks corresponding to Meetup groups has not been studied and described in detail. We quantify the structure of these networks by calculating several local, mesoscopic and global topological measures. The standard topological measure, used for quantifying the connectivity patterns in the networks, is the node degree and its distribution [39]. The degree of a node in Members and Events partitions is defined with Eqs. 1 and 2, while in the weighted monopartite network the node degree is equal to the number of its different links, i.e.

$$q_{i_1^M}^W = \sum_{i_2^M=1}^{N_M} \mathcal{H}(W_{i_1^M i_2^M}) , \qquad (10)$$

where $\mathcal{H}(W_{i_1^M i_2^M})$ is Heaviside function ($\mathcal{H}(x) = 1$ if $x > 0$ and 0 otherwise). The degree distribution $P(q)$ is a probability that randomly chosen node in a network, bipartite or monopartite, has a value of degree q. The connectivity in bipartite networks is quantified with two-degree distributions, each corresponding to one partition. Besides the degree and its distribution, one more local property is used for the description of local connectivity in a weighted network, node strength. Strength of a node is equal to a sum of weights of all links adjacent to that node, i.e.

$$s_{i_1^M}^W = \sum_{i_2^M=1}^{N_M} W_{i_1^M i_2^M} . \qquad (11)$$

Degree-degree correlations is another local topological measure with which we capture the mixing patterns in a network [39]. Specifically, with this measure, we quantify node's preference to attach to similar nodes in a network. It can be estimated by calculating the average nearest neighbor degree of a node $\langle q_{nn}^\kappa \rangle$ and examining its dependence on node degree. The average nearest neighbor degree of node with degree q^κ is given by

$$\langle q_n^\kappa n \rangle = \frac{\sum_{j^{\kappa'}} q_{j^{\kappa'}}^{\kappa'}}{q^\kappa} . \qquad (12)$$

Here κ and κ' denote two different partitions in bipartite networks, in the case of weighted network $\kappa = \kappa' = W$, and $q_{j^{\kappa'}}^{\kappa'}$ is degree of the nearest neighbor node. In uncorrelated networks, the average nearest neighbor degree does not depend on node's degree. If $\langle q_{nn}^\kappa \rangle$ is increasing function of q we have assortative mixing, i.e. nodes in this network have preference toward nodes with similar or higher degree, while decreasing function of $\langle q_{nn}^\kappa \rangle(q)$ indicates disassortative mixing, large degree nodes are linked to ones with small degree. Social networks are usually characterized by assortative mixing patterns, while biological, technological and techno-social networks typically have disassortative mixing [39]. For weighted networks, one can also define weighted average nearest neighbor degree $\langle q_{nn} \rangle^W$, which measures the effective affinity of a node to connect with high- or low-degree neighbors concerning the connection strength. Weighted average nearest neighbor degree of node i^M is given by

$$\langle q_{nn} \rangle_{i_1^M}^W = \frac{1}{s_{i_1^M}^W} \sum_{i_2^M} W_{i_1^M i_2^M} q_{i_2^M}^M . \qquad (13)$$

Another important measure of topological structure of monopartite networks is clustering coefficient. It measures a probability that two first neighbors of a node are also neighbors. For weighted monopartite social networks we can define non-weighted clustering coefficient of node i^M

$$c_{i_1^M} = \frac{1}{q_{i_1^M}^W (1-q_{i_1^M}^W)} \sum_{i_2^M i_3^M} \mathcal{H}(W_{i_1^M i_2^M}) \mathcal{H}(W_{i_1^M i_3^M}) \mathcal{H}(W_{i_2^M i_3^M}) , \qquad (14)$$

and weighted clustering coefficient

$$c_{i_1^M}^W = \frac{1}{s_{i_1^M}^W (1-q_{i_1^M}^W)} \sum_{i_2^M i_3^M} \frac{W_{i_1^M i_2^M} + W_{i_1^M i_3^M}}{2} \mathcal{H}(W_{i_1^M i_2^M}) \mathcal{H}(W_{i_1^M i_3^M}) \mathcal{H}(W_{i_2^M i_3^M}) . \qquad (15)$$

While the non-weighted clustering coefficient measures the number of closed triples formed by node and its neighbours, the $c_{i^M}^W$ measures local cohesiveness of the node's neighbourhood by taking into account the intensity of interactions between local triplets [40]. Non-weighted $\langle C \rangle$ (weighted $\langle C^W \rangle$) clustering coefficient of the whole network is calculated as average value of non-weighted(weighted) clustering coefficients of all nodes in the network. One can also calculate the average clustering coefficient of nodes with degree q^W and thus obtain how c and c^W depend on q^W. This dependences reveal a lot about the hierarchical organisation in network [39,40], while the averaged values of clustering coefficients provide global information between weights and topology in the network [40]. For instance, the comparison between $\langle C \rangle$ and $\langle C^W \rangle$ can reveal much about the importance of clustering in network evolution [40]:
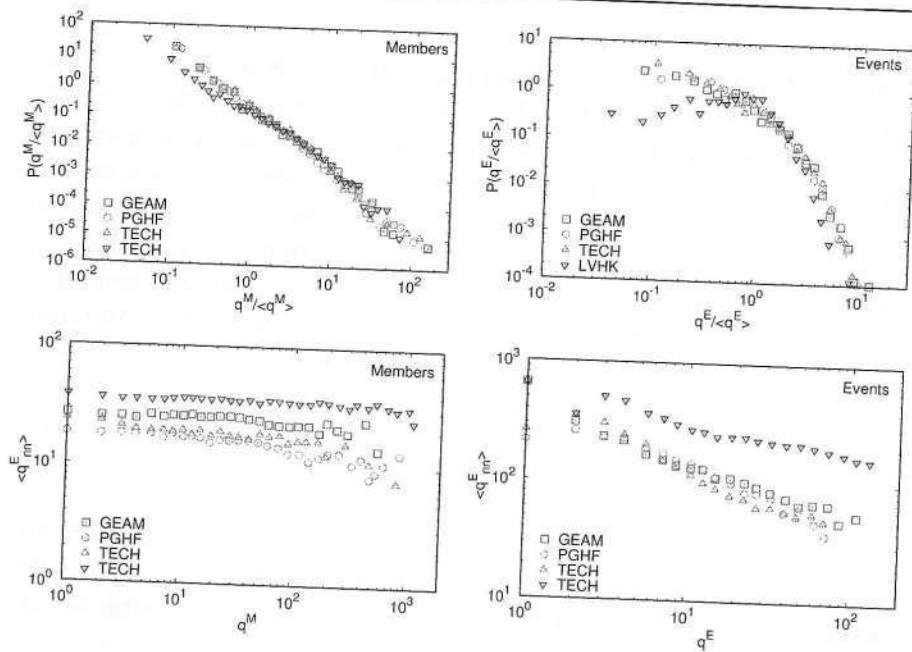
Figure 4. Probability density distribution of rescaled degree for nodes in Members (top left) and Events (top right) partition. The degree is rescaled by the value of an average degree in a given partition. The dependence of average nearest neighbor degree on degree in Members (bottom left) and Events (bottom right) partitions.

$\langle C^W \rangle > \langle C \rangle$ suggests that the processes leading to high value of clustering coefficient have an important role in network evolution.

We first examine the topological structure of bipartite networks corresponding to four considered Meetup groups by calculating the degree distribution for Members, Fig. 4(top left), and Events partition Fig. 4(top right). Degree distribution of Members partition is equivalent to the distribution of the number of attended events studied in [27]. Distributions for all four groups exhibit the same, truncated power-law, behavior

$$P(q^M) \propto (q^M)^{-\gamma} exp\left(-\frac{q^M}{q_0^M}\right), \tag{16}$$

with very similar values of exponent γ ($1.3 \leq \gamma \leq 1.61$) [27]. The observed power-law dependence in Members partition is a result of universal event-driven social dynamics [26,27]. In event-driven dynamics probability for a member to attend the next event depends exclusively on balance between the number of previous participations and non-participations. Our analysis of the evolution of members ego networks in Meetup social groups has shown that through event attendance members widen and strengthen their social connections and that active members repeat attendance in the group events with a small subgroup of their peers [27].

The probability distribution of degrees for Events partition is nothing more than a distribution of the event sizes. For all four social groups, this event size distribution follows log-normal behavior

$$P(q^E) = \frac{1}{q^E \sigma^E \sqrt{2\pi}} e^{-\frac{(\log(q^E)-\mu^E)^2}{2(\sigma^E)^2}}. \tag{17}$$

Three groups (GEAM, PGHF, and TECH) have the same value of mean $\mu^E = -0.7(1)$ and standard deviation $\sigma^E = 0.95(5)$, while the degree distribution for Events of LVHK group, with parameters $\sigma^E = 0.59(2)$ and $\mu^E = -0.16(2)$, shows some deviations from log-normal curve for small values of normalized degree. Log-normal distribution has been observed in many socioeconomic systems, including distribution of firm [41] and city [42] sizes, voting [11,43] and citation patterns [12,44]. For all these processes proportional growth is a common mechanism that leads to a log-normal distribution of event sizes. As it was argued in Section 2, the growth of Meetup groups can be described with multiplicative process [45]. Since event participation is the primary activity in a Meetup group, event size directly depends on the group's growth. Thus, the log-normal distribution of event sizes is just another indicator of the proportional growth of Meetup social groups.

We also examine the mixing patterns in these bipartite networks by calculating the average nearest neighbor degree for the nodes in both partitions and study its dependence on their degree, Fig. 4 (bottom). The average nearest neighbor degree for Members partition, i.e. the average size of events attended by a member, is independent of the magnitude of her activity Fig. 4(bottom right). The descending curve in Fig. 4(bottom left) indicates disassortative mixing for events. This observation is expected since in the previous study [27] we have shown that significant events are predominantly attended by members with a small number of participations.
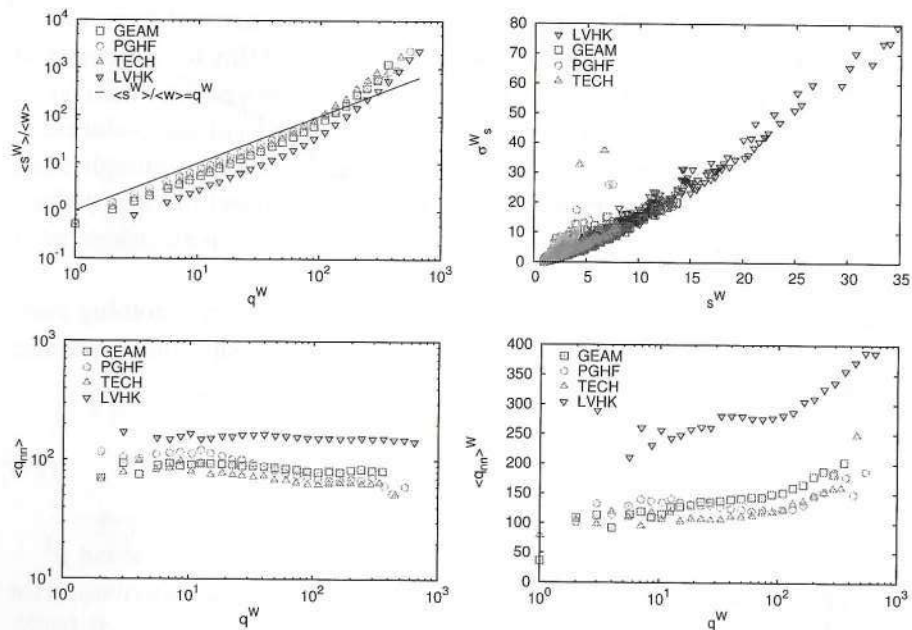
Figure 5. (top left) Dependence of normalized node strength on its degree. (top right) Scatter plot of nodes strength versus standard deviation of link weights adjacent to that node. Average nearest neighbor degree (bottom left) and its weighted counterpart (bottom right) in weighted network.

The observed degree-degree correlations suggest that event-driven dynamics is outlined with correlated social behavior which is beyond random processes characteristic for generation of random bipartite graphs. This event-driven dynamics of Meetup social group shapes and it is formed by the structure of the underlying weighted social network. Real world weighted networks, besides complex topology, also display a considerable heterogeneity in the strength of the connections. Weighted quantities and their correlations with the underlying topological structure of the network provide additional information about the dynamical processes governing network's evolution. We further explore topological features of the weighted network between the Meetup group members, which is obtained by projecting bipartite network onto Members partition and filtering out non-significant links.

Figure 5(left) shows dependence of normalized average strength $\langle s^W \rangle / \langle w \rangle$

of node on the value of its degree q^W. Normalization constant $\langle w \rangle$ is average link weight in the whole network. In the Meetup weighted social network the degree of a node equals to the number of different connections adjacent to that node, while its strength measures the total magnitude of these relationships. In uncorrelated weighted networks, the normalized average strength of the node grows with its degree as $\langle s^W \rangle / \langle w \rangle = q^W$. We see from 5(top left) that for all four groups the dependence of normalized average node's strength deviates from uncorrelated approximation. Furthermore, normalized average node's strength is proportional to its degree for nodes with degree $q^W \leq 50$, while for nodes with large values of degrees it grows super-linearly. This indicates that average weight of links adjacent to a node $\langle w_{iM} \rangle = \frac{1}{q_{i_1^M}^M} \sum_{i_2^M} W_{i_1^M i_2^M}$ grows with

node degree. In fact, the standard deviation of weights of links adjacent to a node also increases with its strength, Fig. 5 (top right), which suggests that very active users are characterized by significant heterogeneity of strength of their social connections. The described dependencies reveal one of the essential characteristics of event-driven dynamics: at the beginning of their involvement in group activities members mostly widen their social circle, while later, as their participation in group activities progresses, they tend to pay more attention to strengthening their existing social relations with particular members.

It was shown above that mixing patterns for Members partition in the bipartite network of Meetup groups are uncorrelated, so it is not surprising that we observe the lack of degree-degree correlations in corresponding weighted social networks Fig. 5 (bottom left). On the other hand, the dependence of weighted average nearest neighbor degree $\langle q_{nn} \rangle^W$ on node degree shows weak assortative mixing. The weighted average nearest neighbor degree is defined in a way to put more emphasis on the properties of strongest connections. Thus we see that members with a large number of connections establish the most solid ones with other high-degree members, forming in this way strongly connected core of very active members. This finding is further confirmed with a comparison of non-weighted $C(q^W)$ and weighted clustering $C^W(q^W)$ clustering coefficient shown in Fig. 6. For all four groups, non-weighted clustering coefficient and its weighted counterpart have the same value for low- and middle-degree nodes. For nodes with $q^W \simeq 50$, the difference between these two topological measures becomes evident, and it further grows with the degree. The ratio between non-weighted and weighted clustering coefficient of very active users suggests that their involvement in group activities is not conditioned by the behaviour of
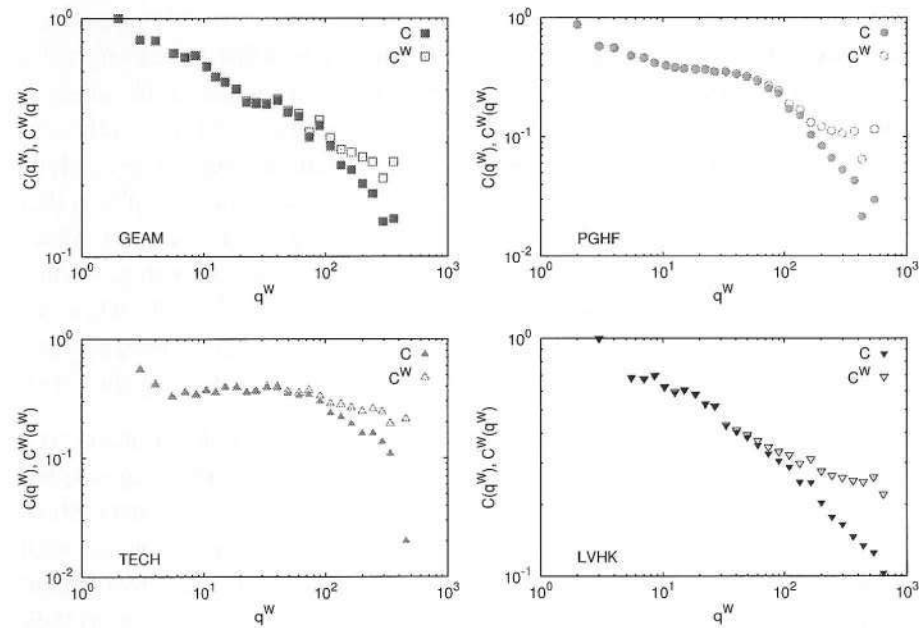
Figure 6. Dependence of non-weighted and weighted clustering coefficient on node degree for four studied Meetup social groups.

other members from their social circles but rather by the activity of the whole subgroups, as it was argued in [27].

This type of activity leads to the occurrence of communities which are manifested as mesoscopic inhomogeneities in the structure of weighted social networks, i.e. as the subgroups of strongly inter-connected nodes. Their existence, observed in many online and offline social systems [7,9,17,33], is one of the strongest indicators of self-organising collective dynamics. We detect these communities in the weighted social network of the 10% most active members using LOUVAIN community detection method [46,47] and its implementation in Gephi [48]. This approach detects communities in a network by finding the configuration of groups of nodes for which the modularity function has the highest value [46]. For all four groups we find that they have between four and five communities of similar sizes, see the Table 2 for details. For the illustration, we show the weighted social network of 10% most active individuals in LVHK and its community structure in Fig. 2(right).

Table 2. Details about weighted social subnetworks formed between 10% of most active members for considered networks. $\overline{N_W}$ and $\overline{L_W}$ are the number of members and links in subnetwork, activity range is the lowest and highest number of attended events by selected group of members, and N_C is the number of detected communities using LOUVAIN method

Acronym	$\overline{N_W}$	$\overline{L_W}$	Activity range	N_C
GEAM	538	15575	19-1494	4
PGHF	500	6828	15-1191	4
TECH	317	5350	21-1175	5
LVHK	607	29327	48-1521	4

All results presented in this section show that these four groups, although essentially different when it comes to their topic, type of activity and personality of their members, have very similar, practically universal, internal structure. This further confirms the conclusion from our previous research [26, 27] that the forces that influence the behavior of the individuals and thus govern the collective dynamics of the group are very fundamental, i.e. they do not depend on specific type of social group. In the next section we further support these findings by analysing group dynamics.

4. Dynamics of Meetup Social Groups

Specificity of Meetup social groups is their event-driven dynamics. Members of the group meet at a precise time and in a precisely determined place to participate in some joint activity. Through these activities, they form and strengthen social connections, which then influence the future dynamics of the group, i.e. the frequency and size of future events. The individual participations in group events can be translated into time series that either describes the activity of each member or the collective activity of the subgroup or the whole group. The quantitative analysis of these time series provides a characterization of system dynamics and reveals valuable information about its organization and evolution [7,8,10,17,33]. In this section, we analyze activity patterns of individual members as well as the total subgroup and group activity to find universal features of Meetup group dynamics.
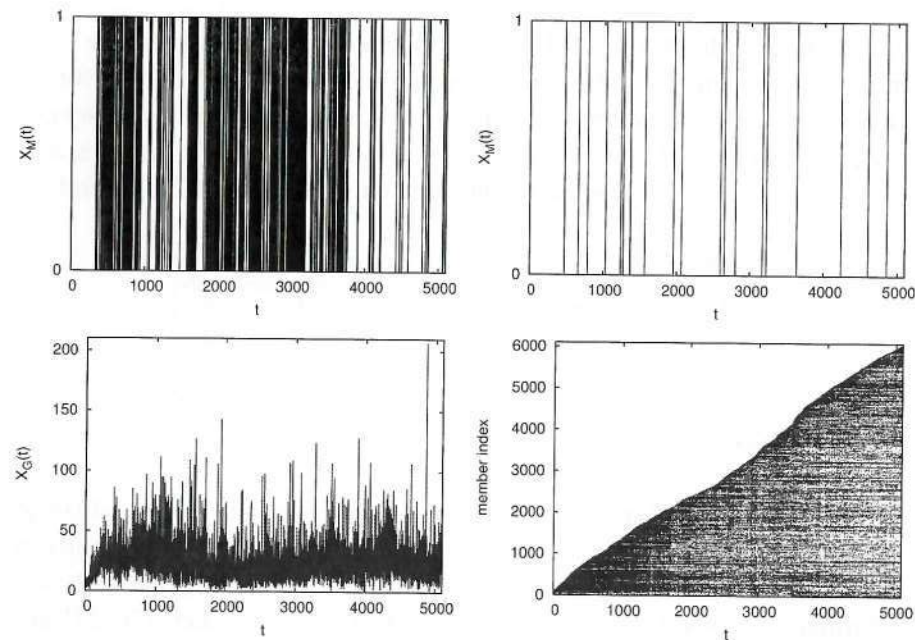
Figure 7. Time series of activity of one of the very active members (top left) and low active members (top right) of LVHK group. (bottom left) Time series of total activity in LVHK group. (bottom right) LVHK members activity pattern. The x axes in Fig.7 corresponds to event-time while y axes shows member's index. Members are ordered according to their first occurrence time in group activities, and dot denotes member's participation in event.

The number of organized events per week varies not just between different Meetup groups, but also within the group Fig. 1 (right). This frequency is influenced by various external and internal factors. To separate factors of different nature, we measure the activity in the group against the number of held events and thus omit real time from the further analysis. This way we obtain different time series $X(t)$ where t is a serial number of event. Time series of individual member's activity $X_{iM}(t)$, shown in Fig. 7 (top), is an array of 0 and 1 (a member can either attend or not attend the event). The time series that describes the activity of the whole group $X_G(t)$ is obtained through aggregation of the time series for individual members. $X_G(t)$ is nothing else but the time series of event sizes Fig. 7(bottom left).

First, we study the patterns of member's participation in group's events. Based on time series of individual member's activity one can calculate the time lag between two attended events ΔT^M, measured in event-time, and obtain the time series of inter-event times for each member $\{\Delta T_1^{iM}, \ldots, \Delta T_{q^M-1}^{iM}\}$. As it was shown in [27] the distributions of these time interval for all four Meetup groups follow the truncated power-law behavior with power-law exponents in interval $[1.06, 1.38]$. This suggests that temporal patterns of members activity have fractal structure [8, 33] known as bursty behavior [49]. Bursty behavior, characterized by intervals with persistent activity followed by the long inactive intervals, is universal characteristic of human behavior and it has been observed in various social systems [22, 33, 50]. This type of activity results in particular pattern of members attendance of group events, shown for LVHK group in Fig. 7 (bottom right).
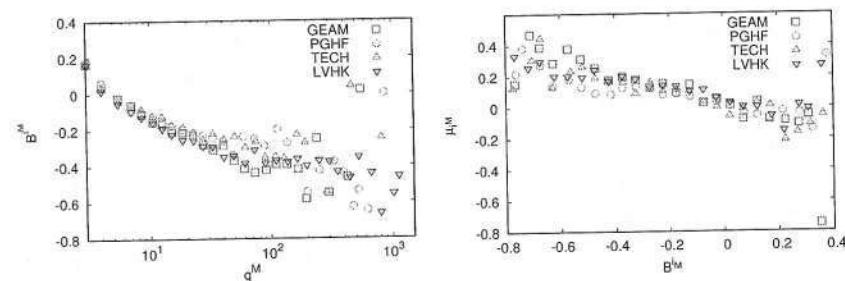
Figure 8. (left) Dependence of member's burstiness on the number of attended events. (right) Dependence of memory coefficient on burstiness.

It was shown in [49] that there are two different mechanisms responsible for the bursty nature of the signal: the inter-event time distribution and memory. Two measures used for distinguishing between these two mechanisms are burstiness parameter B and memory coefficient μ. The burstiness parameter is given by

$$B_{iM} = \frac{\langle \Delta T^{iM} \rangle - \sigma_{\Delta T^{iM}}}{\langle \Delta T^{iM} \rangle + \sigma_{\Delta T^{iM}}}, \qquad (18)$$

where $\langle \Delta T^{iM} \rangle$ and $\sigma_{\Delta T^{iM}}$ are average value and standard deviation of time lag between two attended events respectively. B_{iM} takes values between -1 and

1 and it is used for quantifying the inter-event time distribution of the signal: $B_{iM} = 1$ corresponds to very bursty signal, $B_{iM} = 0$ to random signal, and $B_{iM} = -1$ corresponds to very regular (periodic) signal. The memory coefficient μ, a correlation-based measure, measures autocorrelations between consecutive inter-event times $\left(\Delta T_k^{iM}, \Delta T_{k+1}^{iM}\right)$ for one member. It is calculated as

$$\mu_{iM} = \frac{1}{q^M - 2} \sum_{k=1}^{q^M-2} \frac{(\Delta T_k^{iM} - \langle \Delta T^{iM} \rangle_1)(\Delta T_{k+1}^{iM} - \langle \Delta T^{iM} \rangle_2)}{\sigma_1 \sigma_2}, \qquad (19)$$

where $\langle \Delta T^{iM} \rangle_1$ $\left(\langle \Delta T^{iM} \rangle_2\right)$ and $\sigma_1 (\sigma_2)$ are sample mean and sample standard deviation of ΔT_k^{iM} $\left(\Delta T_{k+1}^{iM}\right)$ respectively.

We quantify the temporal patterns of individual member's activities by calculating the burstiness parameter B_{iM} and memory coefficient μ_{iM} [49]. Figure 8 (left) shows how burstiness of individual members depends on the size of their activity. Figure 8 indicates that burstiness parameter decreases as the involvement of member in group grows, suggesting that very active members attend events on more regular basis than members with a small or moderate number of participations. This is in line with previous findings [26, 27] where it was shown that probability of members future involvement in group activities depends nonlinearly on the ratio between the number of previous participations and non-participations in group events. Specifically, for very active members their probability to attend the next event is very close to one and does not vary much with time, which results in relatively regular patterns of participation (see Fig. 7 (top left)). On the other hand, a member with a few attendances has the typical pattern shown in Fig. 7(top right): inter-event time distribution is more a Poissonian like with burstiness parameter close to zero. The dependence of memory coefficient μ_{iM} on burstiness parameter, shown in Fig. 8 (right), further confirms that members probability of attending the next group event depends on the number of previously attended and not-attended events. Members with the largest number of attended events, i.e. a negative value of burstiness parameter, have the highest and positive value of memory coefficient. This suggests that the origin of burstiness of their activity is memory and that regular members have more predictable group activity patterns than temporary members. These findings show that human event-participation dynamics is very different from one typically observed in social systems where human behavior is characterized with high burstiness parameter and memory coefficient close to zero.

As we saw in Section 3, members with medium and high number of atten-
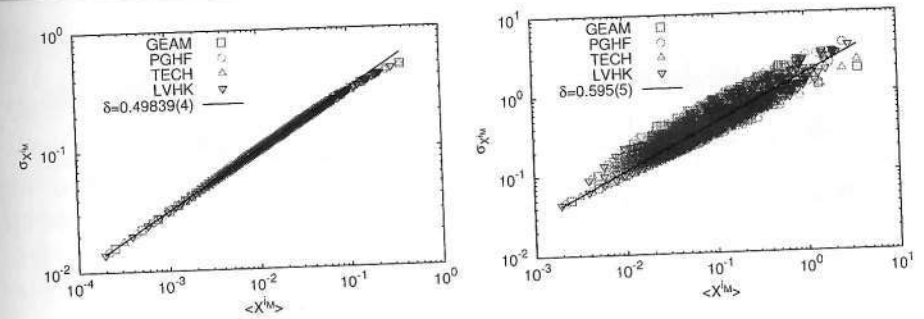
Figure 9. The relationship between fluctuations $\sigma_{X_{iM}}$ and average $\langle X_{iM} \rangle$ activity of member time series at the scale of one (left) and ten events (right).

dance have large circles of acquaintances within which we can distinguish subgroups of strongly connected friends. It was argued in [27] that these connections and member's need to maintain them are the primary drivers behind her involvement in group activities. On the other hand, it was shown that dynamics of various social groups is externally driven by the arrival of new members [9, 17]. We use formalism introduced in [51] to determine in quantitative manner which of the factors, endogenous or exogenous, are responsible for Meetup group dynamics. Based on time series of activity, for every member, we calculate time average activity $\langle X_{iM} \rangle$ and standard deviation $\sigma_{X_{iM}}$. The analysis of dynamics in various systems has shown that relationship between standard deviation and time average of activity follows the scaling law

$$\sigma_{X^{iM}} \propto \langle X_{iM} \rangle^{\delta}. \qquad (20)$$

The value of exponent $\delta \approx 0.5$ corresponds to system where endogenous factors dominate group dynamics, $\delta \approx 1$ to system driven by external forces, while the values of δ between these two extremal values denote mixed dynamics. Figure 9 (left) shows a perfect scaling for all four groups with value of exponent $\delta \approx 0.5$, which indicates that Meetup group dynamics is practically random on the time scale of one event. If we do the same analysis on the time series obtained by aggregating members activity over ten events (which corresponds to one week in real time), we see that the scaling exponent δ becomes larger than 0.5 suggesting correlated members activity at longer time scales, Fig. 9 (right).

The community structure of the network of the 10% most active members is a direct consequence of correlated co-occurrence of members at group events.

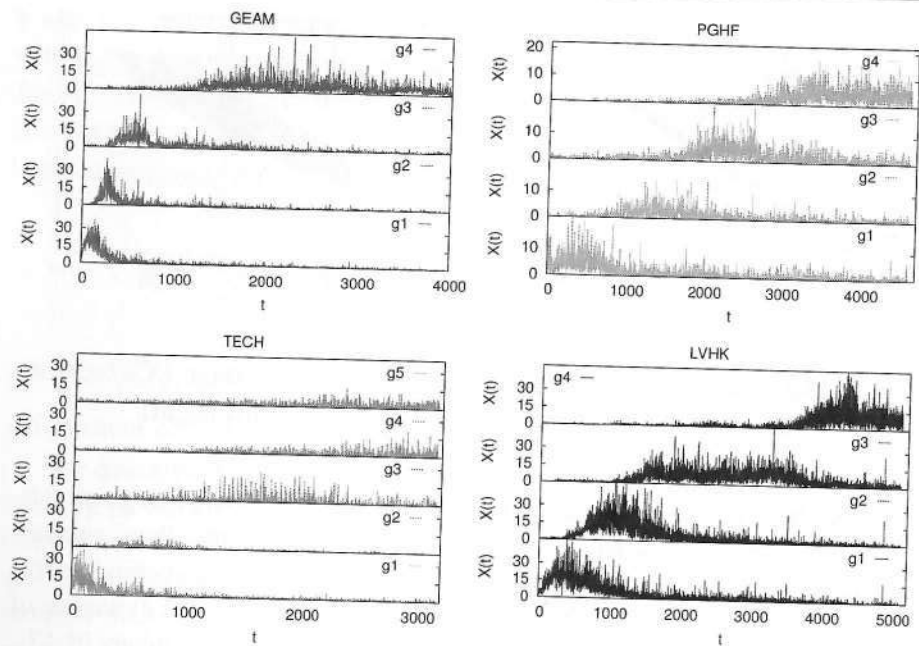Figure 10. Total activity of subgroups found in weighted networks of the 10% most active members using LOUVAIN method.

Figure 10 shows the temporal evolution of activity in these communities. The time series shown in Fig. 10 are obtained by aggregating time series of individual actions of all members belonging to one community. We see that all these communities have similar paths of evolution: at the beginning, the activity of their members is low, almost sporadic; then we observe a gradual growth, followed by the peak of community activity, after which the activity decreases until the community eventually ceases to exist. This rise and fall of social groups have been observed in both online and offline social systems [8, 9, 52]. The ability of group to dynamically alternate their composition is one of the main features of persistent large groups. Obviously, the activity of members belonging to one community is strongly correlated; we see that there is only small overlap between the activity of different communities. This further indicates that significant social relations are clustered in time and that these relations are main driving force behind the dynamics of Meetup social groups [26, 27].

Conclusion

Data from Meetup groups contain invaluable information about the human event-driven offline activity. We have demonstrated how the combination of computational methods from complex network theory and statistical physics can be used to explore the structure and dynamics of four different Meetup social groups. Specifically, we have shown how the data from these networks can be mapped onto bipartite networks of Members and Events, and how to project these networks onto Members partition to obtain social relations. These networks have several specific features that are universal for Meetup and, presumably other event-driven, social networks. In particular, both network types have very rich topology which is directly related to event-driven group dynamics. Bipartite networks are characterized by high degree heterogeneity in both partitions, which is not surprising, since power-law dependencies are the distinctive feature of members attendance patterns in event-driven social groups [26, 27]. What is surprising is a lack of degree-degree correlations in Members community, which suggests that members attend events of different sizes with equal probability. The weak degree-degree correlations observed for Events partition indicate that small events are predominately attended by active members of Meetup group, which confirms the findings of event size importance in [27]. The analysis of features of weighted networks of significant social links shows that network topology alone is not sufficient for capturing and understanding the social mechanism that govern the Meetup group dynamics and that link weights and their distributions need to be taken into account. The analysis of weighted measures shows that connection weights become important for dynamics of members with more than 50 connections. It also indicates that very active members tend to form strongly connected subgroups. The communities, identified as topological mesoscopic inhomogeneities in weighted social networks of the 10% most active members, are distinctive sign that these groups have self-organizing dynamics.

Meetup group collective dynamics is shaped by the behavior of individual member and interactions between them. The analysis of patterns of members behavior has again shown that human dynamics is bursty. What is different in the case of Meetup dynamics is that origin of this burstiness is memory, which is different from what is observed in other social dynamics. The positive value of memory coefficient for active users shows that their activity is correlated in time. We have shown that this correlation can be observed at the scale of ten and more event, which corresponds to one week in real time. This correlated behavior is

also the principal mechanism which leads to formation of communities in social network of the 10% most active users.

It has been shown earlier that Meetup group dynamics is universal, i.e. does not depend on group category, location or type of activity [27]. The comparison of this dynamics with conference participation patterns indicates that this universality is independent of members motivation to take part in groups activities, for instance, professional versus leisure. Here we show that this universality is also observed in the structure of social networks. Members of all four Meetup groups form networks which have a remarkably similar structure: they all exhibit similar degree distributions, have the same degree-degree correlations, and very similar community structure with similar number of communities.

The methodology that we demonstrated here by applying it to Meetup social groups is a standard part of social computational science, an emerging interdisciplinary scientific field. The presented methods and tools can be used in study of other social networks whose activity is localized in time and space.

Acknowledgment

References

[1] Lazer D, Pentland AS, Adamic L, Aral S, Barabasi AL, Brewer D, et al. Life in the network: the coming age of computational social science. *Science*. 2009; 323(5915):721.

[2] Conte R, Gilbert N, Bonelli G, Cioffi-Revilla C, Deffuant G, Kertesz J, et al. Manifesto of computational social science. *European Physical Journal-Special Topics*. 2012; 214:325.

[3] Castellano C, Fortunato S, Loreto V. Statistical physics of social dynamics. *Rev Mod Phys*. 2009 May;81:591–646.

[4] Sen P, Chakrabarti BK. *Sociophysics: an introduction*. Oxford University Press; 2013.

[5] Schelling TC. Dynamic models of segregation. *Journal of mathematical sociology*. 1971; 1(2):143–186.

[6] Putterman L. Cooperation and punishment. *Science*. 2010; 328(5978):578–579.

[7] Mitrović M, Paltoglou G, Tadić B. Networks and emotion-driven user communities at popular blogs. *The European Physical Journal B-Condensed Matter and Complex Systems*. 2010; 77(4):597–609.

[8] Mitrović M, Paltoglou G, Tadić B. Quantitative analysis of bloggers' collective behavior powered by emotions. *Journal of Statistical Mechanics: Theory and Experiment*. 2011; 2011(02):P02005. Available from: http://stacks.iop.org/1742-5468/2011/i=02/a=P02005.

[9] Mitrović M, Tadić B. Dynamics of bloggers' communities: bipartite networks from empirical data and agent-based modeling. *Physica A: Statistical Mechanics and its Applications*. 2012;391(21):5264–5278.

[10] Šuvakov M, Mitrović M, Gligorijević V, Tadić B. How the online social networks are used: dialogues-based structure of MySpace. *Journal of The Royal Society Interface*. 2013;10(79):20120819.

[11] Chatterjee A, Mitrović M, Fortunato S. Universality in voting behavior: an empirical analysis. *Scientific Reports*. 2013;3:1049.

[12] Radicchi F, Fortunato S, Castellano C. Universality of citation distributions: Toward an objective measure of scientific impact. *Proceedings of the National Academy of Sciences*. 2008;105(45):17268–17272.

[13] Dunbar RIM. Coevolution of neocortical size, group size and language in humans. *Behavioral and Brain Sciences*. 1993 12;16(4):681–694.

[14] Hill RA, Dunbar RIM. Social network size in humans. *Human Nature*. 2003;14(1):53–72.

[15] Dunbar RIM. Mind the gap: or why humans aren't just great apes. *Proceedings of the British Academy*. 2008;154:403–423.

[16] Dunbar RIM. Constraints on the evolution of social institutions and their implications for information flow. *Journal of Institutional Economics*. 2011;7(3):345–371.

[17] Mitrović Dankulov M, Melnik R, Tadić B. The dynamics of meaningful social interactions and the emergence of collective knowledge. *Scientific Reports*. 2015;5:12197.

[18] Aral S, Walker D. Identifying influential and susceptible members of social networks. *Science*. 2012;337(6092):337–341.

[19] González-Bailán S, Borge-Holthoefer J, Moreno Y. Broadcasters and hidden influentials in online protest diffusion. *American Behavioral Scientist*. 2013;57(7):943–965.

[20] Lin YR, Chi Y, Zhu S, Sundaram H, Tseng BL. Facetnet: a framework for analyzing communities and their evolutions in dynamic networks. In: *Proceedings of the 17th International Conference on World Wide Web*. WWW '08; 2008. p. 685–694.

[21] Garas A, Garcia D, Skowron M, Schweitzer F. Emotional persistence in online chatting communities. *Scientific Reports*. 2012;2:402.

[22] Yasseri T, Sumi R, Rung A, Kornai A, Kertsz J. Dynamics of conflicts in wikipedia. *PLoS ONE*. 2012 06;7(6):1–12.

[23] Montazeri A, Jarvandi S Soghraand Haghighat, Vahdani A Mariamand Sajadian, Ebrahimi M, Haji-Mahmoodi M. Anxiety and depression in breast cancer patients before and after participation in a cancer support group. *Patient education and counseling*. 2001;45:195–198.

[24] Davison KP, Pennebaker JW, Dickerson SS. Who talks? The social psychology of illness support groups. *American Psychologist*. 2000;55:205–217.

[25] Tam Cho WK, Gimpel JG, Shaw DR. The tea party movement and the geography of collective action. *Quarterly Journal of Political Science*. 2012;7:105–133.

[26] Smiljanić J, Chatterjee A, Kauppinen T, Mitrović Dankulov M. A theoretical model for the associative nature of conference participation. *PLoS ONE*. 2016 02;11(2):1–12.

[27] Smiljanić J, Dankulov MM. Associative nature of event participation dynamics: A network theory approach. *PloS one*. 2017;12(2):e0171565.

[28] Meetup;. Accessed: 2017-06-30. https://www.meetup.com/.

[29] Meetup Dataset;. Accessed: 2016-06-27. https://figshare.com/articles/Meetup_Datasets/2066904.

[30] Qiao Z, Zhang P, Zhou C, Cao Y, Guo L, Zhang Y. Event recommendation in event-based social Networks. In: *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence*. AAAI'14; 2014. p. 3130–3131.

[31] Zhang W, Wang J, Feng W. Combining latent factor model with location features for event-based group recommendation. In: *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD '13; 2013. p. 910–918.

[32] Liu X, He Q, Tian Y, Lee WC, McPherson J, Han J. Event-based social networks: linking the online and offline social worlds. In: *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD '12; 2012. p. 1032–1040.

[33] Mitrović M, Tadić B. Bloggers behavior and emergent communities in Blog space. *Eur Phys J B*. 2010;73(2):293–301.

[34] Dianati N. Unwinding the hairball graph: Pruning algorithms for weighted complex networks. *Physical Review E*. 2016 Jan;93:012304.

[35] Saracco F, Di Clemente R, Gabrielli A, Squartini T. Randomizing bipartite networks: the case of the World Trade Web. *Scientific Reports*. 2015;5:10595.

[36] Saracco F, Straka MJ, Di Clemente R, Gabrielli A, Caldarelli G, Squartini T. Inferring monopartite projections of bipartite networks: an entropy-based approach. *New Journal of Physics*. 2017;19(5):053022.

[37] Bianconi G, Coolen ACC, Perez Vicente C. Entropies of complex networks with hierarchically constrained topologies. *Physical Review E*. 2008;78(1):016114.

[38] Anand K, Bianconi G. Entropy Measures for Networks: Toward an Information Theory of Complex Topologies. *Phys Rev E*. 2009;80:045102(R).

[39] Boccaletti S, Latora V, Moreno Y, Chavez M, Hwang DU. Complex networks: structure and dynamics. *Physics reports*. 2006;424(45):175–308.

[40] Barrat A, Barthlemy M, Pastor-Satorras R, Vespignani A. The architecture of complex weighted networks. *Proceedings of the National Academy of Sciences of the United States of America*. 2004;101(11):3747–3752.

[41] Stanley MH, Buldyrev SV, Havlin S, Mantegna RN, Salinger MA, Stanley HE. Zipf plots and the size distribution of firms. *Economics letters*. 1995;49(4):453–457.

[42] Gabaix X, Ioannides YM. The evolution of city size distributions. *Handbook of regional and urban economics*. 2004;4:2341–2378.

[43] Fortunato S, Castellano C. Scaling and universality in proportional elections. *Physical Review Letters*. 2007;99(13):138701.

[44] Chatterjee A, Ghosh A, Chakrabarti BK. Universality of citation distributions for academic institutions and journals. *PloS one*. 2016;11(1):e0146762.

[45] Mitzenmacher M. A brief history of generative models for power law and lognormal distributions. *Internet mathematics*. 2004;1(2):226–251.

[46] Blondel VD, Guillaume JL, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*. 2008;2008(10):P10008.

[47] Lambiotte R, Delvenne JC, Barahona M. *Laplacian dynamics and multiscale modular structure in networks*. arXiv preprint arXiv:08121770. 2008;.

[48] Bastian M, Heymann S, Jacomy M. *Gephi: an open source software for exploring and manipulating networks*. 2009;.

[49] Goh KI, Barabási AL. Burstiness and memory in complex systems. *Europhysics Letters*. 2008;81(4):48002.

[50] Karsai M, Kaski K, Barabási AL, Kertész J. Universal features of correlated bursty behaviour. *Scientific Reports*. 2012;2.

[51] De Menezes MA, Barabási AL. Fluctuations in network dynamics. *Physical Review Letters*. 2004;92(2):028701.

[52] Palla G, Barabási AL, Vicsek T. Quantifying social group evolution. *Nature*. 2007;446(7136):664–667.

Multiplex and Multilevel Networks

Edited by

Stefano Battiston

Department of Banking and Finance, University of Zurich, Switzerland

Guido Caldarelli

IMT, Piazza S. Francesco 19, Lucca, Italy
ISC, Istituto dei Sistemi Complessi, CNR, Italy
Catchy, Laboratory of Big Data Analysis, Roma, Italy

Antonios Garas

Chair of Systems Design, ETH Zurich, Switzerland

OXFORD

UNIVERSITY PRESS

Contents

3

Classifying Networks with *dk*-Series

Marija Mitrović Dankulov[1], Guido Caldarelli[2], Santo Fortunato[3], and Dmitri Krioukov[4]

[1] Scientific Computing Laboratory, Center for the Study of Complex Systems, Institute of Physics Belgrade, University of Belgrade, Pregrevica 118, 11080 Belgrade, Serbia
[2] IMT Alti Studi, Lucca, Italy
[3] Center for Complex Networks and Systems Research, School of Informatics and Computing and Indiana University Network Science Institute (IUNI), Indiana University, Bloomington, USA
[4] Department of Physics, Department of Mathematics, Department of Electrical & Computer Engineering, Northeastern University, Boston, MA, USA

3.1 Introduction

The theory of complex networks provides powerful tools for studying complex systems in various disciplines such as biology, social sciences, computer sciences, mathematics, and physics [210]. One of the main research directions in network science studies the structural properties of networks and how they affect the dynamical processes and functions of systems represented by these various networks [17]. The standard assumption is that a self-organizing system should evolve to a network structure that makes these dynamical processes, or network functions, efficient [209, 198, 50]. Thus, understanding the network structure also reveals the mechanisms underlying the evolution of the system represented by the network.

Topological structure of networks can be characterized by a great number of various measures describing the system organization at different levels. Measures such as degree, average neighbor degree, clustering coefficient, concentrations of small subgraphs, betweenness, the distribution of shortest paths, and spectral properties have been used to describe in quantitative manner features that are characteristic of wide classes of complex networks. It has been shown that many complex networks have fat-tailed degree distributions [16], possess the small world property [308], and are often organized in

communities [96]. There is a common belief that the evolution of networks with similar structural properties is governed by the same mechanism. For instance, preferential attachment is often used for modeling networks with fat-tailed degree distributions [38]. Thus, classifying networks according to their topological properties is of great importance for identifying these mechanisms and better understanding the evolution, and indirectly the function and dynamics, of various complex systems.

However, classifying network via such graph measures is problematic, as there is no systematic way to determine which of them should be used. Besides, the measures are interdependent, that is, they positively or negatively correlate with each other in a complex way [303, 124, 289, 98, 63]. For these reasons, it is quite difficult to classify the structure of networks in a unique way using these topological measures and thus identify the evolution mechanisms characteristic for each class. For instance, the small world property has been found in many real networks, including social networks and interareal cortical networks in the primate brain. Yet in social networks, which are sparse graphs, this property is due to randomness in the linking patterns between the nodes, whereas in cortical networks it is trivially the consequence of network's high density. Therefore, the evolution of these two networks has been driven by different mechanisms, so that they cannot be assigned to the same network class.

One way to address the problem of interdependence among network properties is to find which of them are significant for a given network, and thus for its function. The standard procedure for the identification of a significant property X and its dependence on some other property Y is to generate a set of random graphs that have property Y but are random in all other respects, and then to check whether the property X is also characteristic of these graphs. If this is the case, then obviously property X is not interesting and relevant for the network function and dynamical processes running on it. We conclude that property X is a statistical consequence of property Y, and Y fully describes the structure of the network. Mechanisms generating network property Y can be thus considered to be relevant for the network evolution and dynamics. If X is not a typical property of these random graphs, one cannot conclude anything about the relevance of property X. The only conclusion that would follow from this is that property X is independent of property Y, but that does not mean that it is also independent of some other network property.

The identification of significant network features using the procedure described above raises another, equally important, question about the choice of null models. Since there are infinitely many network properties Y, there are infinitively many null models defined by property Y, and these can be used to test the statistical significance of any other property X [297]. For example, for most properties X, including motifs [195], their significance is tested with respect to random graphs with the same degree distribution. Although the choice of degree distribution as a Y property can seem natural, given the fundamental role played by it [210], there is no evidence that this choice of null model is less arbitrary than others. In general, there can be some other property which can be explanatory for both Y (here, the degree distribution) and X. Thus, one needs to identify the right reference property or properties Y in the null model that should be used for the testing of the (statistical) significance of property X.

In a recent work, Orsini *et al.* [219] proposed a way to identify such basic properties, enabling us to do a complete systematic description and unique classification of the structure of real networks. It is based on a set of properties known as the *dk-series* [182], a converging series of basic interdependent degree- and subgraph-based properties that characterize the local network structure at an increasing level of detail. It has been shown [182, 219] that the *dk*-series also defines a corresponding series of null models or random graph ensembles. These random graph models have exactly the same distribution of subgraphs of size d for all d-ples of nodes with degree $(k_1, k_2, k_3, \ldots, k_d)$ as in the real network. Or, to be precise, they are random graphs with fixed average degree, degree distribution, degree correlations, clustering, and so on. In Ref. [219], the authors used this methodology to quantify the randomness of six real single-layer networks, of very different function and dynamics. They showed that random graphs with fixed degree distribution, degree–degree correlations, average clustering, and degree-dependent average clustering reproduce all relevant topological properties for most networks. Here, we apply this approach to three networks and show that they differ in the randomness of their structure. We show that although many network properties can be reduced to specific degree- and subgraph-based characteristics, some of them cannot be explained with *dk*-series.

In recent years, a lot of attention in network science has been devoted to networks in which the same set of nodes are connected with multiple links of different types. These networks are referred to as multiplex or multilayer networks, since they consist of correlated single-layer networks composed of links of the same type. Many of the topological measures used to describe the structure of single-layer networks have been adapted in order to characterize the structure of layers and correlations between them [37]. We show how *dk*-series can be extended to describe in a systematic way the structure of multiplex networks using *dk*-annotated series [79].

3.2   *dk*-Series for single-layer networks

As indicated in the previous section, one needs to find an ordered set of reference properties of networks $Y_0, Y_1, \ldots$, satisfying some criteria. The first criterion is *inclusiveness*: every subsequent property provides more details about the network structure than its predecessor. Formally, this is equivalent to the requirement that networks with property Y_d, $d > 0$, should also have all properties prior to it, that is, all properties $Y_{d'}$, where $0 \le d' < d$. The second criterion is *convergence*, that is, the minimal set of properties has to be finite, that is, the last property in series Y_D should fully characterize the adjacency matrix of any given graph. The Y-series that satisfies these conditions allows us to claim that, for any property X that is deemed important in a given real network, we can find a minimal d^* such that the property Y_{d^*} explains property X. The convergence of the series ensures the existence of some d^*, while the inclusiveness means that random networks with Y_d $(d = (d^* + 1), \ldots, D)$ also have property X, so if we go to higher values of d, the random network has property X along with other significant properties. This enables the classification of network structure in a systematic manner. Several

approaches, including motifs [195], graphlets [314], and similar constructions [212], try to fully characterize the structure of networks by using relatively small set of properties, but they all violate the inclusiveness condition. On the other hand, one can still define many Y-series satisfying both conditions. We chose *dk-series* [182], the most natural choice due to their simplicity and the fact that they are a combination of subgraph- and degree-based characteristics of networks.

In these series, properties Y_d are *dk*-distributions. Each *dk*-distribution is actually a collection of distributions, stating how the subgraphs of size d are distributed over nodes with degrees $k, k', k'', \ldots, k^d$ in a graph G. Note here that the isomorphic subgraphs of G involving nodes of different degrees are thus counted separately. Specifically, the $0k$ "distribution" is simply the average degree $\bar{k} = \frac{2M}{N}$, where N and M are the number of nodes and links in a given graph, while the $1k$ distribution is the number of subgraphs of size 1, nodes, with the degree k, that is, the standard degree distribution

$$P(k) = \frac{N(k)}{N},\tag{3.1}$$

where $N(k)$ is the number of nodes of degree k. The $2k$ distribution counts how many nodes of degrees k and k' are forming subgraphs with two nodes, and is known as joint degree matrix $P(k, k')$

$$P(k, k') = \frac{\mu(k, k')M(k, k')}{2M},\tag{3.2}$$

where

$$\mu(k, k') = \begin{cases} 2 & \text{if } k = k', \\ 1 & \text{otherwise.} \end{cases}\tag{3.3}$$

The $3k$ distribution is a set of two distributions corresponding to two non-isomorphic subgraphs of size 3: wedges $\wedge$ and triangles $\triangle$. It characterizes the connectivity patterns between triples of nodes of degrees k, k', and k'':

$$P_\wedge(k', k, k'') = \mu(k', k'')\frac{N_\wedge(k', k, k'')}{2W},\tag{3.4}$$

$$P_\triangle(k, k', k'') = v(k, k', k'')\frac{N_\triangle(k, k', k'')}{6T},\tag{3.5}$$

where W and T are the total numbers of wedges and triangles in the network, and

$$v(k, k', k'') = \begin{cases} 6 & \text{if } k = k' = k'', \\ 1 & \text{if } k \neq k' \neq k'', \\ 2 & \text{otherwise,} \end{cases}\tag{3.6}$$

so that both $P_\wedge(k',k,k'')$ and $P_\triangle(k,k',k'')$ are normalized, and $\sum_{k,k',k''} P_\wedge(k',k,k'') = \sum_{k,k',k''} P_\triangle(k,k',k'') = 1$. We could continue in a similar manner to obtain a 4k distribution that consists of six distributions, each corresponding to one of the six non-isomorphic subgraphs of size 4 and so on until we reach $d = N$, the Nk-distribution, which characterizes the whole adjacency matrix of a given graph (see Figure 3.1).

The *dk*-series is directly related to some of the standard topological measures from complex network theory. Besides the already mentioned average degree and degree distribution, which are directly related to the 0k and 1k distributions, the 2k distribution defines the node degree correlations in networks, or network's assortativity. The average neighbor degree $\bar{k}_n n(k)$ is a projection of $P(k,k')$ via

$$\bar{k}_{nn}(k) = \frac{\sum_{k'} k' P(k,k')}{\sum_{k'} P(k,k')}.$$

The average clustering coefficient $\bar{c}$ and degree-dependent clustering coefficient $\bar{c}(k)$ can be calculated based on the number of triangles in the network. Specifically,

$$\bar{c} = \frac{1}{N} \sum_i \frac{2\triangle_i}{k_i(k_i - 1)},$$

and

$$\bar{c}(k) = \frac{6T}{N} \frac{\sum_{k',k''} P_\triangle(k,k',k'')}{k(k-1)P(k)}, \tag{3.7}$$

where $\triangle_i$ is the number of triangles composed of node i and its neighbors, while T is the total number of triangles. In general, the arbitrary *dk*-distribution characterizes both degree correlations between nodes at the hop distances $d' < d$, and the frequencies of d'-sized subgraphs, $d' \leq d$, in graph G.

One can easily see that *dk*-series is inclusive. The $(d+1)k$-distribution contains the same information about the network structure at the level d as the *dk*-distribution, plus some additional information about the degree correlations at the level $d + 1$. Specifically, the 1k-distribution defines the average degree (0k-distribution), via

$$\bar{k} = \sum_{k'} k' P(k'),$$

while the 1k-distribution can be obtained from the 2k distribution as

$$P(k) = \frac{\bar{k}}{k} \sum_{k'} P(k,k').$$

Similarly, the 3k-distribution defines the 2k-distribution by

$$P(k, k') = \frac{1}{k + k' - 2} \sum_{k''} \left\{ \frac{6T}{M} P_\triangle(k, k', k'') \right.$$
$$\left. + \frac{W}{M} \left[P_\wedge(k', k, k'') + P_\wedge(k, k', k'') \right] \right\}. \tag{3.8}$$

The opposite does not hold, that is, knowing the *dk*-distribution will not allow one to infer anything about the $(d+1)k$-distribution, meaning that higher values of d correspond to a greater level of details about the network structure.

The number of non-isomorphic subgraphs, and thus the number of distributions needed to characterize network topology at the level d, grows exponentially with d; hence, the calculation of *dk*-series becomes a computationally intensive task for higher values of d. One could argue that just counting the number of d-sized subgraphs in a given network regardless of their node degrees should be enough for the description of network structure in a systematic manner [195, 314, 212]. The subgraph-based series obtained from the count of d-subgraphs without including information about the degree, which we can call a d-series, satisfies the convergence condition, and the statistics for $d = N$ subgraphs would also fully describe the topology of a given network, but, unlike *dk*-series, they are not inclusive. Careful analysis of the first four elements of a d-series clearly demonstrates its non-inclusiveness. For the d-series, the zeroth element is not defined, while the number of nodes N and number of edges M are the properties corresponding to $d = 1$ and $d = 2$, respectively. These two quantities are independent of each other, that is, knowing the number of edges does not allow one to tell much about the number of nodes in a network. Similarly, the properties at the level $d = 3$, the number of triangles T and wedges W, define neither the size nor the density of the network [219]. This analysis demonstrates that the elements in the d-series are independent of each other and that each of them conveys a different kind of information about network topology.

Figure 3.1(b) illustrates the inclusiveness and convergence of *dk*-series, and also suggests that all graphs with N nodes and M edges constitute a set of random graphs $\mathcal{G}_{0k} = \mathcal{G}_{N,M}$ with the same $0k$ property. Graphs with the same degree sequence, the $1k$ property, form a smaller set of graphs $\mathcal{G}_{1k}$, which is a subset of $\mathcal{G}_{0k}$, and so on. Each set of graphs with a given *dk*-distribution, known as *dk-graphs*, is at the same time a subset of $(d-1)k$- and a superset of $(d+1)k$-graphs. It follows from this that a sequence of *dk*-distributions defines a sequence of random graph ensembles (null models). In order to compare a real network with the random graphs which have the same *dk*-properties, and thus quantify the randomness of its structure of the real network, one needs a maximum entropy ensemble of these graphs or *dk-random graphs* [182]. All graphs in *dk*-random graphs have equal sampling probability $P(G) = 1/\mathcal{N}_d$, where $\mathcal{N}_d$ is the number of *dk*-graphs. Each collection of *dk*-distributions is more informative about the network structure and thus more constraining than $(d-1)k$-distributions, that is, $\mathcal{N}_0 \geq \mathcal{N}_1 \geq \cdots \geq \mathcal{N}_N = 1$. The size of the final ensemble at level N is clearly equal to 1, since it just contains the network with the exact adjacency matrix. The number $\mathcal{N}_d$ is too
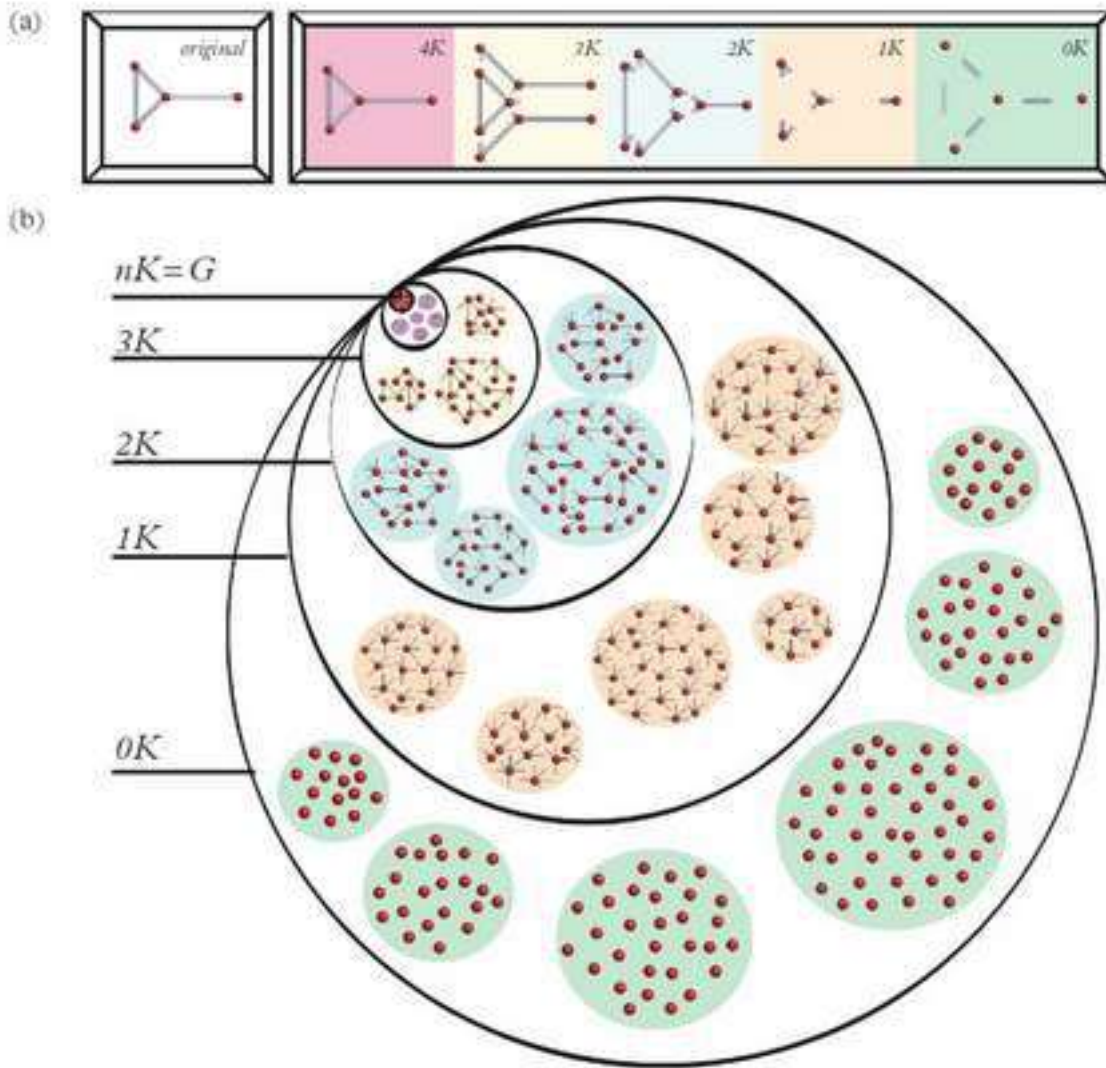
Figure 3.1 *The illustration of* dk-*series. (a) The* dk-*distributions for a graph of size 4. The 4k-distribution is the graph itself. The 3k-distribution consists of its three subgraphs of size 3: one triangle connecting nodes of degrees 2, 2, and 3, and two wedges connecting nodes of degrees 2, 3, and 1. The 2k-distribution is the joint degree distribution in the graph. It specifies the number of links (subgraphs of size 2) connecting nodes of different degrees: one link connects nodes of degrees 2 and 2, two links connect nodes of degrees 2 and 3, and one link connects nodes of degree 3 and 1. The 1k-distribution is the degree distribution in the graph. It lists the number of nodes (subgraphs of size 1) of different degree: one node of degree 1, two nodes of degree 2, and one node of degree 3. The 0k-distribution is just the average degree in the graph, which is 2. (b) The inclusiveness and convergence of the* dk-*series are illustrated via the hierarchy of* dk-*graphs, which are graphs having the same* dk-*distribution of a given graph* G *of size* N. *From Ref. [219].*

large, especially for small values of d (exact or approximate calculations for $d = 0, 1, 2$ can be found in [32, 20]), making the construction of the whole set of *dk*-graphs impossible. Thus, one needs to sample *dk*-random graphs uniformly in order to be able to compare them with a given real network and properly test the significance of different topological properties.

The *dk*-series and *dk*-random graphs enable the systematic and full characterization of the structure of any real network by finding the value of d for which all $d'k$-distributions for $d' > d$ do not contain any additional information of the network structure. This means that any topology metric one can define on network G is captured with *dk*-random graphs. The convergence and inclusiveness properties of *dk*-series ensure that this value d exists, that is, they guarantee that any network property X of any given network G can be reproduced with any desired accuracy by high-enough d. Clearly, all properties are reproduced exactly for $d = N$, but the question is whether there is a value $d < N$ for which all relevant topological properties of a given network are captured with *dk*-random graphs. By finding this value d, one also quantifies the randomness of the structure of a given network. The entropy of *dk*-ensembles is $S_d = \ln \mathcal{N}_d$, and it is a nonincreasing function of d, that is, the *dk*-random graphs are *less random and more structured*, the higher d is. In the following section, we demonstrate how one can classify single-layer networks based on their *dk*-randomness by applying the procedure described in Ref. [219] to three real networks.

3.2.1   Classifying single-layer networks based on their *dk*-randomness

First, we briefly discuss the constructibility of *dk*-random graphs, and the problem of sampling graphs uniformly at random from the sets of *dk*-graphs. Here, we emphasize that, in *dk*-graphs, the *dk*-distribution constraints are sharp, that is, all graphs in *dk*-graphs set have exactly the same *dk*-distribution. Given a real network G, there exist two ways to sample *dk*-random graphs: *dk*-randomize G, generalizing the randomization algorithms in Refs. [186, 187], or construct random graphs with G's *dk*-sequence from scratch [182, 117], also called direct construction [148, 76, 147, 22]. We chose the first option, *dk*-randomization, due to its simplicity and the existence of algorithms that enable the uniform sampling of *dk*-random graphs for values of d greater than 2 (see the detailed discussions about construction algorithms in [182, 219]).

The *dk*-randomization is an edge-swapping procedure where pairs of edges are swapped at random, starting from G, such that the *dk*-distribution is preserved at each swap. Figure 3.2 illustrates permitted swaps of edges for each *dk*-distribution. Specifically, to preserve 0*k*-distribution, average degree, we disconnect a pair of nodes and connect two other, non-neighboring, nodes. The graphs obtained in this procedure are Erdős-Rényi graphs $\mathcal{G}_{N,M}$ of fixed size N and average degree $2M/N$. To preserve the degree sequence (1*k*-distribution), we chose at random a pair of edges and swapped their targeting nodes, while, for the 2*k*-distribution, we swapped edge pairs only if there were at least two nodes of equal degrees adjacent to different edges belonging to this pair. Allowed 3*k*-swaps are then 2*k*-swaps that preserve 3*k*-distribution, the same connectivity patterns between the triplets of nodes with respect to node degrees. From this and the inclusiveness of *dk*-series, it follows that $(d+1)k$-swaps form a subset of *dk*-swaps for $d > 0$ [182]. During all these rewiring procedures, the edge swapping is only allowed if it does not lead to the creation of multiple edges between the same pair of nodes.

There are many concerns regarding the described rewiring procedure [316], two of which are particularly important: (1) the ergodicity of the rewiring process, that is, whether any two pairs of graphs with the same dk-properties are connected with a chain of dk-swaps; (2) the uniformity of the rewiring process, that is, how close to uniform sampling the dk-swap Markov chain is after its mixing time is reached. It has been shown that dk-swapping processes for values of $d = 1, 2$ are ergodic [186, 187, 67], while it is common belief that there is no ergodic edge swapping, of any type, that preserves the $3k$-distribution, and thus dk-distributions for values $d \geq 4$, although a rigorous proof of this is lacking at the moment [219]. When it comes to the uniformity of the dk-swapping process for $d = 0, 1, 2$, it has been shown that if the edge-swap process is done correctly, then the sampling is uniform [2, 11].

Since the $2k$-random graphs do not capture all topological properties for most of the tested real networks [137, 219], we need algorithms that will allow us to go beyond preserving only $2k$-properties. The dk-targeting $d'k$-preserving rewiring, where $d' < d$, has proven to be a good choice for generating random graphs with the same dk-properties as in the considered real network [182, 219]. This procedure incorporates the following modification of the $d'k$-rewiring algorithm: the $d'k$-swap is accepted with probability $\min(1, \exp(-\beta \Delta H))$, where β is the inverse temperature of this simulated annealing process, and ΔH is the change in the L^1 distance between the dk-distribution in the current graph and the targeted dk-distribution before and after the swap. The numerical experiments with $3k$-targeting rewiring have shown that this process does not converge for most real networks [219], due to the extremely constraining nature of the $3k$-distribution. Therefore, it is reasonable to retreat to numeric investigations of $2k$-random graphs in which, in addition to the $2k$-distribution, some substatistics of the $3k$-distribution are fixed. In particular, we consider $2.1k$-random graphs, which have the same $2k$-distribution and value of the average clustering coefficient $\bar{c}$ as the given real network, and $2.5k$-random graphs with the same $2k$-properties and average clustering coefficient $\bar{c}(k)$ of nodes of degree k [117]. Since $2.1k$- and $2.5k$-statistics are fully defined by the $3k$-distribution, and $2.1k$ is defined by $2.5k$, the $3k$-random graphs comprise a subset of $2.5k$-random graphs, which are, in turn, subsets of $2.1k$-random graphs, that is, $\mathcal{N}_2 > \mathcal{N}_{2.1} > \mathcal{N}_{2.5} > \mathcal{N}_3$. As a consequence, if a certain topological property of real networks is captured by $2.5k$-random graphs, it will be also captured by $3k$-random graphs, while the opposite is not generally true.

The scheme of algorithm(s) that we use for creating a set of dk-random graphs for $d = 0, 1, 2, 2.1, 2.5$ is given in Figure 3.2, while their detailed description and a link to a Web page with publicly available software can be found in Ref. [75]. The dk-random graphs for $d = 0, 1, 2$ are created using the standard dk-swapping described above. Although it is known that these procedures for general graphs do not lead to a uniform sampling of dk-random graphs, unlike their modified versions [2, 11], it has been shown that, for power-law distributions, the obtained sample of uniform graphs is very close to uniform. To generate dk-random graphs for $d = 2.1, 2.5$, we start with a $2k$-random graph and apply to it a described $2k$-preserving $2.xk$-targeting ($x = 1, 5$) rewiring process (see Figure 3.2). For this, we use a modified version of the algorithm [63, 219], which ensures the convergence for all networks.
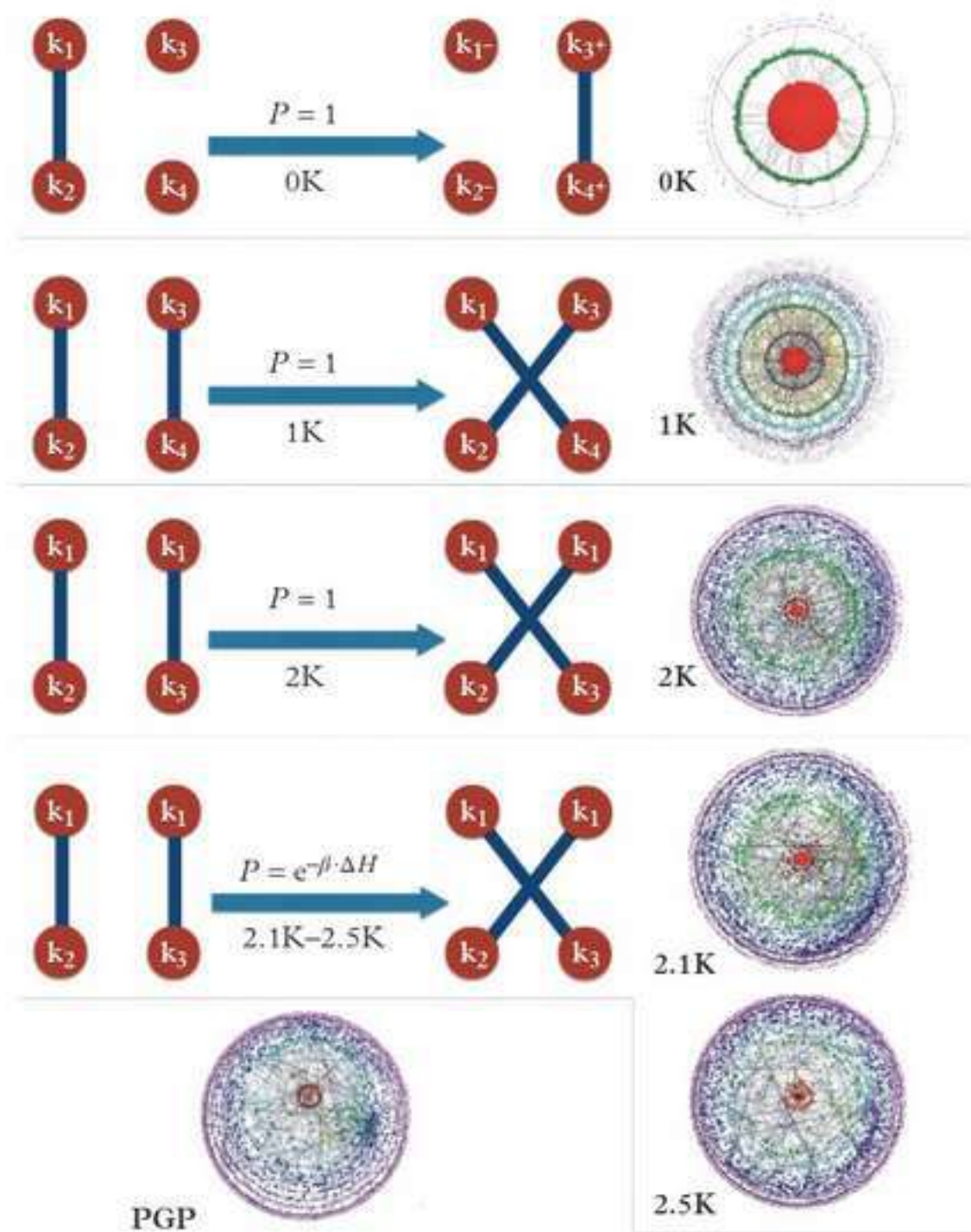
Figure 3.2 dk-*sampling and convergence of dk-series. The left column shows the elementary swaps of* dk-*randomizing (for* d = 0, 1, 2*) and* dk-*targeting (for* d = 2.1, 2.5*) rewiring. The nodes are labeled by their degrees, and the arrows are labeled by the rewiring acceptance probability. In* dk-*randomizing rewiring, random (pairs of) edges are rewired preserving the graph's* dk-*distribution (and consequently its* d'K-*distributions for all* d' < d*). In* 2.1k- *and* 2.5k-*targeting rewiring, the moves preserve the* 2k-*distribution, but each move is accepted with probability* p *designed to drive the graph closer to a target value of average clustering* $\bar{c}$ *(2.1k) or degree-dependent clustering* c(k) *(2.5k):* $p = \min(1, e^{-\beta \Delta H})$, *where* β *is the inverse temperature of this simulated annealing process,* $\Delta H = H_a - H_b$, *and* $H_{a,b}$ *are the distances, after and before the move, between the current and target values of clustering:* $H_{2.1k} = |\bar{c}_{current} - \bar{c}_{target}|$ *and* $H_{2.5k} = \sum_i |\bar{c}_{current}[k_i] - \bar{c}_{target}[k_i]|$. *The right column shows LaNet-vi [29] visualizations of the results of these* dk-*rewiring processes, applied to the Pretty Good Privacy (PGP) network, visualized at the bottom of the left column. The node sizes are proportional to the logarithm of their degrees, while the color reflects node coreness [29]. As* d *grows, the shown* dk-*random graphs quickly become more similar to the real PGP network. From Ref. [219].*

To quantify the randomness of the real network, that is, to determine the value of d for which dk-random graphs capture most of its topological properties, we adopt the following procedure. For a given real network, we calculate its average degree, degree distribution, degree correlations, average clustering coefficient, and averaging clustering coefficient of nodes of degree k; then, based on this, we generate 20 dk-random graphs, using the methodology described in the previous paragraph, for $d = 0, 1, 2, 2.1, 2.5$. Then, for each sample, we compute a variety of network properties and compare their values with the corresponding ones obtained for the real network. The value of d for which the considered properties of dk-random graphs are in reasonable agreement with the ones of a real network determines the randomness of its structure. The higher the value of d, the more structured and less random a given network is.

In Ref. [219] the authors performed an extensive set of numeric experiments with six real, very different networks with respect to their function. Here, we demonstrate the described procedure by applying it to three of these networks: the Internet at the level of autonomous systems (INTERNET) [183], a technosocial web of trust among users of the distributed Pretty Good Privacy (PGP) cryptosystem [39], and a functional MRI (fMRI) map of the human brain (BRAIN) [85]. In the first network, INTERNET, the nodes are so-called autonomous systems (ASs; organizations owning parts of the Internet infrastructure), and there is a link between two ASs if they have a business relationship in which they exchange Internet traffic. The nodes in the second network we consider, PGP, are users' PGP certificates, while the edges denote the existence of trust between two users. We consider here only the largest connected component of the PGP network. The third network considered, BRAIN, is the largest component of an fMRI map of the human brain, where voxels (representing small areas of a resting brain, approximately 36 mm^3 in volume) are represented with nodes, and an edge exists between two voxels if the correlation coefficient of the fMRI activity of the voxels exceeds 0.7. We chose these three networks because they have different values of d for which dk-random graphs capture their structural properties [219]. In particular, most of the considered properties for INTERNET are reproduced with $2k$-random graphs, which makes it the most random network among the three networks, while, to reproduce the same properties of the PGP network, we need graphs with preserved $2.5k$-distribution. Some of the properties of the BRAIN network are not reproduced even with $2.5k$-random graphs, meaning that this network is the least random one among these three networks (Figures 3.3–3.7 and Tables 3.1 and 3.2).

The properties that we use to compare the structure of real networks with the ones of random graphs can be divided into three categories: microscopic, mesoscopic, and macroscopic. The microscopic properties describe the networks' structure at the level of individual nodes and subgraphs of small sizes (see Figures 3.3 and 3.4). Some of these properties, namely, average degree, degree distribution, average degree of nearest neighbors, average clustering coefficient, and average clustering coefficient of nodes of degree k, are fixed by the corresponding dk-distributions. On the other hand, the concentration of subgraphs of size 3 and 4 [219], as well as the distribution of the number of common neighbors shared by a pair of nodes, are not fixed by dk-distributions for $d < 3$. The distribution of common neighbors equals the probability that two connected
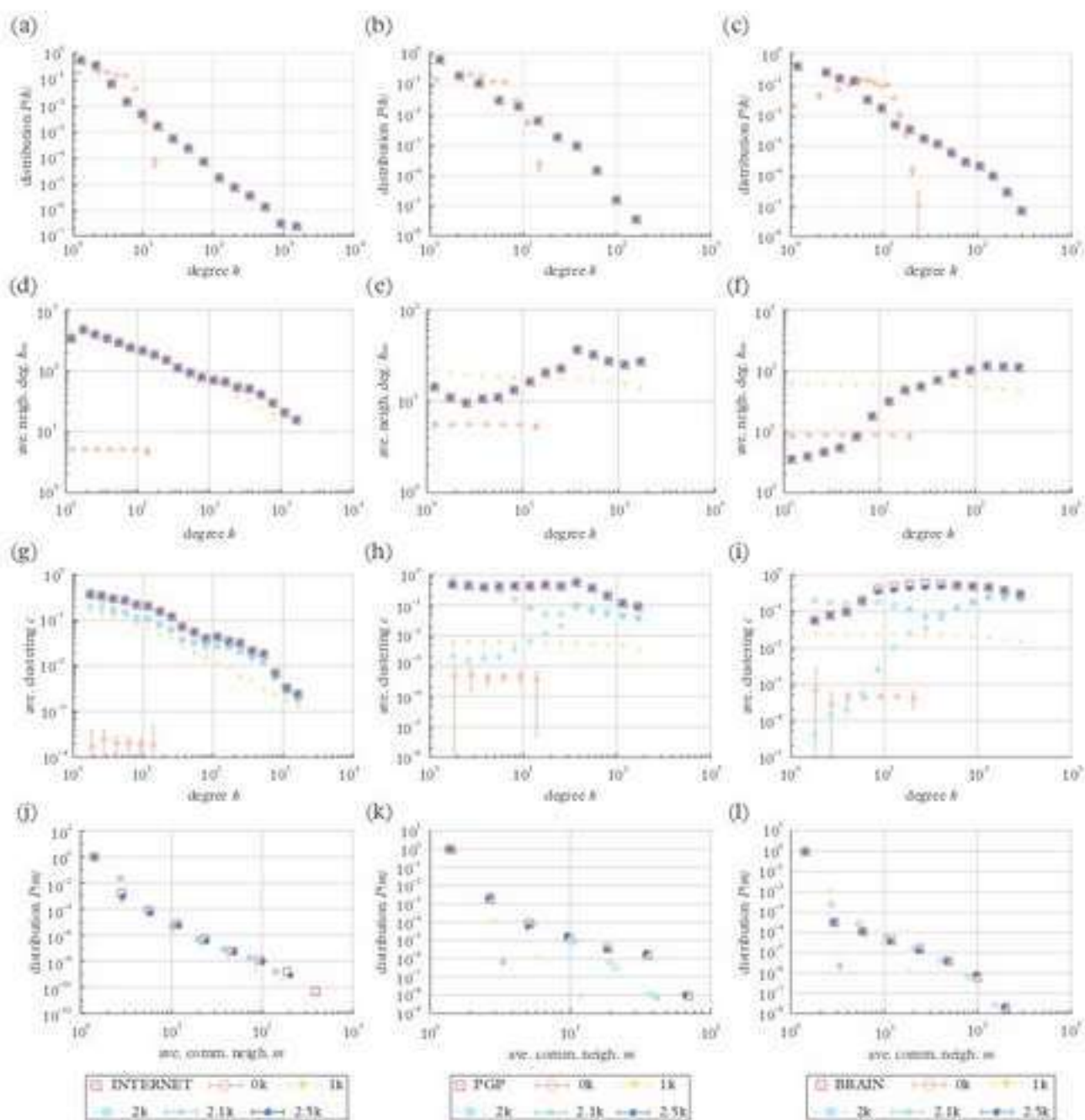
Figure 3.3 *Microscopic properties of real complex networks and their* dk-*random graphs. The first nine panels show topological properties fixed by* dk-*distributions: the degree distribution* P(k) *for (a) INTERNET, (b) PGP, and (c) BRAIN; the average degree of nearest neighbors* $\bar{k}_{nn}$(k) *for (d) INTERNET, (e) PGP, and (f) BRAIN; and the degree-dependent average clustering coefficient* $\bar{c}$(k) *for (g) INTERNET, (h) PGP, and (i) BRAIN. The last three panels show the distribution of the number of common neighbors* P(m), *which is not fixed by* dk-*distributions for* d < 3: *(j) INTERNET, (k) PGP, and (l) BRAIN. Adapted from Ref. [219].*

Figure 3.4 *Density of subgraphs of sizes 3 and 4. The average absolute difference between the subgraph concentration in the* dk-*random graphs and its concentration in the real networks: INTERNET (top), PGP (middle), and BRAIN (bottom). Adapted from Ref. [219].*

nodes have m common neighbors and is exactly fixed by the $3k$-distribution. Mesoscopic properties depend on both local and global network organization. Here, we consider k-coreness [9] and k-density [250] (see Figure 3.5). A node has k-coreness equal to k if it belongs to k-core of the original graph, which is the largest induced subgraph of graph in which every node has degree at least k. Similarly, an edge has k-denseness equal to k if it belongs to the largest induced subgraph of the original graph in which all edges have multiplicity at least k. Macroscopic properties are truly global: betweenness, the distribution of hop lengths of shortest paths, and spectral properties (see Figure 3.6 and Tables 3.1 and 3.2). We measure the distance between real and dk-random graphs with Kolmogorov–Smirnov distances between the distributions of all the considered properties (see Figure 3.7).

For all three networks and for most of the considered properties, we observe a nice convergence as d increases, that is, there is no statistically significant difference between the property in the real network and in its $2.5k$-random graphs. Although this is expected
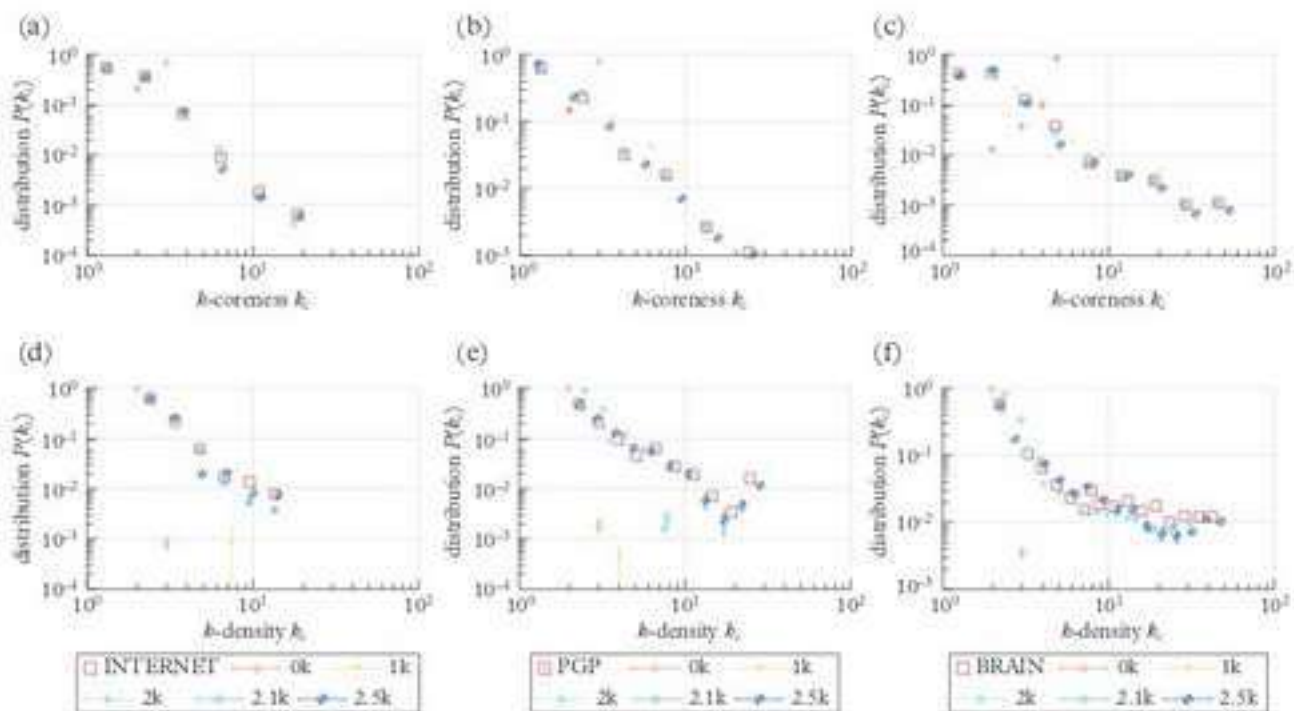
Figure 3.5 *Mesoscopic properties:* k-coreness and k-density distributions. The upper panels show the k-coreness of (a) INTERNET, (b) PGP, and (c) BRAIN networks, and the lower panels show their k-density: (e) INTERNET, (f) PGP, and (g) BRAIN. The k_c-core of a graph G is the maximal subgraph of G in which all nodes have degree at least k_c. A node has k-coreness k_c if it belongs to the k_c-core but not to the $(k_c + 1)$-core. The k_c-dense subgraph is the maximal subgraph of a graph in which all edges have multiplicity $(k_c - 2)$; the multiplicity of an edge is the number of triangles the edge is part of. Adapted from Ref. [219].

for microscopic properties that are fixed with *dk*-distributions, there is no reason to expect convergence in the case of small subgraph frequencies, in the distribution of the number of common neighbors, or for mesoscopic or macroscopic properties. Figure 3.4 shows that the relative difference between subgraph frequencies in real and 2.5*k*-random graphs is very close to zero for subgraphs of sizes 3 and 4. For INTERNET, this property is already captured with 1*k*-graphs, while it is clear that, for the BRAIN and PGP networks, one needs to fix the degree-dependent clustering coefficient in order to observe the same motif count as in the real systems. Mesoscopic properties are reproduced with 2.5*k*-graphs for the BRAIN and PGP networks, and with 2*k*-random graphs for the INTERNET network (see Figure 3.5). While betweenness and average shortest-path distance require 1*k*- and 2.5*k*-random graphs for INTERNET and PGP, respectively, such properties are not captured even with 2.5*k* graphs for BRAIN (see Figure 3.6). Table 3.1 shows that the largest eigenvalue of the adjacency matrix is closely, but not exactly, reproduced by $d = 2.5$ for all three networks. The spectral gap, the difference between the largest and second largest eigenvalue of the adjacency matrix, given in Table 3.2, shows that 2*k*- and 2.1*k*-random graphs are better connected and interlinked, compared to real networks. Figure 3.7 shows that the Kolmogorov–Smirnov distances calculated for the distributions of INTERNET and PGP are either zero or
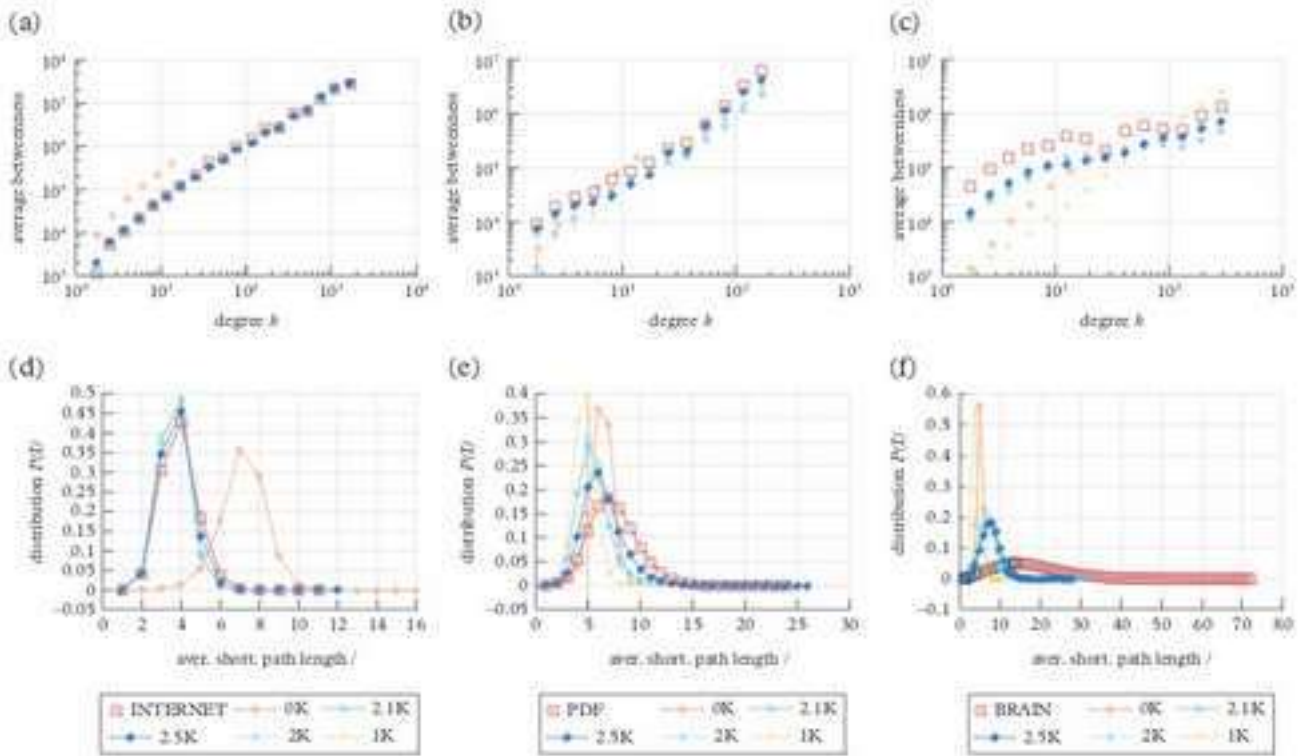
Figure 3.6 *Macroscopic properties: betweenness and average shortest hop distance. The average betweenness* $\bar{b}(k)$ *of nodes of degree* k *is shown in the upper panels: (a) INTERNET, (b) PGP, and (c) BRAIN. The lower panels show the distribution* P(l) *of the length* l *of the shortest paths between all pairs of nodes: (d) INTERNET, (e) PGP, and (f) BRAIN. Adapted from Ref. [219].*

Table 3.1 *Largest eigenvalues of the adjacency matrix for the three networks considered and their corresponding dk-graphs. For the latter, we show averages across different realizations for each d, and their standard deviations in parentheses.*

	Original	0*k*	1*k*	2*k*	2.1*k*	2.5*k*
INTERNET	67.17	5.36 (0.01)	56.02 (0.33)	61.15 (0.03)	61.32 (0.06)	65.34 (0.10)
PGP	42.44	5.77 (0.02)	19.50 (0.24)	34.08 (0.03)	34.40 (0.05)	42.95 (0.12)
BRAIN	119.66	8.91 (0.01)	54.89 (0.26)	113.41 (0.02)	114.09 (0.06)	122.27 (0.20)

Table 3.2 *Spectral gap between the largest and the second-largest eigenvalues of the adjacency matrix. For the dk-graphs, the shown values are the averages across different realizations for each d, while their standard deviations are reported in parentheses.*

	Original	0*k*	1*k*	2*k*	2.1*k*	2.5*k*
INTERNET	17.56	0.70 (0.05)	14.94 (0.53)	18.83 (0.07)	18.55 (0.11)	19.53 (0.25)
PGP	4.25	0.98 (0.04)	5.51 (0.31)	18.01 (0.18)	17.55 (0.21)	4.71 (0.19)
BRAIN	40.97	2.90 (0.06)	35.52 (0.31)	77.53 (0.11)	76.59 (0.27)	42.71 (0.35)
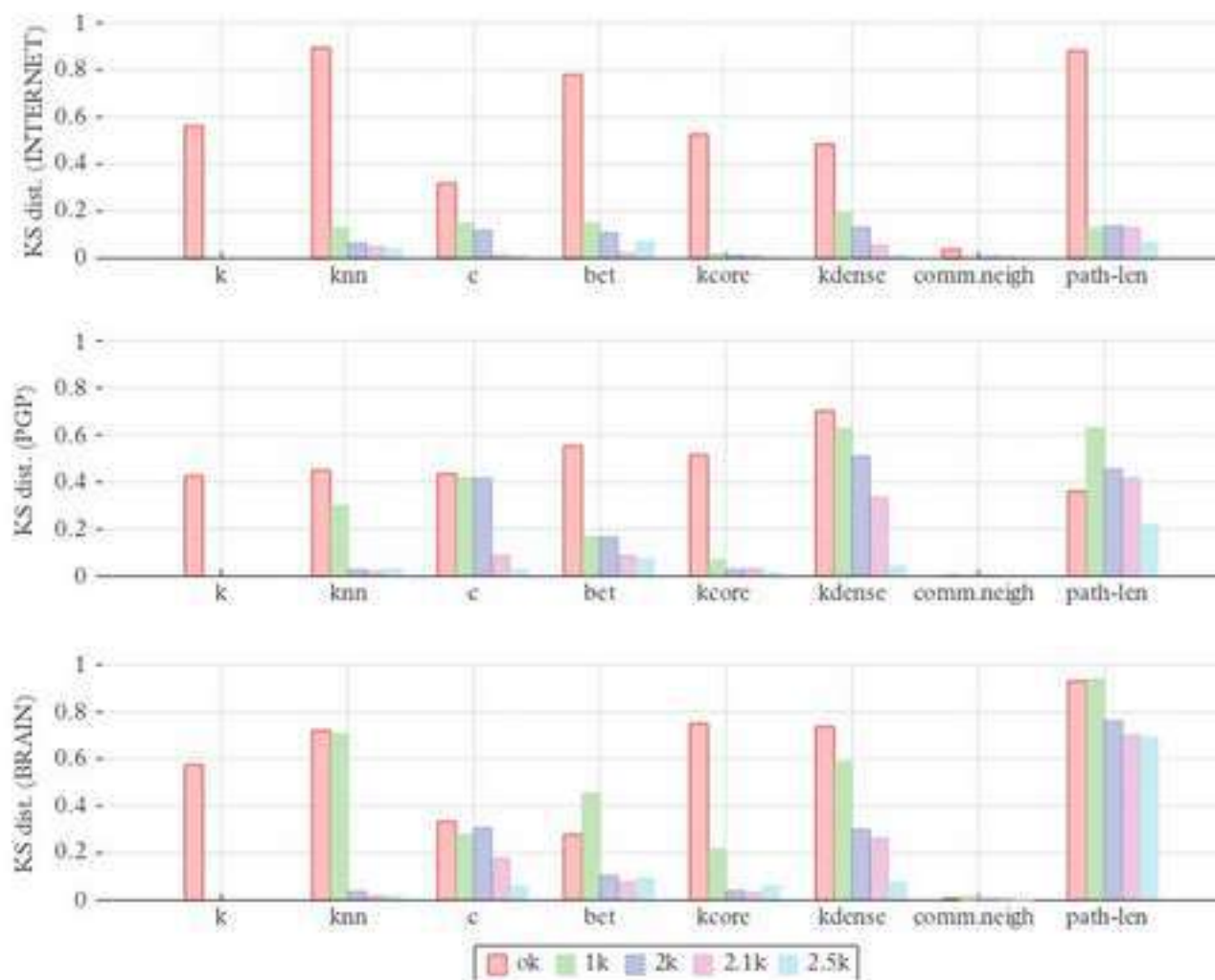
Figure 3.7 *Kolmogorov–Smirnov distance between real networks and their* dk-*random graphs. The Kolmogorov–Smirnov (KS) distances between the distributions of per-node values of a given property in the real networks, and the same distributions in their dk-random graphs for the following properties: degree (k), average degree of nearest neighbors (knn), clustering coefficient (c), number of common neighbors (comm.neigh), k-coreness (kcore), k-density (kdense), betweenness (bet), and shortest-path distance (path-len). Adapted from Ref. [219].*

very close to zero for higher values of d, indicating that these two networks can be well approximated with 2k- and 2.5k-random graphs. On the other hand, for BRAIN, the global properties exhibit slow or no convergence at all, so it is an outlier, that is, its properties can be captured with dk-random graphs with $d \geq 3$. Although many properties can be reproduced with 2.5k-graphs, we find that community structure is not preserved for any of these networks, regardless of the community detection algorithm.

3.3   *dk*-Series in multilayer networks

In this section, we discuss how dk-series generalize to multilayer networks in general and to multiplex networks in particular. The key idea behind the generalized dk-series

is the same as in the case of monolayer networks: a *dk*-series is a series of inclusive and convergent statistics based on the frequencies of degree-labeled subgraphs of increasing size in a given network.

3.3.1 Multilayer networks

Here, we consider the most general case of multilayer networks with the most detailed form of generalized *dk*-statistics.

The structure of a multilayer network is fully specified by the adjacency tensor $A_{\alpha i, \beta j}$, where indices $\alpha, \beta = 1, \ldots, L$ indicate layers, while $i, j = 1, \ldots, N$ indicate nodes: $A_{\alpha i, \beta j} = 1$ if node i at layer α is connected to node j at layer β, and $A_{\alpha i, \beta j} = 0$ otherwise [72] (see also Chapter 1). To simplify the notation, from now on, we assume that networks are undirected and have no loops—$A_{\alpha i, \beta j} = A_{\beta j, \alpha i}$, and $A_{\alpha i, \alpha i} = 0$, for any combination of α, β, i, and j—but the generalization to directed networks is straightforward.

The $0k$-statistics, that is, the number of edges in the network, and the $0k$-"distribution," that is, the average degree, are no longer scalar M and $\bar{k}$ as in the monolayer case, but the $L \times L$-matrices

$$\hat{M} \equiv M_{\alpha\beta} = \sum_{i \leq j} A_{\alpha i, \beta j}, \tag{3.9}$$

$$\hat{\bar{k}} \equiv \bar{k}_{\alpha\beta} = \frac{2 M_{\alpha\beta}}{N}, \tag{3.10}$$

specifying the number of interlayer edges between layers α and β if $\alpha \neq \beta$, or the number of intralayer edges at layer α if $\alpha = \beta$. Similarly, the degree of node i is no longer scalar but the $L \times L$-matrix

$$\hat{k}_i \equiv (k_i)_{\alpha\beta} = \sum_j A_{\alpha i, \beta j} \tag{3.11}$$

specifying the number of node's connections to other nodes in the same layer if $\alpha = \beta$, or to nodes in other layers if $\alpha \neq \beta$. If node i is not present at layer α, then $(k_i)_{\alpha\beta} = 0$. Node degrees are thus also matrices $\hat{k} = k_{\alpha\beta}$. A node that has this degree has $k_{\alpha\beta}$ connections from layer α to other nodes at layer β, which can be equal to α. We note that the degree matrices are not, in general, symmetric, even if the network is undirected.

The $1k$-statistics, that is, the number of nodes $N(\hat{k})$ with degree $\hat{k}$, and the $1k$-distribution are then

$$N(\hat{k}) = \sum_i \delta(\hat{k}_i, \hat{k}), \tag{3.12}$$

$$P(\hat{k}) = \frac{N(\hat{k})}{N}, \tag{3.13}$$

where δ stands for the Kronecker delta, and the distribution is properly normalized: $\sum_{\hat{k}} P(\hat{k}) = 1$. In contrast with the monolayer case, the degree distribution is no longer a univariate distribution but a multivariate joint distribution of L^2 variables. In particular, this distribution contains all the information on the correlation of degrees of the same node at different layers. As in the monolayer case, the 1k-distribution fully defines the 0k-distribution via

$$\bar{\hat{k}} = \sum_{\hat{k}} \hat{k} P(\hat{k}). \tag{3.14}$$

The 2k-statistics, that is, the number of links $N_{\alpha\alpha'}(\hat{k}, \hat{k}')$ between nodes of degrees $\hat{k}$ and $\hat{k}'$ at layers α and α', and the corresponding 2k-distribution are given by the matrices

$$\hat{N}(\hat{k}, \hat{k}') \equiv N_{\alpha\alpha'}(\hat{k}, \hat{k}') = \sum_{i \leq i'} A_{\alpha i, \alpha' i'} \delta(\hat{k}_i, \hat{k}) \delta(\hat{k}_{i'}, \hat{k}'), \tag{3.15}$$

$$\hat{P}(\hat{k}, \hat{k}') = \mu(\hat{k}, \hat{k}') \frac{\hat{N}(\hat{k}, \hat{k}')}{2\hat{M}}, \text{ where} \tag{3.16}$$

$$\mu(\hat{k}, \hat{k}') = \begin{cases} 2, & \text{if } \hat{k} = \hat{k}', \\ 1, & \text{otherwise,} \end{cases} \tag{3.17}$$

is the factor taking care of proper normalization $\sum_{\hat{k},\hat{k}'} P_{\alpha\alpha'}(\hat{k}, \hat{k}') = 1$ for any α, α'. Here and below, all vector, matrix, and tensor multiplication and divisions are element-wise, for example, $\left(\hat{N}/\hat{M}\right)_{\alpha\alpha'} = \hat{N}_{\alpha\alpha'}/\hat{M}_{\alpha\alpha'}$. Instead of a joint distribution of two variables $P(k, k')$ in the monolayer case, we deal with $L(L-1)$ joint distributions of $2L^2$ variables $P_{\alpha\alpha'}(\hat{k}, \hat{k}')$. These distributions contain strictly more information about degree correlations than the 1k-distribution does. In particular, in addition to capturing the degree correlations of the same node ($i = j$) across layers, they also encompass all the degree correlations of distinct connected nodes ($i \neq j$) across both intralayer ($\alpha = \alpha'$) and interlayer ($\alpha \neq \alpha'$) connections. The 2k-distributions define the 1k-distribution similarly to the monolayer case:

$$P(\hat{k}) = \frac{\bar{\hat{k}}}{\hat{k}} \sum_{\hat{k}'} \hat{P}(\hat{k}, \hat{k}'). \tag{3.18}$$

It is evident from the expressions above that the *dk*-series in multilayer networks are different from the *dk*-series in monolayer networks only in that the scalar number of edges and node degrees are replaced by the $L \times L$-matrices $M \to \hat{M}$ and $k \to \hat{k}$, while their *dk*-distributions form tensors of rank d whose indices are layers. If $d = 1$, this tensor is trivial: $\mathbf{P}(\hat{k}) \equiv P_\alpha(\hat{k}) = P(\hat{k})$ for any layer α; but, starting with $d = 2$, any two elements of these *dk*-distribution tensors can, in general, be different, specifying in the $d = 2$ case,

for instance, the degree correlations across pairs of different layers if $\alpha \neq \alpha'$, or within the same layer if $\alpha = \alpha'$. Each element of these dk-tensors is a joint distribution of d degrees, that is, of dL^2 variables.

All higher-order statistics and distributions are then defined exactly as in the monolayer case, albeit with these two modifications. For instance, the $3k$-distribution is defined by Eqs (3.4)–(3.5) and determines the $2k$-distributions via Eq. (3.8), except that scalar degrees in these equations are replaced by degree matrices, and the numbers of wedges and triangles have three indices specifying to which layers the three nodes forming these two subgraphs belong. As in the monolayer case, higher-d dk-distributions determine the degree correlations of nodes at distance $d - 1$, the frequencies of d-cliques, including clustering at $d = 3$, and so on.

3.3.2 Multiplex networks

In node-aligned multiplex networks, all nodes are present in all layers, and all interlayer connections are trivial: every node is connected only to all its copies in all other layers, so that interlayer connections form N disjoint L-cliques: $A_{\alpha i, \beta i} = 1$ for any combination of i and $\alpha \neq \beta$. The dk-series can therefore be excused from keeping track of statistics of interlayer connections, which somewhat simplifies the formalism in the previous section, as described below. This simplification boils down to per-layer projections of the most detailed dk-statistics discussed in the previous section.

Since in multiplex networks $k_{\alpha\beta} = L - 1$ for all nodes and all $\beta \neq \alpha$, all the off-diagonal components of degree matrices $\hat{k}$ are not informative and can thus be dropped, mapping degree $L \times L$-matrices to degree L-vectors composed of the diagonal elements of $\hat{k}$: $\hat{k} \mapsto \boldsymbol{k} \equiv diag(\hat{k})$. The α-component k_α of this vector $\boldsymbol{k}$ specifies the number of intralayer connections of a node at layer α. Similarly, the number-of-edges matrix $\hat{M}$ maps to the vector $\mathbf{M} \equiv diag(M)$, whose components M_α are the numbers of intralayer edges within layer α.

Similarly, it is convenient to project the dk-distribution tensors per layer, forming vectors of distributions $\mathbf{P}(\boldsymbol{k}, \boldsymbol{k}', \dots) \equiv P_\alpha(\boldsymbol{k}, \boldsymbol{k}', \dots)$, consisting of the diagonal elements of the full distribution tensor $P_{\alpha\alpha', \dots}(\boldsymbol{k}, \boldsymbol{k}', \dots)$, that is, $P_\alpha(\boldsymbol{k}, \boldsymbol{k}', \dots) = P_{\alpha\alpha, \dots}(\boldsymbol{k}, \boldsymbol{k}', \dots)$, thus keeping track only of intralayer correlations of degrees of different nodes. The correlations of the degrees of the same node at different layers are still contained in $P(\boldsymbol{k})$.

Given this simplified representation, the dk-statistics, distributions, and their relations are exactly as in the general multilayer case, except that all matrices and tensors are replaced by vectors whose components are layers. For $d = 0, 1, 2$, for instance, we have the following expressions:

$$\bar{\boldsymbol{k}} = \frac{2\mathbf{M}}{N}, \tag{3.19}$$

$$N(\boldsymbol{k}) = \sum_i \delta(\boldsymbol{k}_i, \boldsymbol{k}), \tag{3.20}$$

$$P(\boldsymbol{k}) = \frac{N(\boldsymbol{k})}{N}, \tag{3.21}$$

$$\bar{k} = \sum_{k} kP(k), \tag{3.22}$$

$$\mathbf{N}(k,k') \equiv N_{\alpha}(k,k') = \sum_{i \le i'} A_{\alpha i,\alpha i'} \delta(k_i,k)\delta(k_{i'},k'), \tag{3.23}$$

$$P(k,k') = \mu(k,k')\frac{\mathbf{N}(k,k')}{2\mathbf{M}}, \tag{3.24}$$

$$P(k) = \frac{\bar{k}}{k} \sum_{k'} \mathbf{P}(k,k'), \tag{3.25}$$

which all are lists of *L* per-layer standard monolayer expressions, except that degrees are vectors. Compared to general multilayer networks, the *dk*-distributions in multiplex networks with these simplifications are all *L*-vectors, for *d* > 1, whose components consists of joint distributions of *d* degrees, that is, of *dL* variables.

3.3.3   Application to real networks

The general methodology behind the application of *dk*-series generalized to multilayer networks is the same as in the monolayer case discussed in Section 3.2. Yet, one has to keep in mind that multilayer *dk*-statistics tend to be extremely sparse and thus extremely constraining, even in the multiplex-projected case. This is because, compared to monolayer networks, the *dk*-distributions in multilayer networks contain much more detailed information about degrees, which are no longer scalars but matrices or vectors, and about their correlations within subgraphs of different sizes. Therefore, it is usually convenient to consider summary statistics of these distributions and define graph randomization procedures based on those. These procedures may depend on a particular choice of real network, on its specifics, and on particular questions one is to answer about the network.

For instance, in Ref. [79], generalized *dk*-series were applied to the Internet at the AS level. This network is a multiplex network with two layers. One layer consists of directed customer-provider links, for which, in order to send traffic over them, customer ASs must pay provider ASs, while the other layer consists of undirected peer-to-peer links connecting mostly large Internet service provider ASs, which exchange traffic free of charge over these links, based on bilateral agreements. The specific question addressed for this network was how to generate synthetic random graphs of varying sizes that reproduce specific types of degree correlations that reflect realities of business relationships between ASs in the Internet. For instance, peer-to-peer links tend to exist only between large Internet providers of large AS degree. Large providers tend to have large number of customers, a handful of peers, and few or no providers. Small customer ASs have no customers, no peers, and a small number of providers, and so on.

To properly capture these correlations, three joint distributions were considered. One was the full 1*k*-distribution $P(k)$ specifying the correlations among the numbers of customer, provider, and peer connections that nodes have. Since the

customer–provider layer is a directed network, this distribution is a joint distribution of three variables: in- and out-degrees $k_{1,in}$ and $k_{1,out}$ in the customer–provider layer, and degrees k_2 in the peer-to-peer layer. The other two distributions were the $2k$-distributions of the total degrees $k = \sum_\alpha k_\alpha = k_{1,in} + k_{1,out} + k_2 = |k|_1$ in the two layers. These distributions are projections of the multiplex $2k$-distributions: $P(k,k') = \sum_{k,k'} \mathbf{P}(k,k')\delta(|k|_1,k)\delta(|k'|_1,k')$. After these distributions were determined from data from the real Internet data, the dk-series were used to generate synthetic graphs of any size reproducing all the degree correlations contained in these distributions by first computing the marginals of these distributions $P(k_{1,in})$, $P(k_{1,out})$, $P(k_2)$, and $P(k)$, and the three copulas [206] representing their correlations in their joint distributions $P(k) = P(k_{1,in}, k_{1,out}, k_2)$, $P_1(k,k')$, and $P_2(k,k')$. Joint degree sequences of varying lengths were then sampled from these copulas, and random graphs were constructed using stub-matching procedures. As expected, the degree correlations in these random graphs reproduced the degree correlations in the real Internet. Many other important structural properties of the Internet, including properties specific to the Internet, were reproduced by these $2k$-random graphs as well [79], corroborating the finding that the Internet is nearly $2k$-random with respect to many important properties [182, 219].

We conclude this section by reiterating that, when classifying, that is, determining "how random" a given multilayer or multiplex network is, the full matrix-degree-based dk-series provides a rich set of inclusive and convergent statistics, which contain a variety of summary statistics as different projections of the full dk-distributions. Any combination of these statistics constrained to their values observed in a given network defines a null random graph model, in which any structural property of the network can be tested on its typicality in the model. There seems to be no good-for-all-networks rule of what these projections are, as different multilayer networks may require different projections. Yet, the full joint $1k$-distribution should most likely be always considered, while per-layer projections of dk-distributions with $d > 1$ are likely to be good projection choices for many networks, especially multiplex ones.

3.4   Discussion and conclusion

Topological measures commonly used for characterizing the structure of complex networks are interdependent, but the relation and the extent of their mutual correlations are often unknown. For this reason, systematic classification of networks via standard topological measures is not feasible. Here, we show how the problem of interdependence can be overcome by finding the set of base properties that can be used to explain all other relevant topological features of a network's structure. We describe the methodology proposed in Ref. [219] and show that most topological properties, which are deemed relevant for dynamics and function of networks, can be reproduced by random graphs with fixed degree distribution, degree-degree correlations, an average clustering coefficient and a degree-dependent average clustering coefficient, as in a given real network.

There is no reason to expect that non-local properties, namely, mesoscopic and macroscopic properties, cannot be reproduced by random graphs with local constraints.

And, for some networks and some properties, this is true. Our numerical experiments show that global features of brain networks, for example, the shortest-path length and betweenness distributions, differ drastically between the original network and *dk*-random graphs. This suggests that brain network evolution was subjected to some global constraints, which is reflected in its structure. The human brain consists of two weakly connected parts, corresponding to two brain hemispheres, a feature that cannot be reproduced with *dk*-random graphs with small *d*. In general, *dk*-random graphs with fixed local properties can not reproduce the community structure of all complex networks studied in Ref. [219], that is, the cluster organization is not robust to *dk*-randomization.

On the other hand, INTERNET and PGP, and other networks considered in Ref. [219], are clearly *dk*-random, with $d \leq 2.5$. Our analysis shows that the most basic properties of these networks, including microscopic, mesoscopic, and macroscopic ones, are a consequence of several local *dk*-properties: degree distribution, degree–degree correlations, and global and degree-dependent average clustering coefficients. This implies that the evolution of these networks was dominated by local dynamical rules and that it can be explained to a certain extent by mechanisms that are responsible for the manifestation of specific *dk*-properties. There already exists a multitude of approaches [80, 151, 302, 10, 225, 34] proposing different mechanisms to explain the emergence of these local topological properties. Clearly, the features that cannot be reproduced by *dk*-random graphs require separate explanations, or maybe some other set of base properties and different systems of null models.

The most basic topological features considered in this work can be considered non-significant, that is, there exists a *dk*-property captured by the corresponding *dk*-random graphs. In general, to tell how statistically significant a particular feature is, one needs to compare this feature in a real network with the same feature in an ensemble of random graphs, that is, a null model. The choice of the null model is free, but one should be careful when choosing the null model, since the significance of a certain feature is strongly dependent on it. The *dk*-random graphs discussed in this chapter can be used for determining the right network topology generator. One should first check whether most topological features of networks can be reproduced in *dk*-random graphs with a low value of *d*. If this is the case, then one may not need any sophisticated mission-specific topology generators. The proposed extension of *dk*-series to multilayer graphs enables the use of similar procedures on wider classes of real networks.

There are certain drawbacks of our approach that have to be mentioned. First, we do not have a proof that the proposed *dk*-random graph generation algorithms for $d = 2.1$ and $d = 2.5$ sample graphs uniformly at random from the ensemble. Second, it is known that the random graph ensembles and edge-rewiring processes employed here suffer from problems such as degeneracy and hysteresis [97, 239, 133]. The ideal solution would be to calculate analytically the expected value of given property in an ensemble. For this, we need an analytical description of null models which is currently only available for soft $d = 0, 1, 2$-random graph models [281, 282, 59, 58]. Unfortunately, the null models for generation of random graphs ensembles with constraints $d > 2$ are still not feasible, and they appear to be beyond the reach in the near future. We also

lack algorithms for the generation of *dk*-random graphs for multiplex networks which would allow us to apply similar procedures and quantify the randomness of multilayer networks. Clearly, the solution of these problems will be of great importance for a full understanding of the relationship between the structure, function, and dynamics of real networks.

References

[1] Abrams, D. M., and Strogatz, S. H. (2004). Chimera states for coupled oscillators. *Physical Review Letters*, 93(17):174102.

[2] Coolen, A. C. C., De Martino, A., and Annibale, A. (2009). Constrained Markovian dynamics of random graphs. *Journal of Statistical Physics*, 136(6):1035–67.

[3] Ahn, Y.-Y., Bagrow, J. P., and Lehmann, S. (2010). Link communities reveal multiscale complexity in networks. *Nature*, 466(7307):761–4.

[4] Aizenman, J., Hutchison, M., and Jinjarak, Y. (2013). What is the risk of European sovereign debt defaults? Fiscal space, CDS spreads and market pricing of risk. *Journal of International Money and Finance*, 34(C):37–59.

[5] Albert, R., and Barabási, A.-L. (2002). Statistical mechanics of complex networks. *Reviews of Modern Physics*, 74(1): 47.

[6] Alderson, A. S., and Beckfield, J. (2004). Power and position in the world city system. *American Journal of Sociology*, 109(4):811–51.

[7] Almeida-Neto, M., Guimarães, P., Guimarães, P. R., Loyola, R. D., and Ulrich, W. (2008). A consistent metric for nestedness analysis in ecological systems: reconciling concept and measurement. *Oikos*, 117(8):1227–39.

[8] Alon, N., Caro, Y., Krasikov, I., and Roditty, Y. (1989). Combinatorial reconstruction problems. *Journal of Combinatorial Theory Series B*, 47:153–61.

[9] Alvarez-Hamelin, J., Dall'Asta, L., Barrat, A., and Vespignani, A. (2008). K-core decomposition of Internet graphs: Hierarchies, self-similarity and measurement biases. *Networks and Heterogeneous Media*, 3(2):371–93.

[10] Ángeles Serrano, M., and Boguñá, M. (2005). Tuning clustering in random networks with arbitrary degree distributions. *Physical Review E*, 72(3):36133.

[11] Annibale, A., Coolen, A. C. C., Fernandes, L., Fraternali, F., and Kleinjung, J. (2009). Tailored graph ensembles as proxies or null models for real networks I: tools for quantifying structure. *Journal of Physics A, Mathematical and General*, 42(48):485001.

[12] Anthonisse, J. M. (1971). The rush in a directed graph. *Stichting Mathematisch Centrum. Mathematische Besliskunde*, (BN 9/71):1–10.

[13] Antonopoulos, C. G., Srivastava, S., Pinto, S. E. S., and Baptista, M. S. (2015). Do brain networks evolve by maximizing their information flow capacity? *PLoS Computational Biology*, 11(8):e1004372.

[14] Asllani, M., Busiello, D. M., Carletti, T., Fanelli, D., and Planchon, G. (2014). Turing patterns in multiplex networks. *Physical Review E*, 90(4):042814.

[15] Barabási, A.-L. (2016). *Network Science*. Cambridge University Press, Cambridge.

[16] Barabási, A.-L., and Albert, R. R. (1999). Emergence of scaling in random networks. *Science*, 286(5439):509–12.

[17] Barrat, A., Barthélemy, M., and Vespignani, A. (2008). *Dynamical Processes on Complex Networks*. Cambridge University Press, Cambridge.

[18] Barrat, A., Cattuto, C., Tozzi, A., Vanhems, P., and Voirin, N. (2014). Measuring contact patterns with wearable sensors: Methods, data characteristics and applications to data-driven simulations of infectious diseases. *Clinical Microbiology and Infection*, 20(1):10–16.

[19] Barthélemy, M. (2011). Spatial networks. *Physics Reports*, 499(1–3):1–101.

[20] Barvinok, A., and Hartigan, J. A. (2013). The number of graphs and a random graph with a given degree sequence. *Random Structures and Algorithms*, 42(3):301–48.

[21] Bascompte, J., Jordano, P., Melián, C. J., and Olesen, J. M. (2003). The nested assembly of plant–animal mutualistic networks. *Proceedings of the National Academy of Sciences of the United States of America*, 100(16):9383–7.

[22] Bassler, K. E., Del Genio, C. I., Erdős, P. L., Miklós, I., and Toroczkai, Z. (2015). Exact sampling of graphs with prescribed degree correlations. *New Journal of Physics*, 17(8): 083052.

[23] Bastian, M., Heymann, S., and Jacomy, M. (2009). Gephi: An open source software for exploring and manipulating networks. In Adar, E., Hurst, M., Finin, T., Glance, N. S., Nicolov, N., and Tseng, B. L. (eds), *Proceedings of the Third International Conference on Weblogs and Social Media, ICWSM 2009, San Jose, California, USA, May 17–20*, pp. 361–2. AAAI Press, Palo Alto.

[24] Bastolla, U., Fortuna, M. A., Pascual-Garcia, A., Ferrera, A., Luque, B., and Bascompte, J. (2009). The architecture of mutualistic networks minimizes competition and increases biodiversity. *Nature*, 458(7241):1018–20.

[25] Batagelj, V., and Mrvar, A. (2004). Pajek: Analysis and visualization of large networks. In Jünger, M., and Mutzel, P. (eds), *Graph Drawing Software*, pp. 77–103. Springer, Berlin, Heidelberg.

[26] Battiston, F., Iacovacci, J., Nicosia, V., Bianconi, G., and Latora, V. (2016). Emergence of multiplex communities in collaboration networks. *PLOS ONE*, 11(1):e0147451.

[27] Battiston, F., Nicosia, V., and Latora, V. (2014). Structural measures for multiplex networks. *Physical Review E*, 89(3):32804.

[28] Bavelas, A. (1950). Communication patterns in task-oriented groups. *The Journal of the Acoustical Society of America*, 22(6):725–30.

[29] Beiró, M. G., Alvarez-Hamelin, J. I., and Busch, J. R. (2008). A low complexity visualization tool that helps to perform complex systems analysis. *New Journal of Physics*, 10(12):125003.

[30] Bernard, H. R. (2000). *Social Research Methods: Qualitative And Quantitative Approaches*. Sage Publications, Los Angelos.

[31] Bettencourt, L. M., Lobo, J., Helbing, D., Kühnert, C., and West, G. B. (2007). Growth, innovation, scaling, and the pace of life in cities. *Proceedings of the National Academy of Sciences of the United States of America*, 104(17):7301–6.

[32] Bianconi, G. (2009). Entropy of network ensembles. *Physical Review E*, 79(3):036114.

[33] Bianconi, G. (2013). Statistical mechanics of multiplex networks: Entropy and overlap. *Physical Review E*, 87(6):62806.

[34] Bianconi, G., Darst, R. K., Iacovacci, J., and Fortunato, S. (2014). Triadic closure as a basic generating mechanism of communities in complex networks. *Physical Review E*, 90(4):42806.

[35] Blondel, V. D., Decuyper, A., and Krings, G. (2015). A survey of results on mobile phone datasets analysis. *EPJ Data Science*, 4(1):10.

[36] Blonder, B., and Dornhaus, A. (2011). Time-ordered networks reveal limitations to information flow in ant colonies. *PLOS ONE*, 6(5):e20298.

[37] Boccaletti, S., Bianconi, G., Criado, R., del Genio, C., Gómez-Gardeñes, J., Romance, M., Sendiña-Nadal, I., Wang, Z., and Zanin, M. (2014). The structure and dynamics of multilayer networks. *Physics Reports*, 544(1):1–122.

[38] Boccaletti, S., Latora, V., Moreno, Y., and Chavez, M. (2006). Complex networks: Structure and dynamics. *Physics Reports*, 424(4):175–308.

[39] Boguñá, M., Pastor-Satorras, R., Díaz-Guilera, A., and Arenas, A. (2004). Models of social networks based on social distance attachment. *Physics Review E*, 70(5):56122.

[40] Bollobás, B. (1990). Almost every graph has reconstruction number three. *Journal of Graph Theory*, 14(1):1–4.

[41] Bollobás, B. (2006). *The Art of Mathematics: Coffee Time in Memphis*. Cambridge University Press, Cambridge.

[42] Bonacich, P. (1972). Factoring and weighting approaches to status scores and clique identification. *Journal of Mathematical Sociology*, 2(1):113–20.

[43] Bonacich, P. (1972). Technique for analyzing overlapping memberships. *Sociological Methodology*, 4:176–85.

[44] Bordenave, C., Feige, U., and Mossel, E. (2016). Shotgun assembly of random jigsaw puzzles. arXiv:1605.03086.

[45] Brandes, U. (2001). A faster algorithm for betweenness centrality. *Journal of Mathematical Sociology*, 25(2):163–77.

[46] Brin, S., and Page, L. (1998). The anatomy of a large-scale hypertextual web search engine. *Computer Networks and ISDN Systems*, 30(1):107–17.

[47] Bródka, P., and Kazienko, P. (2014). *Multi-Layered Social Networks*. Springer, Berlin.

[48] Buldyrev, S. V., Parshani, R., Paul, G., Stanley, H. E., and Havlin, S. (2010). Catastrophic cascade of failures in interdependent networks. *Nature*, 464(7291):1025–8.

[49] Burda, Z., Duda, J., Luck, J. M., and Waclaw, B. (2009). Localization of the maximal entropy random walk. *Physical Review Letters*, 102(16):160602.

[50] Burda, Z., Krzywicki, A., Martin, O. C., and Zagorski, M. (2011). Motifs emerge from function in model gene regulatory networks. *Proceedings of the National Academy of Sciences of the United States of America*, 108(42):17263–8.

[51] Bureau van Dijk. (2010). *Orbis*. Available at https://www.bvdinfo.com/en-us/our-products/company-information/international-products/orbis (accessed May 31, 2018).

[52] Bustos, S., Gomez, C., Hausmann, R., and Hidalgo, C. A. (2012). The dynamics of nestedness predicts the evolution of industrial ecosystems. *PLOS ONE*, 7(11):e49393.

[53] Cairncross, F., and Cairncross, F. C. (2001). *The Death of Distance: How the Communications Revolution Is Changing our Lives*. Harvard Business School, Boston.

[54] Caldarelli, G. (2007). *Scale-Free Networks: Complex Webs in Nature and Technology*. Oxford University Press, Oxford.

[55] Callaghan, T., Mucha, P. J., and Porter, M. A. (2007). Random walker ranking for NCAA Division I-A football. *American Mathematical Monthly*, 114(9):761–77.

[56] Cardillo, A., Gómez-Gardeñes, J., Zanin, M., Romance, M., Papo, D., Del Pozo, F., and Boccaletti, S. (2013). Emergence of network features from multiplexity. *Scientific Reports*, 3:1344.

[57] Cattuto, C., Schmitz, C., Baldassarri, A., Servedio, V. D. P., Loreto, V., Hotho, A., Grahl, M., and Stumme, G. (2007). Network properties of folksonomies. *AI Communications*, 20(4):245–62.

[58] Chatterjee, S., and Diaconis, P. (2013). Estimating and understanding exponential random graph models. *Annals of Statistics*, 41(5):2428–61.

[59] Chatterjee, S., Diaconis, P., and Sly, A. (2011). Random graphs with a given degree sequence. *Annals of Applied Probability*, 21(4):1400–35.

[60] Christen, P. (2012). *Data Matching: Concepts and Techniques for Record Linkage, Entity Resolution, and Duplicate Detection*. Data-Centric Systems and Applications. Springer-Verlag, Berlin, Heidelberg.

[61] Chung, F. R. K. (1999). Review of spectral graph theory. *ACM SIGACT News*, 30(2):14–16.

[62] Cohen, A. M., Hersh, W. R., Dubay, C., and Spackman, K. (2005). Using co-occurrence network structure to extract synonymous gene and protein names from MEDLINE abstracts. *BMC Bioinformatics*, 6(1):103.

[63] Colomer-de Simón, P., Serrano, M. Á., Beiró, M. G., Alvarez-Hamelin, J. I., and Boguñá, M. (2013). Deciphering the global organization of clustering in real complex networks. *Scientific Reports*, 3:2517.

[64] Content, J. and Frenken, K. (2016). Related variety and economic development: A literature review. *European Planning Studies*, 24(12):2097–12.

[65] Costanzo, M., Baryshnikova, A., Bellay, J., Kim, Y., Spear, E. D., Sevier, C. S., Ding, H., Koh, J. L. Y., Toufighi, K., Mostafavi, S., et al. (2010). The genetic landscape of a cell. *Science*, 327(5964):425–31.

[66] Cozzo, E., Kivelä, M., De Domenico, M., Solé-Ribalta, A., Arenas, A., Gómez, S., Porter, M. A., and Moreno, Y. (2015). Structure of triadic relations in multiplex networks. *New Journal of Physics*, 17(7):73029.

[67] Czabarka, É., Dutle, A., Erdős, P. L., and Miklós, I. (2015). On realizations of a joint degree matrix. *Discrete Applied Mathematics*, 181:283–8.

[68] da Fontoura Costa, L., and Travieso, G. (2007). Exploring complex networks through random walks. *Physical Review E*, 75(1):16102.

[69] Daqing, L., Kosmidis, K., Bunde, A., and Havlin, S. (2011). Dimension of spatially embedded networks. *Nature Physics*, 7(6):481–4.

[70] Darst, R. K., Granell, C., Arenas, A., Gómez, S., Saramäki, J., and Fortunato, S. (2016). Detection of timescales in evolving complex systems. arXiv:1604.00758.

[71] De Domenico, M., Porter, M. A., and Arenas, A. (2015). MuxViz: A tool for multilayer analysis and visualization of networks. *Journal of Complex Networks*, 3(2):159–76.

[72] De Domenico, M., Solé-Ribalta, A., Cozzo, E., Kivelä, M., Moreno, Y., Porter, A., Gómez, S., and Arenas, A. (2013). Mathematical formulation of multilayer networks. *Physical Review X*, 3(4):41022.

[73] De Domenico, M., Solé-Ribalta, A., Gómez, S., and Arenas, A. (2014). Navigability of interconnected networks under random failures. *Proceedings of the National Academy of Sciences of the United States of America*, 111(23):8351–6.

[74] De Domenico, M., Solé-Ribalta, A., Omodei, E., Gómez, S., and Arenas, A. (2015). Ranking in interconnected multilayer networks reveals versatile nodes. *Nature Communications*, 6:6868.

[75] de Simon, P. C. (2014). RandNetGen: A random network generator. Available at http://github.com/polcolomer/RandNetGen (accessed May 31, 2018).

[76] Del Genio, C. I., Kim, H., Toroczkai, Z., and Bassler, K. E. (2010). Efficient and exact sampling of simple graphs with given arbitrary degree sequence. *PLOS ONE*, 5(4):e10012.

[77] Dickison, M., Havlin, S., and Stanley, H. E. (2012). Epidemics on interconnected networks. *Physical Review E*, 85(6):66109.

[78] Diestel, R. (2000). *Graph Theory*. Springer-Verlag, Berlin, Heidelberg.

[79] Dimitropoulos, X., Krioukov, D., Riley, G., and Vahdat, A. (2009). Graph annotations in modeling complex network topologies. *ACM Transactions on Modeling and Computer Simulation*, 19(4):17.

[80] Dorogovtsev, S. N., Mendes, J., and Samukhin, A. (2001). Size-dependent degree distribution of a scale-free growing network. *Physics Review E*, 63(6):62101.

[81] Dunbar, R. I. M. (2011). Constraints on the evolution of social institutions and their implications for information flow. *Journal of Institutional Economics*, 7(Special Issue 3): 345–71.

[82] Eagle, N., Pentland, A. S., and Lazer, D. (2009). Inferring friendship network structure by using mobile phone data. *Proceedings of the National Academy of Sciences of the United States of America*, 106(36):15274–8.

[83] Eckmann, J.-P., Moses, E., and Sergi, D. (2004). Entropy of dialogues creates coherent structures in e-mail traffic. *Proceedings of the National Academy of Sciences of the United States of America*, 101(40):14333–7.

[84] Edmonds, P. (1997). Choosing the word most typical in context using a lexical co-occurrence network. In *Proceedings of the Eighth Conference on European Chapter for Computational Linguistics*, pp. 507–9. Association for Computational Linguistics, Stroudsburg.

[85] Eguiluz, V., Chialvo, D., Cecchi, G., Baliki, M., and Apkarian, A. V. (2005). Scale-free brain functional networks. *Physics Review Letters*, 94(1):18102.

[86] Ellis, D., Friedgut, E., Kindler, G., and Yehudayoff, A. (2016). Geometric stability via information theory. *Discrete Analysis*, 2016:10.

[87] ESPON (2010). *FOCI: Future Orientation for Cities (Report)*. Available at http://www.espon.eu/main/ (accessed May 31, 2018).

[88] Estrada, E. (2011). *The Structure of Complex Networks: Theory and Applications*. Oxford University Press, Oxford.

[89] Eurostat. (2008). *NACE Rev 2: Statistical Classification of Economic Activities in the European Community*. Available at http://ec.europa.eu/eurostat/documents/3859598/5902521/KS-RA-07-015-EN.PDF (accessed May 31, 2018).

[90] Expert, P., Evans, T. S., Blondel, V. D., and Lambiotte, R. (2011). Uncovering space-independent communities in spatial networks. *Proceedings of the National Academy of Sciences of the United States of America*, 108(19):7663–8.

[91] Fagerland, M. W. (2012). t-Tests, non-parametric tests, and large studies: A paradox of statistical practice? *BMC Medical Research Methodology*, 12(78):1–7.

[92] Feldman, R., and Sanger, J. (2006). *Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*. Cambridge University Press, New York.

[93] Ferrer i Cancho, R., and Ricard, V. S. (2001). The small world of human language. *Proceedings of the Royal Society B*, 268(1482):2261–5.

[94] Fienberg, S. E., Meyer, M. M., and Wasserman, S. S. (1985). Statistical analysis of multiple sociometric relations. *Journal of the American Statistical Association*, 80(389):51–67.

[95] Flaounas, I., Ali, O., Turchi, M., Snowsill, T., Nicart, F., De Bie, T., and Cristianini, (2011). NOAM: News outlets analysis and monitoring system. In *Proceedings of the 2011 ACM SIGMOD International Conference on Management of Data*, pp. 1275–8. ACM, New York.

[96] Fortunato, S. (2010). Community detection in graphs. *Physics Reports*, 486(3–5):75–174.

[97] Foster, D., Foster, J., Paczuski, M., and Grassberger, P. (2010). Communities, clustering phase transitions, and hysteresis: Pitfalls in constructing network ensembles. *Physics Review E*, 81(4):46115.

[98] Foster, D. V., Foster, J. G., Grassberger, P., and Paczuski, M. (2011). Clustering drives assortativity and community structure in ensembles of networks. *Physics Review E*, 84(6):66117.

[99] Fournet, J., and Barrat, A. (2014). Contact patterns among high school students. *PLOS ONE*, 9(9):e107878.

[100] Freeman, L. C. (1977). A set of measures of centrality based on betweenness. *Sociometry*, 40(1):35–41.

[101] Freeman, L. C. (1979). Centrality in social networks conceptual clarification. *Social Networks*, 1(3):215–39.

[102] Freilich, S., Kreimer, A., Meilijson, I., Gophna, U., Sharan, R., and Ruppin, E. (2010). The large-scale organization of the bacterial network of ecological co-occurrence interactions. *Nucleic Acids Research*, 38(12):3857–68.

[103] Frenken, K., Van Oort, F., and Verburg, T. (2007). Related variety, unrelated variety and regional economic growth. *Regional Studies*, 41(5):685–97.

[104] Gallos, L. K., and Argyrakis, P. (2003). Distribution of infected mass in disease spreading in scale-free networks. *Physica A*, 330(1):117–23.

[105] Gallotti, R., Porter, M. A., and Barthélemy, M. (2016). Lost in transportation: Information measures and cognitive limits in multilayer navigation. *Science Advances*, 2(2):e1500445.

[106] Gao, J., Buldyrev, S. V., Stanley, H. E., and Havlin, S. (2011). Networks formed from interdependent networks. *Nature Physics*, 8(1):40–8.

[107] Garas, A. (2016). *Interconnected Networks*. Springer, Cham.

[108] Garas, A., Argyrakis, P., Rozenblat, C., Tomassini, M., and Havlin, S. (2010). Worldwide spreading of economic crisis. *New Journal of Physics*, 12(11):113043.

[109] Gauvin, L., Panisson, A., Barrat, A., and Cattuto, C. (2015). Revealing latent factors of temporal networks for mesoscale intervention in epidemic spread. arXiv:1501.02758.

[110] Gauvin, L., Panisson, A., and Cattuto, C. (2014). Detecting the community structure and activity patterns of temporal networks: A non-negative tensor factorization approach. *PLOS ONE*, 9(1):e86028.

[111] Gauvin, L., Panisson, A., Cattuto, C., and Barrat, A. (2013). Activity clocks: Spreading dynamics on temporal networks of human contact. *Scientific Reports*, 3:3099.

[112] Gemmetto, V., Barrat, A., and Cattuto, C. (2014). Mitigation of infectious disease at school: Targeted class closure vs school closure. *BMC Infectious Diseases*, 14(1):1–10.

[113] Génois, M., Vestergaard, C. L., Cattuto, C., and Barrat, A. (2015). Compensating for population sampling in simulations of epidemic spread on temporal contact networks. *Nature Communications*, 6:8860.

[114] Génois, M., Vestergaard, C. L., Fournet, J., Panisson, A., Bonmarin, I., and Barrat, A. (2015). Data on face-to-face contacts in an office building suggest a low-cost vaccination strategy based on community linkers. *Network Science*, 3(03):326–47.

[115] Gfeller, D., and De Los Rios, P. (2007). Spectral coarse graining of complex networks. *Physical Review Letters*, 99(3):38701.

[116] Ghoshal, G., Zlatic, V., Caldarelli, G., and Newman, M. E. J. (2009). Random hypergraphs and their applications. *Physical Review E*, 79(6):66118.

[117] Gjoka, M., Kurant, M., and Markopoulou, A. (2013). 2.5K-graphs: From sampling to generation. In *INFOCOM, 2013 Proceedings IEEE*, pp. 1968–76. IEEE, Piscataway.

[118] Goldenberg, J., and Levy, M. (2009). Distance is not dead: Social interaction and geographical distance in the internet era. arXiv:0906.3202.

[119] Gómez, S., Díaz-Guilera, A., Gómez-Gardeñes, J., Pérez-Vicente, C. J., Moreno, Y., and Arenas, A. (2013). Diffusion dynamics on multiplex networks. *Physical Review Letters*, 110(2):028701.

[120] Gómez-Gardeñes, J., Zamora-López, G., Moreno, Y., and Arenas, A. (2010). From modular to centralized organization of synchronization in functional areas of the cat cerebral cortex. *PLOS ONE*, 5(8):e12313.

[121] González, M. C., Hidalgo, C. A., and Barabási, A.-L. (2008). Understanding individual human mobility patterns. *Nature*, 453(7196):779–82.

[122] Granell, C., Gómez, S., and Arenas, A. (2013). Dynamical interplay between awareness and epidemic spreading in multiplex networks. *Physical Review Letters*, 111(12):128701.

[123] Granovetter, M. S. (1973). The strength of weak ties. *American Journal of Sociology*, 78(6):1360–80.

[124] Guimerà, R., Sales-Pardo, M., and Amaral, L. A. (2007). Classes of complex networks defined by role-to-role connectivity profiles. *Nature Physics*, 3(1):63–9.

[125] Halu, A., Mondragón, R. J., Panzarasa, P., and Bianconi, G. (2013). Multiplex PageRank. *PLOS ONE*, 8(10):e78293.

[126] Harary, F. (1964). On the reconstruction of a graph from a collection of subgraphs. In *Theory of Graphs and its Applications: Proceedings of the Symposium Held in Smolenice in June 1963)*, pp. 47–52. Publishing House of the Czechoslovak Academy of Sciences, Prague.

[127] Herfindahl, O. C. (1950). Concentration in the US steel industry. PhD thesis, Columbia University, New York.

[128] Hicks, J., Traag, V. A., and Reinanda, R. (2015). Turning digitised newspapers into networks of political elites. *Asian Journal of Social Science*, 43(5):567–87.

[129] Hidalgo, C. A., and Hausmann, R. (2009). The building blocks of economic complexity. *Proceedings of the National Academy of Sciences of the United States of America*, 106(26):10570–5.

[130] Hirschman, A. O. (1980). *National Power and the Structure of Foreign Trade*, volume 105. University of California Press, Berkeley.

[131] Hizanidis, J., Kouvaris, N. E., Gorka, Z.-L., Díaz-Guilera, A., and Antonopoulos, C. G. (2016). Chimera-like states in modular neural networks. *Scientific Reports*, 6: 19845.

[132] Hoover, E. M. (1948). *The Location of Economic Activity*. McGraw-Hill Book Company, Inc., London.

[133] Horvát, S., Czabarka, É., and Toroczkai, Z. (2015). Reducing degeneracy in maximum entropy models of networks. *Physics Review Letters*, 114(15–17):158701.

[134] Isard, W. (1956). *Localization and Space Economy: A General Theory Relating to Industrial Location, Market Areas, Land Use, Trade And Urban Structure*. MIT Press, Cambridge, MA.

[135] Isella, L., Stehlé, J., Barrat, A., Cattuto, C., Pinton, J.-F., and den Broeck, W. (2011). What's in a crowd? Analysis of face-to-face behavioral networks. *Journal of Theoretical Biology*, 271(1):166–80.

[136] Jackson, M. O. (2010). *Social and Economic Networks*. Princeton University Press, Princeton.

[137] Jamakovic, A., Mahadevan, P., Vahdat, A., Boguñá, M., and Krioukov, D. (2009). How small are building blocks of complex networks. arXiv:0908.1143.

[138] Jo, H.-H., Baek, S. K., and Moon, H.-T. (2006). Immunization dynamics on a two-layer network model. *Physica A*, 361(2):534–42.

[139] Jo, H.-H., Karsai, M., Karikoski, J., and Kaski, K. (2012). Spatiotemporal correlations of handset-based service usages. *EPJ Data Science*, 1(1):10.

[140] Jo, H.-H., Pan, R. K., Perotti, J. I., and Kaski, K. (2013). Contextual analysis framework for bursty dynamics. *Physical Review E*, 87:062131.

[141] Jo, H.-H., Saramäki, J., Dunbar, R. I. M., and Kaski, K. (2014). Spatial patterns of close relationships across the lifespan. *Scientific Reports*, 4:6988.

[142] Juršič, M., Sluban, B., Cestnik, B., Grčar, M., and Lavrač, N. (2012). Bridging concept identification for constructing information networks from text documents. In Berthold M. R. (ed.), *Bisociative Knowledge Discovery*, pp. 66–90. Springer, Berlin, Heidelberg.

[143] Karsai, M., Kivelä, M., Pan, R. K., Kaski, K., Kertész, J., Barabási, A.-L., and Saramäki, J. (2011). Small but slow world: How network topology and burstiness slow down spreading. *Physical Review E*, 83:025102.

[144] Katz, L. (1953). A new status index derived from sociometric analysis. *Psychometrika*, 18(1):39–43.

[145] Kelly, P. J. (1957). A congruence theorem for trees. *Pacific Journal of Mathematics*, 7(1):961–8.

[146] Kermack, W. O., and McKendrick, A. G. (1927). A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society A*, 115(772):700–21.

[147] Kim, H., Del Genio, C. I., Bassler, K. E., and Toroczkai, Z. (2012). Constructing and sampling directed graphs with given degree sequences. *New Journal of Physics*, 14(2): 23012.

[148] Kim, H., Toroczkai, Z., Erdős, P. L., Miklós, I., and Székely, L. A. (2009). Degree-based graph construction. *Journal of Physics A*, 42(39):392001.

[149] Kivelä, M., Arenas, A., Barthélemy, M., Gleeson, J. P., Moreno, Y., and Porter, M. A. (2014). Multilayer networks. *Journal of Complex Networks*, 2(3):203–71.

[150] Kleinberg, J. M. (1999). Authoritative sources in a hyperlinked environment. *Journal of the ACM*, 46(5):604–32.

[151] Klemm, K., and Eguiluz, V. (2002). Highly clustered scale-free networks. *Physics Review E*, 65(3):36123.

[152] Klimek, P., Diakonova, M., Eguiluz, V., Miguel, M. S., and Thurner, S. (2016). Dynamical origins of the community structure of multi-layer societies. *New Journal of Physics*, 18(8):1–8.

[153] Kok, S., and Domingos, P. M. (2008). Extracting semantic networks from text via relational clustering. In Daelemans, W., Goethals, B., and Morik, K. (eds), *Machine Learning and Knowledge Discovery in Databases*, pp. 624–39. Springer, Berlin, Heidelberg.

[154] Kolda, T. G., and Bader, B. W. (2009). Tensor decompositions and applications. *SIAM Review*, 51(3):455–500.

[155] Kosmidis, K., Havlin, S., and Bunde, A. (2008). Structural properties of spatially embedded networks. *Europhysics Letters*, 82(4):48005.

[156] Kossinets, G., and Watts, D. J. (2006). Empirical analysis of an evolving social network. *Science*, 311(5757):88–90.

[157] Kouvaris, N. E., Hata, S., and Diaz-Guilera, A. (2015). Pattern formation in multiplex networks. *Scientific Reports*, 5:10840.

[158] Kralj Novak, P., Grčar, M., and Mozetič, I. (2015). *Analysis of Financial News with NewsStream*. Technical Report IJS-DP-11892. Jožef Stefan Institute, Ljubljana.

[159] Krings, G., Calabrese, F., Ratti, C., and Blondel, V. D. (2009). Urban gravity: A model for inter-city telecommunication flows. *Journal of Statistical Mechanics*, 2009(7):L07003.

[160] Krugman, P. (1993). First nature, second nature, and metropolitan location. *Journal of Regional Science*, 33(2):129–44.

[161] Krzywinski, M., Birol, I., Jones, S. J. M., and Marra, M. A. (2012). Hive plots: Rational approach to visualizing networks. *Briefings in Bioinformatics*, 13(5):627–44.

[162] Kumpula, J. M., Onnela, J. P., Saramäki, J., Kaski, K., and Kertész, J. (2007). Emergence of communities in weighted networks. *Physical Review Letters*, 99(22):228701.

[163] Kuperman, M., and Abramson, G. (2001). Small world effect in an epidemiological model. *Physical Review Letters*, 86(13):2909.

[164] Kuramoto, Y., and Battogtokh, D. (2002). Coexistence of coherence and incoherence in nonlocally coupled phase oscillators. *Nonlinear Phenomena in Complex Systems*, 5(4):380–5.

[165] Kwak, H., Lee, C., Park, H., and Moon, S. (2010). What is Twitter, a social network or a news media? In *Proceedings of the 19th International World Wide Web Conference*, pp. 591–600. ACM, New York.

[166] Lambiotte, R., Blondel, V., Dekerchove, C., Huens, E., Prieur, C., Smoreda, Z., and Vandooren, P. (2008). Geographical dispersal of mobile communication networks. *Physica A*, 387(21):5317–25.

[167] Lambiotte, R., Delvenne, J.-C., and Barahona, M. (2014). Random walks, Markov processes and the multiscale modular organization of complex networks. *IEEE Transactions on Network Science and Engineering*, 1(2):76–90.

[168] Lazer, D., Pentland, A., Adamic, L. A., Aral, S., Barabasi, A.-L., Brewer, D., Christakis, N., Contractor, N., Fowler, J., Gutmann, M., et al. (2009). Computational social science. *Science*, 323(5915):721–3.

[169] Leban, G., Fortuna, B., Brank, J., and Grobelnik, M. (2014). Event registry: Learning about world events from news. In *Proceedings of the 23rd International World Wide Web Conference*, pp. 107–10. ACM, New York.

[170] Lee, K.-M., Kim, J. Y., Cho, W.-K., Goh, K.-I., and Kim, I. M. (2012). Correlated multiplexity and connectivity of multiplex random networks. *New Journal of Physics*, 14(3):33027.

[171] Lee, S. H., Kim, P.-J., and Jeong, H. (2006). Statistical properties of sampled networks. *Physics Review E*, 73(1):016102.

[172] Leetaru, K. H. (2011). Culturomics 2.0: Forecasting large-scale human behavior using global news media tone in time and space. *First Monday*, 16(9): http://www.firstmonday.dk/ojs/index.php/fm/article/view/3663/3040.

[173] Lengyel, B., Varga, A., Sagvari, B., Jakobi, A., and Kertesz, J. (2015). Geographies of an online social network. *PLOS ONE*, 10(9): e0137248.

[174] Lewis, K., Kaufman, J., Gonzalez, M., Wimmer, A., and Christakis, N. (2008). Tastes, ties, and time: A new social network dataset using Facebook.com. *Social Networks*, 30(4):330–42.

[175] Li, W., Bashan, A., Buldyrev, S. V., Stanley, H. E., and Havlin, S. (2012). Cascading failures in interdependent lattice networks: The critical role of the length of dependency links. *Physics Review Letters*, 108(22):228702.

[176] Lima, A., De Domenico, M., Pejovic, V., and Musolesi, M. (2013). Exploiting cellular data for disease containment and information campaigns strategies in country-wide epidemics. arXiv:1306.4534.

[177] Liu, B. (2015). *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions*. Cambridge University Press, Cambridge.

[178] Liu, H., and Cong, J. (2013). Language clustering with word co-occurrence networks based on parallel texts. *Chinese Science Bulletin*, 58(10):1139–44.

[179] Lloyd, L., Kechagias, D., and Skiena, S. (2005). Lydia: A system for large-scale news analysis. In Consens M., and Navarro G. (eds), *String Processing and Information Retrieval*, pp. 161–6. Springer, Berlin, Heidelberg.

[180] Lovász, L. (1993). Random walks on graphs: A survey. *Combinatorics, Paul Erdos is Eighty*, 2(1):1–46.

[181] Machens, A., Gesualdo, F., Rizzo, C., Tozzi, A. E., Barrat, A., and Cattuto, C. (2013). An infectious disease model on empirical networks of human contact: Bridging the gap between dynamic network data and contact matrices. *BMC Infectious Diseases*, 13(1):185.

[182] Mahadevan, P., Krioukov, D., Fall, K., and Vahdat, A. (2006). Systematic topology analysis and generation using degree correlations. *ACM SIGCOMM Computer Communication Review*, 36(4):135–46.

[183] Mahadevan, P., Krioukov, D., Fomenkov, M., Huffaker, B., Dimitropoulos, X., Claffy, K., and Vahdat, A. (2006b). The Internet AS-level topology: Three data sources and one definitive metric. *ACM SIGCOMM Computer Communication Review*, 36(1):17–26.

[184] Mane, K. K., and Börner, K. (2004). Mapping topics and topic bursts in PNAS. *Proceedings of the National Academy of Sciences of the United States of America*, 101(Suppl. 1):5287–90.

[185] Martinsson, A. (2017). A linear threshold for uniqueness of solutions to random jigsaw puzzles. arXiv:1701.04813.

[186] Maslov, S., Sneppen, K., and Alon, U. (2003). Correlation profiles and motifs in complex networks. In Bornholdt, S., and Schuster, H. G. (eds), Handbook of Graphs and Networks: From the Genome to the Internet. Wiley-VCH, Berlin.

[187] Maslov, S., Sneppen, K., and Zaliznyak, A. (2004). Detection of topological patterns in complex networks: Correlation profile of the internet. *Physica A*, 333(1–4):529–40.

[188] Mastrandrea, R., and Barrat, A. (2016). How to estimate epidemic risk from incomplete contact diaries data? *PLoS Computational Biology*, 12(6):1–19.

[189] Mastrandrea, R., Fournet, J., and Barrat, A. (2015). Contact patterns in a high school: A comparison between data collected using wearable sensors, contact diaries and friendship surveys. *PLOS ONE*, 10(9):1–26.

[190] McPherson, M., Smith-Lovin, L., and Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 27(1):415–44.

[191] Mercer. (2012). Quality of Living City Ranking. Mercer, New York.

[192] Miljkovic, D., Stare, T., Mozetič, I., Podpečan, V., Petek, M., Witek, K., Dermastia, M., Lavrač, N., and Gruden, K. (2012). Signalling network construction for modelling plant defence response. *PLOS ONE*, 7(12):e0051822.

[193] Miller, G. A. (1995). WordNet: A lexical database for English. *Communications of the ACM*, 38(11):39–41.

[194] Miller, M., Sathi, C., Wiesenthal, D., Leskovec, J., and Potts, C. (2011). Sentiment flow through hyperlink networks. In *Proceedings of the Fifth International Conference on Weblogs and Social Media*, pp. 550–3. AAAI Press, Palo Alto.

[195] Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., and Alon, U. (2002). Network motifs: Simple building blocks of complex networks. *Science*, 298(5594):824–7.

[196] OECD/Eurostat. (2005). *Oslo Manual: Guidelines for Collecting and Interpreting Innovation Data*. OECD Publishing, Paris.

[197] Mossel, E., and Ross, N. (2015). Shotgun assembly of labeled graphs. arXiv:1504.07682.

[198] Motter, A. E., and Toroczkai, Z. (2007). Introduction: Optimization in networks. *Chaos*, 17(2):26101.

[199] Mozetič, I., Grčar, M., and Smailovic, J. (2016). Multilingual Twitter sentiment classification: The role of human annotators. *PLOS ONE*, 11(5):e0155036.

[200] Mucha, P. J., Richardson, T., Macon, K., Porter, M. A., and Onnela, J.-P. (2010). Community structure in time-dependent, multiscale, and multiplex networks. *Science*, 328(5980):876–8.

[201] Murase, Y., Jo, H.-H., Török, J., Kertész, J., and Kaski, K. (2015). Modeling the role of relationship fading and breakup in social network formation. *PLOS ONE*, 10(7):e0133005.

[202] Murase, Y., Török, J., Jo, H.-H., Kaski, K., and Kertész, J. (2014). Multilayer weighted social network model. *Physical Review E*, 90(5):052810.

[203] Murase, Y., Uchitane, T., and Ito, N. (2014). A tool for parameter-space explorations. *Physics Procedia*, 57:73–6.

[204] Myers, S. A., Sharma, A., Gupta, P., and Lin, J. (2014). Information network or social network? The structure of the Twitter follow graph. In *Proceedings of the 23rd International World Wide Web Conference*, pp. 493–8. ACM, New York.

[205] Nakao, H., and Mikhailov, A. S. (2010). Turing patterns in network-organized activator-inhibitor systems. *Nature Physics*, 6(7):544–50.

[206] Nelsen, R. B. (2007). *An Introduction to Copulas*. Springer, New York.

[207] Nenadov, R., Pfister, P., and Steger, A. (2016). Unique reconstruction threshold for random jigsaw puzzles. arXiv:1605.03043.

[208] Newman, M. E. (2002). Assortative mixing in networks. *Physical Review Letters*, 89(20):1–5.

[209] Newman, M. E. (2003). The structure and function of complex networks. *SIAM Review*, 45(2):167–256.

[210] Newman, M. E. (2010). *Networks: An Introduction*. Oxford University Press, New York.

[211] Newman, M. E. J. (2005). A measure of betweenness centrality based on random walks. *Social Networks*, 27(1):39–54.

[212] Newman, M. E. J. (2009). Random graphs with clustering. *Physics Review Letters*, 103(5): 1–4.

[213] Nicosia, V., Bianconi, G., Latora, V., and Barthelemy, M. (2013). Growing multiplex networks. *Physical Review Letters*, 111(5):58701.

[214] Noh, J. D., and Rieger, H. (2004). Random walks on complex networks. *Physical Review Letters*, 92(11):118701.

[215] Ohlin, B. G. (1933). *Interregional and International Trade*. Harvard University Press, Cambridge, MA.

[216] Onnela, J.-P., Arbesman, S., González, M. C., Barabási, A.-L., and Christakis, N. A. (2011). Geographic constraints on social network groups. *PLOS ONE*, 6(4):e16939.

[217] Onnela, J.-P., Saramäki, J., Hyvönen, J., Szabó, G., de Menezes, M., Kaski, K., Barabási, A.-L., and Kertész, J. (2007). Analysis of a large-scale weighted network of one-to-one human communication. *New Journal of Physics*, 9(6):179.

[218] Onnela, J.-P., Saramäki, J., Hyvönen, J., Szabó, G., Lazer, D., Kaski, K., Kertész, J., and Barabási, A.-L. (2007). Structure and tie strengths in mobile communication networks. *Proceedings of the National Academy of Sciences of the United States of America*, 104(18): 7332–6.

[219] Orsini, C., Dankulov, M. M., Colomer-de Simón, P., Jamakovic, A., Mahadevan, P., Vahdat, A., Bassler, K. E., Toroczkai, Z., Boguñá, M., Caldarelli, G., et al. (2015). Quantifying randomness in real networks. *Nature Communications*, 6:8627.

[220] Özgür, A., Cetin, B., and Bingol, H. (2008). Co-occurrence network of Reuters news. *International Journal of Modern Physics C*, 19(05):689–702.

[221] Padgett, J. F., and Ansell, C. K. (1993). Robust action and the rise of the Medici, 1400–1434. *American Journal of Sociology*, 98(6):1259–319.

[222] Palchykov, V., Kaski, K., Kertész, J., Barabási, A.-L., and Dunbar, R. I. M. (2012). Sex differences in intimate relationships. *Scientific Reports*, 2:370.

[223] Pan, J., and Singleton, K. J. (2008). Default and recovery implicit in the term structure of sovereign CDS spreads. *The Journal of Finance*, 63(5):2345–84.

[224] Panaggio, M. J., and Abrams, D. M. (2015). Chimera states: Coexistence of coherence and incoherence in networks of coupled oscillators. *Nonlinearity*, 28(3):R67.

[225] Papadopoulos, F., Kitsak, M., Serrano, M. Á., Boguñá, M., and Krioukov, D. (2012). Popularity versus similarity in growing networks. *Nature*, 489(7417):537–40.

[226] Pearson, K. (1895). Note on regression and inheritance in the case of two parents. *Proceedings of the Royal Society of London*, 58(347–52):240–2.

[227] Pebody, L. (2004). The reconstructibility of finite abelian groups. *Combinatorics, Probability and Computing*, 13(6):867–92.

[228] Pebody, L. (2007). Reconstructing odd necklaces. *Combinatorics, Probability and Computing*, 16(4):503–14.

[229] Pebody, L., Radcliffe, A. J., and Scott, A. D. (2003). Finite subsets of the plane are 18-reconstructible. *SIAM Journal on Discrete Mathematics*, 16(2):272–5.

[230] Pfitzner, R., Scholtes, I., Garas, A., Tessone, C. J., and Schweitzer, F. (2013). Betweenness preference: Quantifying correlations in the topological dynamics of temporal networks. *Physical Review Letters*, 110(19):198701.

[231] Piškorec, M., Sluban, B., and Šmuc, T. (2015). MultiNets: Web-based multilayer network visualization. In Bifet, A., May, M., Zadrozny, B., Gavalda, R., Pedreschi, D., Bonchi, F.,

Cardoso, J., and Spiliopoulou, M. (eds), *Machine Learning and Knowledge Discovery in Databases*, pp. 298–302. Springer, Cham.

[232] Popović, M., Štefančič, H., Sluban, B., Novak, P. K., Grčar, M., Mozetič, I., Puliga, M., and Zlatič, V. (2014). Extraction of temporal networks from term co-occurrences in online textual sources. *PLOS ONE*, 9(12):e99515.

[233] United Nations Global Pulse. (2013). *Mobile Phone Network Data for Development*. Available at http://www.slideshare.net/unglobalpulse/mobile-data-for-development-primer-october-2013 (accessed May 31, 2017).

[234] Pumain, D., Paulus, F., Vacchiani-Marcuzzo, C., and Lobo, J. (2006). An evolutionary theory for interpreting urban scaling laws. *Cybergeo*, Systèmes, Modélisation, Géostatistiques (343): http://journals.openedition.org/cybergeo/2519.

[235] Radcliffe, A. J., and Scott, A. D. (1998). Reconstructing subsets of Z_n. *Journal of Combinatorial Theory, Series A*, 83(2):169–87.

[236] Radicchi, F., and Arenas, A. (2013). Abrupt transition in the structural formation of interconnected networks. *Nature Physics*, 9(11):717–20.

[237] Raghavan, V., Bollmann, P., and Jung, G. S. (1989). A critical investigation of recall and precision as measures of retrieval system performance. *ACM Transactions on Information Systems*, 7(3):205–29.

[238] Ranco, G., Aleksovski, D., Caldarelli, G., Grčar, M., and Mozetič, I. (2015). The effects of Twitter sentiment on stock price returns. *PLOS ONE*, 10(9):e0138441.

[239] Roberts, E. S., and Coolen, A. C. C. (2014). Random graph ensembles with many short loops. *ESAIM: Proceedings and Surveys*, 47(2014):97–115.

[240] Rohr, R. P., Saavedra, S., and Bascompte, J. (2014). On the structural stability of mutualistic systems. *Science*, 345(6195):1253497.

[241] Rossi, L., and Magnani, M. (2015). Towards effective visual analytics on multiplex and multilayer networks. *Chaos, Solitons & Fractals*, 72:68–76.

[242] Rosvall, M., and Bergstrom, C. T. (2007). An information-theoretic framework for resolving community structure in complex networks. *Proceedings of the National Academy of Sciences of the United States of America*, 104(18):7327–31.

[243] Rosvall, M., Esquivel, A. V. Lancichinetti, A., West, J. D., and Lambiotte, R. (2014). Memory in network flows and its effects on spreading dynamics and community detection. *Nature Communications*, 5:4630.

[244] Lennert, M., Van Hamme, G., Patris, C., Smetkowski, M., Płoszaj, A., Gorzelak, G., Kozak, M., Olechnicka, A., Wojnar, K., Hryniewicz, J., et al. (eds), *FOCI: Future Orientation for Cities*. ESPON Final report. Available at http://hal.archives-ouvertes.fr/hal-00734406/document (accessed June 4, 2018).

[245] Rozenblat, C. (2015). Inter-cities' multinational firm networks and gravitation model. *Annals of the Association of Economic Geographers*, 61(3):219–37.

[246] Rozenblat, C., Zaidi, F., and Bellwald, A. (2017). The multipolar regionalization of cities in multinational firms' networks. *Global Networks*, 17(2):171–94.

[247] Rozenfeld, H. D., Kirk, J. E., Bollt, E. M., and Ben-Avraham, D. (2005). Statistics of cycles: How loopy is your network? *Journal of Physics A*, 38(21):4589.

[248] Saavedra, S., Stouffer, D. B., Uzzi, B., and Bascompte, J. (2011). Strong contributors to network persistence are the most vulnerable to extinction. *Nature*, 478(7368):233–5.

[249] Sabidussi, G. (1966). The centrality index of a graph. *Psychometrika*, 31(4):581–603.

[250] Saito, K., and Yamada, T. (2006). Extracting communities from complex networks by the *k*-dense method. In Tsumoto, S., Clifton, C. W., Zhong, N., Wu, X., Liu, J., Wah, B. W., and Cheung, Y.-M. (eds), *Sixth IEEE International Conference on Data Mining: Workshops (ICDMW'06)*, pp. 300–4. IEEE, Los Alamitos.

[251] Salton, G. (1989). *Automatic Text Processing: The Transformation, Analysis, and Retrieval of Information by Computer*. Addison-Wesley, Boston.

[252] Samukhin, A. N., Dorogovtsev, S. N., and Mendes, J. F. F. (2008). Laplacian spectra of, and random walks on, complex networks: Are scale-free architectures really important? *Physical Review E*, 77(3):36115.

[253] Sánchez-García, R. J., Cozzo, E., and Moreno, Y. (2014). Dimensionality reduction and spectral properties of multilayer networks. *Physical Review E*, 89(5):52815.

[254] Saramäki, J., Leicht, E. A., López, E., Roberts, S. G. B., Reed-Tsochas, F., and Dunbar, R. I. M. (2014). Persistence of social signatures in human communication. *Proceedings of the National Academy of Sciences of the United States of America*, 111(3):942–7.

[255] Sassen, S. (2001). *The Global City: New York, London, Tokyo*. Princeton University Press, Princeton.

[256] Scholl, T., Garas, A., and Schweitzer, F. (2015). The spatial component of R&D networks. arXiv:1509.08291.

[257] Scholtes, I., Wider, N., and Garas, A. (2015). Higher-order aggregate networks in the analysis of temporal networks: Path structures and centralities. *European Physical Journal B*, 89(3):1–15.

[258] Scholtes, I., Wider, N., Pfitzner, R., Garas, A., Tessone, C. J., and Schweitzer, F. (2014). Causality-driven slow-down and speed-up of diffusion in non-Markovian temporal networks. *Nature Communications*, 5:5024.

[259] Secrier, M., Pavlopoulos, G. A., Aerts, J., and Schneider, R. (2012). Arena3D: Visualizing time-driven phenotypic differences in biological systems. *BMC Bioinformatics*, 13(1):45.

[260] Seidman, S. B. (1983). Network structure and minimum degree. *Social Networks*, 5(3): 269–87.

[261] Seyed-Allaei, H., Bianconi, G., and Marsili, M. (2006). Scale-free networks with an exponent less than two. *Physical Review E*, 73(4):46113.

[262] Shalgi, R., Lieber, D., Oren, M., and Pilpel, Y. (2007). Global and local architecture of the mammalian microRNA–transcription factor regulatory network. *PLoS Computational Biology*, 3(7):e131.

[263] Shanahan, M. (2010). Metastable chimera states in community-structured oscillator networks. *Chaos*, 20(1):13108.

[264] Shang, Y., Li, Y., Lin, H., and Yang, Z. (2011). Enhancing biomedical text summarization using semantic relation extraction. *PLOS ONE*, 6(8):1–10.

[265] Shaw, M. E. (1954). Group structure and the behavior of individuals in small groups. *Journal of Psychology*, 38(1):139–49.

[266] Simpson, E. H. (1949). Measurement of diversity. *Nature*, 163(4148):688.

[267] Sinatra, R., Gómez-Gardeñes, J., Lambiotte, R., Nicosia, V., and Latora, V. (2011). Maximal-entropy random walks in complex networks with limited information. *Physical Review E*, 83(3):30103.

[268] Skiena, S. S. (1998). *The Algorithm Design Manual*. Springer-Verlag New York, New York.

[269] Sluban, B., Grčar, M., and Mozetič, I. (2016). Temporal multi-layer network construction from major news events. In Cherifi H., Gonçalves B., Menezes R., Sinatra R. (eds), *Complex Networks VII: Proceedings of the 7th Workshop on Complex Networks CompleNet*, pp. 29–41. Springer, Cham.

[270] Sluban, B., Smailović, J., Battiston, S., and Mozetič, I. (2015). Sentiment leaning of influential communities in social networks. *Computational Social Networks*, 2(1):1–21.

[271] Sluban, B., Smailović, J., and Mozetič, I. (2016). Understanding financial news with multi-layer network analysis. In Battiston, S., De Pellegrini, F., Caldarelli, G., and Merelli, E. (eds), *Proceedings of ECCS 2014*, pp 193–207. Springer, Cham.

[272] Smailović, J., Grčar, M., Lavrač, N., and Žnidaršič, M. (2014). Stream-based active learning for sentiment analysis in the financial domain. *Information Sciences*, 285:181–203.

[273] Smailović, J., Kranjc, J., Grčar, M., Žnidaršič, M., and Mozetič, I. (2015). Monitoring the Twitter sentiment during the Bulgarian elections. In Gaussier, E., Cao, L., Gallinari, P., Kwok, J., Pasi, G., and Zaiane, O. (eds), *Proceedings of the 2015 IEEE International Conference on Data Science and Advanced Analytics, DSAA 2015*, pp. 1–10. IEEE, Piscataway.

[274] Solá, L., Romance, M., Criado, R., Flores, J., del Amo, A. G., and Boccaletti, S. (2013). Eigenvector centrality of nodes in multiplex networks. *Chaos*, 23(3):33131.

[275] Solé-Ribalta, A., De Domenico, M., Gómez, S., and Arenas, A. (2014). Centrality rankings in multiplex networks. In *Proceedings of the 2014 ACM Conference on Web Science*, pp. 149–55. ACM, New York.

[276] Solé-Ribalta, A., De Domenico, M., Gómez, S., and Arenas, A. (2016). Random walk centrality in interconnected multilayer networks. *Physica D*, 323–4:73–9.

[277] Solé-Ribalta, A., De Domenico, M., Kouvaris, N. E., Díaz-Guilera, A., Gómez, S., and Arenas, A. (2013). Spectral properties of the Laplacian of multiplex networks. *Physical Review E*, 88(3):032807.

[278] Solé-Ribalta, A., Gómez, S., and Arenas, A. (2016). Congestion induced by the structure of multiplex networks. *Physical Review Letters*, 116(10):108701.

[279] Song, C., Wang, D., and Barabási, A.-L. (2013). Connections between human dynamics and network science. arXiv:1209.1411.

[280] Sowa, J. F. (1991). *Principles of Semantic Networks: Explorations in Representation of Knowledge*. Morgan Kaufmann Series in Representation and Reasoning. Morgan Kaufmann, San Mateo.

[281] Squartini, T., and Garlaschelli, D. (2011). Analytical maximum-likelihood method to detect patterns in real networks. *New Journal of Physics*, 13(8):083001.

[282] Squartini, T., Mastrandrea, R., and Garlaschelli, D. (2015). Unbiased sampling of network ensembles. *New Journal of Physics*, 17(2):23052.

[283] Starnini, M., Machens, A., Cattuto, C., Barrat, A., and Pastor-Satorras, R. (2013). Immunization strategies for epidemic processes in time-varying contact networks. *Journal of Theoretical Biology*, 337:89–100.

[284] Stehlé, J., Voirin, N., Barrat, A., Cattuto, C., Colizza, V., Isella, L., Regis, C., Pinton, J.-F., Khanafer, N., den Broeck, W., et al. (2011). Simulation of an SEIR infectious disease model on the dynamic contact network of conference attendees. *BMC Medicine*, 9(1):87.

[285] Stopczynski, A., Sekara, V., Sapiezynski, P., Cuttone, A., Madsen, M. M., Larsen, J. E., and Lehmann, S. (2014). Measuring large-scale social networks with high resolution. *PLOS ONE*, 9(4):e95978.

[286] Stumpf, M. P. H., Wiuf, C., and May, R. M. (2005). Subnets of scale-free networks are not scale-free: Sampling properties of networks. *Proceedings of the National Academy of Sciences of the United States of America*, 102(12):4221–4.

[287] Su, H.-N., and Lee, P.-C. (2010). Mapping knowledge structure by keyword co-occurrence: A first look at journal papers in Technology Foresight. *Scientometrics*, 85(1):65–79.

[288] Tacchella, A., Cristelli, M., Caldarelli, G., Gabrielli, A., and Pietronero, L. (2012). A new metrics for countries' fitness and products' complexity. *Scientific Reports*, 2:723.

[289] Takemoto, K., Oosawa, C., and Akutsu, T. (2007). Structure of n-clique networks embedded in a complex network. *Physica A*, 380:665–72.

[290] Tetali, P. (1991). Random walks and the effective resistance of networks. *Journal of Theoretical Probability*, 4(1):101–9.

[291] Tetlock, P. C., Saar-Tsechansky, M., and Macskassy, S. (2008). More than words: Quantifying language to measure firms' fundamentals. *Journal of Finance*, 63(3):1437–67.

[292] The Economist Intelligence Unit. (2013). *Global Liveability Ranking and Report August 2013*. Available at http://www.eiu.com/public/topical_report.aspx?campaignid=Liveability2013 (accessed May 31, 2018).

[293] Thébault, E., and Fontaine, C. (2010). Stability of ecological communities and the architecture of mutualistic and trophic networks. *Science*, 329(5993):853–6.

[294] Timár, G., Dorogovtsev, S. N., and Mendes, J. F. F. (2016). Scale-free networks with exponent one. *Physical Review E*, 94(2):22302.

[295] Török, J., Murase, Y., Jo, H.-H., Kertész, J., and Kaski, K. (2016). What does Big Data tell? Sampling the social network by communication channels. *Physical Review E*, 94:052319.

[296] Traag, V. A., Reinanda, R., and van Klinken, G. (2015). Elite co-occurrence in the media. *Asian Journal of Social Science*, 43(5):588–612.

[297] Treviño S., III, Nyberg, A., Del Genio, C. I., and Bassler, K. E. (2015). Fast and accurate determination of modularity and its effect size. *Journal of Statistical Mechanics*, 2015(2):P02003.

[298] Ugander, J., Karrer, B., Backstrom, L., and Marlow, C. (2011). The anatomy of the Facebook social graph. arXiv:1111.4503

[299] Ulam, S. M. (1960). *A Collection of Mathematical Problems*. Interscience Tracts in Pure and Applied Mathematics. Interscience Publishers, New York, London.

[300] Vanhems, P., Barrat, A., Cattuto, C., Pinton, J.-F., Khanafer, N., Régis, C., Kim, B.-A., Comte, B., Voirin, N., Pinton, J.-F., et al. (2013). Estimating potential infection transmission routes in hospital wards using wearable proximity sensors. *PLOS ONE*, 8(9), e73970.

[301] Varshney, L. R., Chen, B. L., Paniagua, E., Hall, D. H., and Chklovskii, D. B. (2011). Structural properties of the *Caenorhabditis elegans* neuronal network. *PLoS Computational Biology*, 7(2):e1001066.

[302] Vázquez, A. (2003). Growing network with local rules: Preferential attachment, clustering hierarchy, and degree correlations. *Physics Review E*, 67(5):56104.

[303] Vázquez, A., Dobrin, R., Sergi, D., Eckmann, J.-P., Oltvai, Z. N., and Barabási, A.-L. (2004). The topological relationship between the large-scale attributes and local interaction patterns of complex networks. *Proceedings of the National Academy of Sciences of the United States of America*, 101(52):17940–5.

[304] Viswanathan, G., Buldyrev, S. V., Havlin, S., Da Luz, M., Raposo, E., and Stanley, H. E. (1999). Optimizing the success of random searches. *Nature*, 401(6756):911–14.

[305] Vitali, S., Glattfelder, J. B., and Battiston, S. (2011). The network of global corporate control. *PLOS ONE*, 6(10):e25995.

[306] Wall, R. S. (2009). *NETSCAPE: Cities and Global Corporate Networks*. Erasmus University Rotterdam, Erasmus Research Institute of Management, Rotterdam.

[307] Wasserman, S., and Faust, K. (1994). *Social Network Analysis: Methods and Applications*. Structural Analysis in the Social Sciences, volume 8. Cambridge University Press, Cambridge.

[308] Watts, D. and Strogatz, S. (1998). Collective dynamics of "small-world" networks. *Nature*, 393(6684):440–2.

[309] Welch, B. L. (1947). The generalization of "Student's" problem when several different population variances are involved. *Biometrika*, 34(1–2):28–35.

[310] Wilkinson, D. M., and Huberman, B. A. (2004). A method for finding communities of related genes. *Proceedings of the National Academy of Sciences of the United States of America*, 101(Suppl. 1):5241–8.

[311] Wilson, R. J. (1972). *Introduction to Graph Theory*, volume 111. Academic Press, New York.

[312] Wolfrum, M. (2012). The Turing bifurcation in network systems: Collective patterns and single differentiated nodes. *Physica D*, 241(16):1351– 7.

[313] Yang, S.-J. (2005). Exploring complex networks by walking on them. *Physical Review E*, 71(1):16107.

[314] Yaveroğlu, Ö. N., Malod-Dognin, N., Davis, D., Levnajic, Z., Janjic, V., Karapandza, R., Stojmirovic, A., and Pržulj, N. (2014). Revealing the hidden language of complex networks. *Science Reports*, 4:4547.

[315] Zhao, K., Stehlé, J., Bianconi, G., and Barrat, A. (2011). Social network dynamics of face-to-face interactions. *Physical Review E*, 83(5):056109.

[316] Zlatic, V., Bianconi, G., Diaz-Guilera, A., Garlaschelli, D., Rao, F., and Caldarelli, G. (2009). On the rich-club effect in dense and weighted networks. *European Physical Journal B*, 67(3):271–5.

[317] Zlatić, V., Ghoshal, G., and Caldarelli, G. (2009). Hypergraph topological quantities for tagged social networks. *Physical Review E*, 80(3):36118.

[318] Zollo, F., Novak, P. K., Del Vicario, M., Bessi, A., Mozetič, I., Scala, A., Caldarelli, G., and Quattrociocchi, W. (2015). Emotional dynamics in the age of misinformation. *PLOS ONE*, 10(9):e0138740.

This series focuses on the impact of transport planning policy and implementation on the wider society and on the participation of the users. It discusses issues such as: gender and public transport, travel for the elderly and disabled, transport boycotts and the civil rights movement etc. Interdisciplinary in scope, linking transport studies with sociology, social welfare, cultural studies and psychology.

Published:

Transport Policy
Learning Lessons from History
Edited by Colin Divall, Julian Hine and Colin Pooley

The Mobilities Paradigm
Discourses and Ideologies
Edited by Marcel Endres, Katharina Manderscheid and Christophe Mincke

Ports as Capitalist Spaces
A Critical Analysis of Devolution and Development
Gordon Wilmsmeier and Jason Monios

Non-motorized transport integration into urban transport planning in Africa
Edited by Winnie Mitullah, Marianne Vanderschuren and Meleckidzedeck Khayesi

Forthcoming:

Visions, Concepts and Experiences of Travel Demand Management
Edited by Gerd Sammer, Michael Bell and Wafaa Saleh

Digital Social Networks and Travel Behaviour in Urban Environments
Edited by Pnina O. Plaut and Dalit Shach-Pinsly

Digital Social Networks
and Travel Behaviour
in Urban Environments

Edited by Pnina O. Plaut and
Dalit Shach-Pinsly

We dedicate this book in tribute to the memory of our dear colleague Prof. Rein Ahas (1966–2018), a member and group leader in EU COST Action TU1305.

Contents

Tables

Contributors

Arnaud Adam is a PhD student in geography who has been hired by the BruNet project, where he studies the spatial organization of the Brussels area by means of community detection and new information and communication technology (ICT) data. He is also conducting sensitivity analyses of the community detection methods.

Giannis Adamos, PhD, graduated from the School of Engineering of University of Thessaly, Department of Civil Engineering, in 2004. In 2006, he acquired his MSc diploma in Applied Engineering and Systems Simulation and, in 2016, his PhD in Transportation Engineering from the Department of Civil Engineering of University of Thessaly. Since 2006, he is a scientific researcher at the Traffic, Transportation and Logistics Laboratory of University of Thessaly. From 2010 to 2015, he worked as a research associate at the Hellenic Institute of Transport of the Centre for Research and Technology Hellas in Thessaloniki, Greece, and from 2008 to 2011, he was a lecturer at the Technological Educational Institute of Larisa, Greece.

Rein Ahas, PhD (1966–2018), was Professor of Human Geography at the University of Tartu and founder of Mobility Lab, University of Tartu, Estonia. His studies mainly focused on the spatial mobility, urban studies, ICT and tourism and environmental change. His special field of interest was developing new methods for mobile phone–based research, concentrating to call detail records and smartphone-based active tracking. He was a member of the Eurostat Big Data working party and one of the initiators of the international conference series "Mobile Tartu". He was Research Professor of the Estonian Academy of Sciences from 2013 to 2015, and he was on the editorial board of the *Journal of Location Based Service* (Taylor & Francis), *Big Data and Society* (SAGE) and *Travel Behavior and Society* (Elsevier).

María del Mar Alonso-Almeida is Lecturer of Business Administration in Autonomous, Faculty of Economics and Business Administration, University of Madrid in Spain. She is interested in exploring the travel behaviour and social networks from a tourism viewpoint. She is interested in the study of transformations in destinations and organizations due to social media, as well

as in the analysis of economic impact and the adaptability of new economic phenomenon.

Constantinos Antoniou is Full Professor and holds the chair of Transportation Systems Engineering at the Technical University of Munich, Germany. He has a diploma in Civil Engineering from the National Technical University of Athens (1995), an MS in Transportation (1997) and a PhD in Transportation Systems (2004), both from the Massachusetts Institute of Technology. His research focuses on modelling and simulation of transportation systems, intelligent transport systems, calibration and optimization applications, road safety and sustainable transport systems, and in his 20-plus years of experience, he has held key positions in a number of research projects in Europe, the United States and Asia, while he has also participated in a number of consulting projects.

Antònia Casellas is Professor and Researcher in the Geography Department at Universitat Autònoma of Barcelona (UAB) and Coordinator of the Master's Degree in Interdisciplinary Studies in Environmental, Economic and Social Sustainability at Institut de Ciència I Tecnologia Ambientals–UAB. Her research focuses on the interaction among economic viability, governance and urban morphology. She holds a PhD in Urban Planning and Policy Development, a Master in City and Regional Planning (MCRP) from Rutgers University (USA) and two bachelor degrees, in Philosophy from Universitat de Barcelona and in Communication Science from the UAB.

Emmanouil (Manos) Chaniotakis is Research Associate and PhD candidate at the Technical University of Munich. Until September 2016, he was a Research Associate at the Hellenic Institute of Transport of the Centre for Research and Technology Hellas. He holds a diploma in Rural and Surveying Engineering (2011) from the Aristotle University of Thessaloniki and an MSc in Transportation Infrastructure and Logistics (2014) from the Delft University of Technology. His research focuses on modelling and simulation of transportation systems, demand modelling and big data analysis in transportation.

Mario Cools holds a master's degree in applied economics, with a major quantitative business economics and a minor operations research (University of Antwerp, 2004) and a master's degree in applied statistics (Hasselt University, 2005). He is appointed as full-time associate professor at the Faculty of Applied Sciences of the University of Liège, where he is charge of research domain "diagnosis and analysis of transport and its externalities" of the research unit LEMA. LEMA is a research group attached to the Urban & Environmental Engineering research unit at the University of Liège, and develops its researches in the domains of (1) architectural and urban modelling with the goal to improve the sustainability of the built environment, (2) urban planning and urban forms and (3) diagnosis and analysis of transport and its externalities.

João de Abreu e Silva is Associate Professor at Instituto Superior Técnico, Universidade de Lisboa. His main research interests include travel behaviour

and transport land-use interactions. João has participated in several COST Actions and in projects related with real-time data collection and data fusion. He has authored more than 40 articles in international peer-reviewed journals and has been involved in several international scientific networks, being currently member of the ADB10 and ADD30 Committees of the Transportation Research Board. He is Vice-Chair of the World Society for Transport and Land Use Research (WSTLUR) and a member of the Steering Committee for the Travel Survey Conference. He is editor of the *Journal of Transport and Land Use*.

Loukas Dimitriou is Assistant Professor in Department of Civil and Environmental Engineering, University of Cyprus (UCY) and Founder and Head of the LaB for Transport Engineering, UCY. His research interests focus in the application of advanced computational intelligence methods, concepts and techniques for understanding the complex phenomena involved in realistic transport systems and, further, in developing design and control strategies such as to optimize their performance. The methodological paradigms that he proposes utilize and/or combine elements from data science, behavioural analytics, complex systems modelling and advanced optimization and are applied to traditional fields of transport, such as demand modelling, travel behaviour and systems organization, optimization and control. He has more than 120 publications in peer-reviewed journals, proceedings of conferences and book chapters, while he is a regular reviewer in almost 50 international journals. Also, he an active member of international scientific organizations and committees.

Domokos Esztergar-Kiss is a research fellow at Budapest University of Technology and Economics. Since 2014 he is the international project coordinator of the Faculty of Transportation Engineering and Vehicle Engineering. His main fields of research are multimodal journey planning, personalized information services, activity-chain optimization, mobility as a service and sustainable mobility planning. He has organized the MT-ITS 2015 (Models and Technologies for Intelligent Transportation Systems), EWGT 2017 (Euro Working Group on Transportation meeting) and hEART 2019 (European Association for Research in Transportation) conferences. He is member of COST Actions TU1004 (Modelling Public Transport Passenger Flows in the Era of Intelligent Transport Systems), TU1305 (Social Networks and Travel Behaviour) and CA16222 (Wider Impacts and Scenario Evaluation of Autonomous and Connected Transport). He is involved in several Horizon 2020 and Interreg projects, such as MoveCit (Engaging employers from public bodies in establishing sustainable mobility), LinkingDanube (Transnational, multimodal traveller information and journey planning for environmentally friendly mobility in the Danube Region), MaaS4EU (Engaging employers from public bodies in establishing sustainable mobility and mobility planning) and electric travelling (platform to support the implementation of electromobility in smart cities based on information and communication technology [ICT] applications). He has prepared reviews for several international peer reviewed journals

(International Journal of Intelligent Transportation Systems Research, EURO Journal on Transportation and Logistics, Transportation Research Part A: Policy and Practice, Periodica Polytechnica Transportation Engineering).

Olivier Finance is a post-doctoral researcher in geography at UC Louvain (Center for Operations Research and Econometrics), hired in the BruNet project. He is developing an interactive visualization tool called atlas.brussels, to wrap up main results obtained during the project. He is moreover associate member at UMR8504 Géographie-cités (Paris).

Slaven Gasparovic is a geographer, and he works as Assistant Professor at the University of Zagreb, Faculty of Science, Department of Geography. He received his MA and PhD at the same university. He worked at several primary and secondary schools as a geography teacher. His research deals with transportation geography, especially with social aspect of transportation (i.e. transportation disadvantage, transportation based social exclusion, transportation poverty, transportation of vulnerable groups), as well as interrelation between ICT, social networks and travel behaviour. He has participated in several projects (e.g. CH4LLENGE, Sinergi, bilateral projects). He has been member of several COST Actions.

Maria Candelaria Gil-Fariña has been a full professor since 1997 in the Department of Applied Economics and Quantitative Methods – Universidad de la Laguna, Tenerife. A graduate in Economic Sciences and Management, with specialization in International and Regional Economics (1990), she has a PhD in Economic Sciences (1995). Her topics of research include Padé approximation in numerical analysis and multivariate time-series analysis; algebraic properties of Padé approximation and applications to economics; discrete and continuous time models applied to economics; the application of wavelets to economic (agricultural, energy, water) and financial (mortgages, gross domestic product) data; the estimation of water consumption in specific territories; sensitivity analysis for banana production according to the given incentives; and financial data modelling.

Bilin Han is a PhD candidate of the Urban Planning and Transportation Group of the Eindhoven University of Technology. Her research concerns modelling of work schedule arrangements and work participation in dual-earner households with children under social influence.

Matthew Hanchard is a Research Associate at the University of Glasgow working at the intersection of digital sociology, data science and urban studies. His current research with Beyond the Multiplex (AHRC: AH/P005780/1) focuses on developing a computational ontology from mixed-methods research to develop specialised film audiences. Matthew also acts as co-investigator for a Glasgow-Sydney University collaborative project called SmartPublics which focuses on the cultural, political, social and governance implications of 5G-enabled smart technologies in public spaces.

Bin Jiang is Professor in GeoInformatics and Computational Geography at University of Gävle, Sweden. He worked in the past with Hong Kong Polytechnic University, and University College London's Centre for Advanced Spatial Analysis. He is the primary developer of the software tool Axwoman for topological analysis of very large street networks. He invented the new classification scheme head/tail breaks for scaling analysis of big data. He is the founding chair of the International Cartographic Association Commission on Geospatial Analysis and Modeling. He used to be Associate Editor of international journal *Computer, Environment and Urban Systems* (2009–2014), and is currently Academic Editor of the open-access journal *PLOS ONE* and Associate Editor of *Cartographica*. His research interests centre on geospatial analysis and modelling of urban structure and dynamics, for example agent-based modelling, scaling hierarchy and topological analysis applied to street networks, cities and geospatial big data. Inspired by Christopher Alexander's work, he developed a mathematical model of beauty, which helps address why a design is beautiful, and how much beauty the design has.

Tom Erik Julsrud, PhD, is a sociologist and works as Senior Research Scientist at CICERO Center for International Climate Research and The Institute of Transport Economics (TOI) in Oslo. His current research interests include social practice theory, sustainable consumption, sociotechnical innovation theory, trust and collaborative consumption. He has recently published the book *Trust in network organizations*, at Fagbokforlaget (2018).

Sven Kesselring is a sociologist. He holds a professorship in Sustainable Mobilities at Nürtingen-Geislingen University, Germany (HfWU). He is the director of the Sustainable Mobilities master program at HfWU, the founder of the Cosmobilities Network, and its speaker since 2004. Cosmobilities is a global network of mobility researchers which promote social science based mobilities research. His main research areas are sustainable mobility policies, the transformations of automobilities, corporate mobility regimes and the social theory of the mobile risk society.

Jinhee Kim is Assistant Professor of the Department of Urban Planning and Engineering at Yonsei University, Republic of Korea. He was Assistant Professor of the Urban Planning and Transportation Group of the Eindhoven University of Technology, the Netherlands, until 2018. He earned a PhD in Urban Planning and Engineering, emphasizing on transportation planning from Yonsei University, August 2013. He has research interests in modelling behavioural process in activity-travel decisions, including the effects of social networks on decision-making process, and developing advanced discrete-choice models. He is a member of the International Association for Travel Behaviour Research and a life member of the Korean Society of Transportation.

Agnieszka Lukasiewicz (PhD candidate) is Researcher in Economic Division, the Road and Bridge Research Institute, Poland. She is an expert in stakeholders as well as managing large infrastructure projects analysis. Agnieszka's

doctoral thesis is about financial dimensions of stakeholders' influence on large road infrastructure projects. She was also a project manager of national research and development project Collaboration of standard proceedings with stakeholders' relationships in the process of planning and realizing road and rail infrastructure projects. She has an experience in participation in various research projects related to transportation. She also has practice in cooperating with different kind of institutions, local authorities and informational documents preparation, especially in context of foreign direct investment (2000–2005 employed in Polish Information and Foreign Investment Agency). She also has experience in strategy building for enterprises and governmental institutions (e.g. Ukrainian Road Research Strategy project pro-investment strategy for Polish ICT sector, foreign investor relations based on electronic communication channels for Polish Information and Foreign Investment Agency and local authorities). She has taken part in many international projects, FP7, FP6, COST Actions and recently HORIZON 2020 (BENEFIT project).

Jasna Mariotti is Lecturer in Architecture at Queen's University Belfast. She studied architecture in Skopje and at Delft University of Technology, where she graduated with honours in 2007. In 2014 she completed a PhD on post-socialist cities and their urban transformations at the Faculty of Architecture, University of Ljubljana. Her current research focuses on the relationship between urban history, planning and architecture, urban governance and city retailing in the 20th and 21st centuries. Since 2010, Jasna Mariotti has worked on projects and urban interventions in different scales. Previously she was architect and urban designer in WEST 8 in Rotterdam, working on master plans and public space design.

Yufan Miao is a PhD researcher in the Future Cities Laboratory (FCL), Singapore–ETH Centre. He received his master's degree in Computational Science from Uppsala University and bachelor's in Geomatics from University of Gävle under the supervision of Prof. Bin Jiang. His current research interest is cognitive design computing as part of the Big-Data Informed Urban Design project in the research scenario of responsive cities in FCL. He aims to enrich and apply methods from cognitive computing to automate and optimise the urban design process for the facilities of urban designers.

Marija Mitrović Dankulov is Assistant Research Professor at the Scientific Computing Laboratory and deputy head of Innovation Centre at the Institute of Physics Belgrade, University of Belgrade. She has extensive knowledge and experience in theoretical and computational physics. Her main research interests are statistical physics of complex systems, with the emphasis on physics of socio-economic systems, and theory of complex networks. She completed her PhD in statistical physics at the Faculty of Physics, University of Belgrade in 2012. After her PhD studies, during which she was employed at the Department of Theoretical Physics, Institute Jožef Stefan, Slovenia, she undertook postdoctoral work at Department of Biomedical Engineering and Computational Science, Aalto University, Finland.

Gaëtan Montero, PhD student in geography (hired by the BruNet project). He particularly studies morphological issues and their link with social networks in the urban agglomeration of Brussels.

Itzhak Omer is Professor and Researcher in the Geography and Human Environment Department at Tel Aviv University, and also directs the Urban Space Analysis Laboratory, where he and his students work on the relationship between the urban environment and the spatial behaviour of individuals and developing of spatial analysis methods and simulation models, for studying movement and activities in the city. The areas of his academic interest include urban modelling, spatial behaviour and cognition, urban morphology, urban movement, urban systems and social geography of the city.

Pnina O. Plaut is Associate Professor of Transport and Urban Planning and Policy at the Faculty of Architecture and Town Planning, Technion, Israel Institute of Technology. The Vice Dean for Graduate Studies and Research, the former chair of the Graduate Program in Urban and Regional Planning and Director of the Center for City and Regional Studies at the Technion, Prof. Plaut holds a BSc in Civil Engineering, an MSc in Urban and Regional Planning from the Technion and a PhD in City and Regional Planning from the University of California, Berkeley. Prof. Plaut is the Initiator and Chair of the EU Horizon 2020 COST Action TU1305 titled: Social Networks and Travel Behaviour. Her fields of interests include transportation and land-use planning and policy, the impacts of infrastructure (transportation and telecommunications) on economic development and urban/regional structure and the environment. Her special research interests are smart cities and new mobility services and the nexus among ICT social networks, travel behaviour and urban environments.

Anniki Puura is a PhD student and Junior Research Fellow of Human Geography at the University of Tartu. Her research is focused on relationships between social networks and spatial mobility. In her studies, she mainly uses mobile phone–based data sets.

Soora Rasouli is Full Professor and Chair of the Urban Planning and Transportation Group, Eindhoven University of Technology. She received her master of Road and Transportation in 2005 at Tehran University and completed her PhD at the Eindhoven University of Technology. She specializes in developing travel behaviour models under uncertainty, regret-based models and social network analysis, in general, and (travel) behaviour, in particular.

Dalit Shach-Pinsly is an architect and urban designer who received her PhD at the Faculty of Architecture and Town Planning, Technion – IIT. She is currently a researcher and senior lecturer at the Technion – IIT. She was a Partner Investigator in the H2020-MG-9.2–2014 MIND-SETS – Mobility Innovations for a New Dawn in Sustainable Transport Systems. She is participating in the EU FP7 TUD COST Action TU1305 as the co-initiator and Management Committee member. She was a senior researcher of FP7-DESURBS-EU-Designing

Safer Urban Spaces. Formerly she joined the College of Built Environments at the University of Washington as a post-doctoral fellow sponsored by the competitive Marie Curie EU IOF Fellowship (2008–2011). Her fields of interests include built environment; measuring and evaluating diverse qualitative aspects of the urban environment, such as security and safety, visibility and privacy; master plan, analysis, methods and tools; and urban regeneration.

Fariya Sharmeen is a scholar of urban planning with specializations in built environment, travel behaviour and social networks. She is Assistant Professor of Sustainable Transportation at Radboud University, Nijmegen. Her research interests consist social contexts of urban and transportation planning and urban regeneration. Her PhD research was focused on the relationship between social network dynamics and travel behaviour. She is the recipient of Royal Geographic Society Postgraduate Award in 2013. Previously she worked as Lecturer at Bangladesh University of Engineering and Technology, Bangladesh, and as researcher at Cities Institute, UK, and Utrecht University, the Netherlands.

Siiri Silm, PhD, is Research Scientist of Human Geography at the University of Tartu and acting Head of Mobility Lab, University of Tartu. Her fields of research include human spatial-temporal behaviour, analyses of urban space, social networks and segregation (ethnic, age-related). She has participated since 2004 in developing mobile phone–based methodology and conducting research.

Francois Sprumont is a PhD student at the MobiLab Transport Research Group of the University of Luxembourg. After a bachelor in geography (Namur, Belgium), he obtained a master's degree in Spatial Planning from the University of Luxembourg. At that period (2009), he was among the 10 best students of the university and, as an award, participated to a one-month summer school in Jinan, China. More recently, he obtained his own grant from the FNR (Fonds national de la recherche) for the STABLE (Sustainable Transport behaviour considering Activity chains of BelvaL commutErs) research proposal.

Silvana Stefani is Full Professor of Mathematics Applied to Economics and Finance, Department of Statistics and Quantitative Methods, Università Milano – Bicocca, Italy. Her main research activities include discrete mathematics applied to economics and finance: financial graphs; spectral properties of matrices and graphs, networks and centrality; energy and environmental markets: modelling optimal production and hedging strategies in commodity and (conventional and renewable) energy markets, including mean reverting strategies in commodity markets; environmental markets and European Union Allowances; forward and futures premia; and ranking and journal classification, applying fuzzy statistical techniques to bibliometry.

Isabelle Thomas is Research Director at FRS-FNRS and full professor at the Université catholique de Louvain (Belgium). She is the 2016–2019 research

director of the Centre of Operation Research and Econometrics (CORE). Her research fields are anchored in quantitative geography (economic/transport/urban geography) often in a multidisciplinary context. She has (co-)authored numbers of publications in books and international scientific journals. She is and was involved in several national and international research projects and scientific commissions; she is member of the editorial board and referee for several journals in geography and spatial economics.

Harry Timmermans holds a PhD degree at the Catholic University of Nijmegen. Since 1976 he has been affiliated with the Department of the Built Environment of the Eindhoven University of Technology, the Netherlands. In 1985 he was appointed chaired professor of Urban Planning. His main research interests concern the study of human judgement and choice processes. In 2018, he received an honorary doctorate at Hasselt University, Belgium. Currently, he is also associated with the Department of Air Transportation Management, Nanjing University of Aeronautics and Astronautics.

Maria Tsami is Spatial Planning and Development Engineer (MSc) holding also an MSc in computer sciences. She is a scientific researcher and PhD candidate of the Traffic, Transportation and Logistics Laboratory at the Department of Civil Engineering of the University of Thessaly, and a scientific researcher of the Hellenic Institute of Transport. Her core research activity focuses on dynamic transit modelling, taking into account information and quality-of-service indicators.

Ann Verhetsel is Full Professor Economic Geography at the University of Antwerpen; she is Head of the Department of Transport and Regional Economics. She fulfils her appointment (research, education and services) in areas at the interface of geography, economics and planning. Ongoing research assignments under her leadership include projects and PhDs in the fields of mobility, smart specialization and new regional economic policy, networks within the logistics industry, the geographical dimension of venture capital, urban logistics and the location of retail.

Anne Vernez Moudon is Professor Emerita of Architecture, Landscape Architecture, and Urban Design and Planning; Adjunct Professor of Epidemiology and Civil and Environmental Engineering at the University of Washington, Seattle, where she also directs the Urban Form Lab. Dr. Moudon holds a BArch. (Honors) from the University of California, Berkeley, and a Doctor ès Science from the École Polytechnique Fédérale of Lausanne, Switzerland. Dr. Moudon was President of the International Seminar on Urban Morphology; Faculty Associate at the Lincoln Institute of Land Policy in Cambridge, MA; Fellow of the Urban Land Institute in Washington, D.C.; and a national advisor to the Robert Wood Johnson Foundation program on active living research. She is Professeur des Universités and Chercheur Associé, NEMESIS team of IPLESP (Pierre Louis Institute of Epidemiology and Public Health) in Paris. Dr. Moudon lectured at universities in Europe, Latin America and Asia. She

published extensively in urban design, transportation and public health journals. Her books include *Built for Change: Neighborhood Architecture in San Francisco* (MIT Press 1986), *Public Streets for Public Use* (Columbia University Press 1991) and *Monitoring Land Supply with GIS* (Wiley & Sons 2000).

Bridgette Wessels is currently Professor in Social Inequality at the University of Glasgow (UK). Her research focuses on the innovation and use of digital services in social context. She has addressed the Internet and World Wide Web in everyday life, the public sphere, community life, telehealth, digital divides and social networks. Her books include *Communicative Civic-ness: Social Media and Political Culture* (Routledge 2018), *Open Data and Knowledge Society* (Amsterdam University Press 2017), *Understanding the Internet: A Socio-Cultural Perspective* (Palgrave 2010) and *Social Change: Process and Context* (Palgrave 2014).

1 Social networks theory

Definitions and practice

Marija Mitrović Dankulov, María del Mar Alonso-Almeida, Fariya Sharmeen and Agnieszka Lukasiewicz

Introduction

Humans are social beings. Development and organization of societies and the behaviour of individuals are governed and shaped by social interactions (Wasserman & Faust, 1994). That is why it is not surprising that the structure and evolution of social networks are subject of various scientific disciplines, from sociology, economics and transportation studies to mathematics, computer science and physics. Sociologists introduced the term and the idea of social networks at the end of the 19th century to study the emergence of different social phenomena. Further development of this field during the first half of the twentieth century was led by scientists working in the field of psychology, anthropology and mathematics. During this period, scientists started with systematic recordings and analysis of social interactions in small groups, while mathematicians worked on developing a formalism to quantify the structure of social networks graph theory. The development of Information and Communication Technologies (ICT) at the end of the twentieth century and the availability of large data sets about human behaviour attracted the attention of physicists and computer scientists. These disciplines brought new concepts and powerful quantitative methods which further advanced our knowledge about the structure and dynamics of social networks (Sen & Chakrabarti, 2014).

This chapter provides a brief literature review of works on the topic of social networks in transportation studies, economics and physics. All these disciplines use a concept of social network and define it similarly. The differences between these disciplines are research problems and questions related to social networks, as well as their approaches. In the first part of the chapter, we provide a definition and classification scheme of social networks and describe some quantitative measurements of the network structure. In the second part of the chapter, we summarize the most relevant application of social network theory in transportation studies, economics and physics. We pay particular attention to the results, which indicate the connection between the structure of social networks and mobility patterns and travel behaviour of individuals.

Social networks theory

A social network is a theoretical concept used for the quantitative and qualitative description of social entities and relations between them. The social entities, actors, can be individuals, corporate or collective social units. A tie establishes a linkage between a pair of actors, and it can express a relation between two social entities like talking, kinship, friendship or business relations. In mathematical terms, a social network is a set of nodes (vertices), representing actors, and edges (links), representing social ties. Social ties can represent a direct relation, as in friendship or sexual partners' networks, or actors can interact indirectly, through artefacts, for example, a network of bloggers (Sen & Chakrabarti, 2014).

Social network analysis is used for exploring and quantifying patterns of relationships that arise among interacting social bodies, mostly individuals. An explicit assumption of such an approach is that indirect relationships in social groups matter. A particular backbone of social network analysis is that it provides standardized mathematical methods for calculating measures of sociality across levels of social organization, ranging from the population and group levels to the individual level (Freeman, 1984; McCowan et al., 2011). The concept of graphically representing social relationships is not new (Foster, Rapoport, & Orwant, 1963). Nevertheless, recent developments and widespread accessibility of network software have enabled easier visualization and exploration of complex social structures.

Social networks can be roughly divided into three classes based on the type of their ties: single-layer, temporal and multiplex networks (Boccaletti, Latora, Moreno, Chavez, & Hwang, 2006; Holme & Saramäki, 2012; Boccaletti et al., 2014). Single-layer networks are used for representation social systems whose actors interact through only one type of interaction, for example network of co-workers in one company. Depending on whether there are one or two types of actors, these networks can be monopartite or bipartite. Weighted networks are used to represent the systems where interactions can be of different strength. Finally, the interactions can be symmetric (undirected networks), as in the network of co-workers where the relationship is mutual or asymmetric (directed networks), where relationships are not reciprocal, such as student–mentor social network. Temporal networks are used for the representation of networks where ties and nodes are active at certain points in time (Holme & Saramäki, 2012), for example in mobile phone communication networks where phone call has a limited duration. Multiplex or multilayer networks are composed of a multiplicity of overlapping single-layer networks that capture different types of social connection, for instance actors can use different means of communication (phone calls, short message services, or online media) where each layer of a communication network has its own properties and dynamics. The size of this chapter does not allow us to cover all three types of network representations. A detailed description of methods and tools for the quantitative description of these networks can be found in review articles (Boccaletti et al., 2006; Holme & Saramäki, 2012; Boccaletti et al., 2014). Here we present only several quantitative measures used

for the description of the topological structure of single-layer binary undirected networks: degree distribution, clustering coefficient and its dependence on node degree, of assortativity index, the dependence of average first neighbour degree on node degree and shortest path. It was shown that these properties are essential for the description of the topological structure of most real complex networks, including social networks (Orsini et al., 2015).

A quantitative description of social and complex networks requires the right set of tools. Graph theory is a natural framework for the mathematical representation of social and complex networks (Boccaletti et al., 2006). A network or a graph consists of two sets: a set of nodes (vertices) and a set of links (edges) that connects those nodes. Two connected nodes are said to be *adjacent* or *neighbouring*. The *node degree* is the number of its first neighbours.

One of the essential topological features of a network is degree distribution $P(q)$ defined as the probability that randomly chosen node has a degree q. The degree distribution is used for quantifying network heterogeneity at the local level and can be estimated as the fraction of nodes in the network having a degree q. The degree distribution is sufficient for a complete description of the structure of uncorrelated complex networks (Boccaletti et al., 2006; Orsini et al., 2015). However, most of the real, complex networks are correlated in the sense nodes with certain values of degree are more likely to be linked to each other. The degree correlations are characterized by conditional probability $P(q|q')$ which equals a probability that there is a link between nodes with degrees q and q'. The direct evaluation of conditional probability from the data is often not possible. Degree–degree correlations in a network can be estimated using average-nearest-neighbours degree and its dependence-on-node degree. For uncorrelated networks, the average-nearest-neighbours degree is independent of node degree. Correlated networks can be divided into two classes: assortative networks for which the average degree of nearest neighbours grows with q, and is disassortative where the opposite behaviour is observed.

Clustering, or transitivity, is another topological property of the networks which is particularly important for social networks. The clustering coefficient of a node is the probability that two randomly chosen neighbours of a node are also neighbours. It is estimated as a fraction of existing links out of all possible links between neighbours of a node. By averaging clustering coefficients over all nodes, one obtains the network clustering coefficient. Node and network clustering coefficients take the values between 0 and 1. Networks with a high value of clustering coefficient are considered to be clustered.

The shortest path has been one of the most important properties for characterization of network structure. A path between two nodes is an alternating sequence of nodes and edges, in which no node is visited more than once. The path of the minimal length is known as the shortest path. The shortest path of the largest length in the network is known as network diameter. The average shortest path of a network is defined as the mean of geodesic lengths over all pairs of nodes in the network. Most of the real, complex networks have relatively small average shortest-path length compared to their size, which is why they are often called

small-world networks. A recent study (Orsini et al., 2015) has shown that small-world property can be explained using degree distribution, degree–degree correlation and dependence of clustering coefficient the on node degree.

Real complex networks are characterized by inhomogeneities on the mesoscopic level, also known as communities. The notion of community and the term itself have been proposed in the social sciences (Wasserman & Faust, 1994). In single-layer networks community is defined as a group of nodes more densely connected than with the rest of the network. The detection and quantitative description of the community structure of complex networks have attracted much attention in the past two decades (Fortunato, 2010). The networks can be characterized by different community structure. The community structure can be simple, with distinct communities. However, typical real-world networks have overlapping communities or communities that are hierarchically embedded. A detailed description of different algorithms and methods for finding different types of communities in static single-layered complex networks can be found in the paper by Fortunato (2010).

Social networks in various disciplines

Social networks in transportation studies

In the field of transportation, seminal empirical research on social networks has been documented. Among them are the studies by Wellman, Carrasco and colleagues (Carrasco, Hogan, Wellman, & Miller, 2008; Carrasco, Miller, & Wellman, 2008; Carrasco & Miller, 2005); Axhausen, Frei and Kowald (Axhausen, 2008; Frei & Axhausen, 2008; Kowald & Axhausen, 2012); and Timmermans, Van den Berg, Arentze, Sharmeen and colleagues (van den Berg, Arentze, & Timmermans, 2008, 2009, 2010, 2011, 2012; Sharmeen, Arentze, & Timmermans, 2013, 2014a, 2014b, 2015a, 2015b, 2016, 2017), based in Toronto, Zurich and Eindhoven, respectively. Most of those studies collected primary survey data, asking respondents to report a section of their social networks. Those studies came together into a comparative analysis of personal social network features in different spatial settings (Kowald et al., 2013) and were also combined in a recent book providing a much-needed overview of the relevant studies (Kowald & Axhausen, 2015).

Among the first attempts to understand social networks in transportation studies, is the connected lives study where name generators were employed to collect social network data (Carrasco & Miller, 2005; Carrasco et al., 2008a, 2008b). They focused on social activity generation, explicitly incorporating social network characteristics of each network member (alter) as well as the characteristics of the overall social structure. For a better understanding of the spatial distribution of social activity, they incorporated activity space anchor points, based in the home, institution and public spaces. Simultaneously, they characterized those places based on recurrence – whether these are regular places or not. The role of ICT on social interaction was also investigated.

On the other hand, Axhausen (2008) argued that social network membership influences a person's mental map, and therefore, the geographical network should have an impact on travel behaviour. In this study, he discussed in detail the survey instruments and data requirements for social network studies. In a subsequent study, Frei and Axhausen (2008) elaborated how geographical distances in personal social networks influence travel behaviour. They found that face-to-face contact frequency decreases with increasing distance whereas ICT (email) frequency increases.

Van den Berg et al. (2008, 2009, 2010, 2011, 2012, 2013) went in detail to investigate the interactions among ICT, socio-demographics, land use and social interactions. They reported a series of analyses of their social interaction diary and social network data collected in Eindhoven in 2008. The study employed a social interaction diary, followed by a name generator survey, to collect social network and interaction data. They extensively examined the impact of ICT on social travel behaviour and reported that the results differ significantly from a previous study conducted by Molin, Arentze, and Timmermans (2008), who used data about social networks collected in the 1980s, also in the Dutch context, implying that the inter-relations of social network and travel demand have changed over last two decades. They further investigated social travel for the elderly (2011) and the effect of club memberships on social interaction (2012). They reported that the elderly were as mobile as the young population in terms of frequency of social trips. The only difference in travel-mode choice was reported. They also delineated reciprocity in social network size and club memberships and, as expected, involvement in clubs and voluntary organizations increased social trips.

In most of the studies (including those mentioned earlier), social interaction diaries were commonly used, some in a different manner than the other. For example Silvis, Niemeier, and D'Souza (2006) used a similar social interaction diary, where instead of collecting information about network members, they provided an option for the network member to volunteer for the survey. Starting with three seed respondents, they collected information about 24 individuals over three phases of survey design. They concluded that individuals did not mind making longer trips for socializing and visiting family. They studied primarily concentrated on trip generation influence of social network. They also mapped the geography of trip frequencies.

In a similar direction, built environment effects of social interactions were investigated using time geography approach. In a recent study, Farber, Neutens, Carrasco, and Rojas (2014) calculated a Social Interaction Potential (SIP) metric that estimated the potential for an individual to engage in social activities given a specific time and space window. The metric was evaluated using data from cities Ghent and Concepcion. It was developed to assess the relationships between spatial structure and the potential opportunities for face-to-face contacts. The study provided a framework to assess the sociability of urban environments.

Sociability has been investigated from the perspective of solo versus joint activity planning. Ettema and Kwan (2010) analysed the company of social activities among ethnic groups in the Netherlands. They tested many hypotheses, contextual

to social and recreational travel, and found that individuals had multiple networks (such as family, friends, associational and professional) which potentially perform multiple roles.

Thus far, the interplay between social networks and travel behaviour has been explored in travel behaviour research, in a particular space and time context. These studies are focused on static concepts of social networks, whereas social networks are dynamic. To develop a comprehensive understanding of network influence on individual/household's travel behaviour, it is imperative to examine those in a dynamic setting. Given the recency of the inclusion of social network components in travel behaviour studies, it is understandable the forthcoming arena of exploration.

The major setback of investigating social network evolution is perhaps data deficiency. Ideally, a panel study would be required, which is difficult with limited resources. Alternatively, social influences can be computed using stated preference surveys of expressed choices using non-linear utility functions (Kim, Rasouli, & Timmermans, 2014a, 2014b). Reviews of such efforts have also been detailed out in recent papers (Kim et al., 2018). The use of simulated or synthetic datasets has also shown potential to account for network dynamics. Dugundji and Gulyás (2008) aimed to incorporate social influence on transportation mode choice. They developed a multi-agent simulation model of household interactions, looking at how they decided on transportation mode alternatives by carefully distinguishing social and spatial network interdependencies. Páez, Scott, and Volz (2008) described a discrete choice model to account for social influence on decision making as an advancement over auto-correlation analysis. Social network simulation was developed based on the structure analysis tradition of sociology by developing an informal support network. In an earlier publication, Páez and Scott (2007) applied a similar methodology for decisions about telecommuting research. They pointed out certain limitations: first, the limited scope for empirical analysis due to limited mobility data with social network information and second, the dynamics of social networks were not taken into account. For example, although changing residential location meant new neighbours, schoolmates or gym mates, the social network was kept static in their model. Finally, they mentioned that they incorporated no influence from the "rest of the world". It is relatively difficult to incorporate that.

Han, Arentze and Timmermans (2011) presented a dynamic model that simulated habitual behaviour versus exploitation and exploration as a function of discrepancies between dynamic, context-dependent aspiration levels and expected utilities. Principles of social learning and knowledge transfer were used in modelling the impact of social networks, and related information exchange, adaptations of mutual choice sets and formation of common aspiration levels.

Dynamics concerning short-term decision making have been studied, as well as opening the scope for empirical analysis. Hackney and Axhausen (Axhausen & Hackney, 2006; Hackney, 2009) developed a multi-agent representation, incorporating dynamics of the social network, by addition and deletion links, based on feedback through activity-choice sets. They accounted for homophily and associations and assumed some maximum number of contacts per agent.

However, the structure of the social network and its characteristics have not yet been incorporated. To that end, Arentze and Timmermans (2008) developed a theoretical and modelling framework to capture the essence of social networks, social interactions and activity-travel behaviour. The core assumption was that the utility that a person derives from social interaction is a function of dynamic social and information needs, on one hand, and of similarity between the relevant characteristics of the persons involved, on the other. The model is consistent with the traditional social network theories (such as homophily and transitivity) developed in the social science literature. The process model has been tested using arbitrary agents. It led to the conclusion that an individual's social network had an equilibrium size dependent on several factors and changes over time. Although this study does formulate a theory and model of social network dynamics, no empirical data were collected to estimate the parameters of the model and test their theory. The same conclusion holds for Ronald, Dignum, Jonker, Arentze, and Timmermans (2012), which can best be seen of that line of work extension focusing on numerical simulation.

An elaboration of the framework is developed in Kowald and Axhausen (2012). Two Swiss data sets were used to simulate connected personal networks and encounters between actors. The model investigated leisure relationships and provided insights on the connectedness between actors and the factors affecting the leisure relationships between them. Illenberger, Flöteröd, Kowald, and Nagel (2009) conducted a similar simulation with a different approach. They tested network indicators, including edge-length distribution and network-degree distribution, in their model but did not account for properties such as homophily. Although those simulation frameworks were promising, most remain relatively simple. Thus, it remained a computational challenge to integrate large networks and complex social dynamics.

Hence, the dimensions and components of a social network may affect transportation choices in many ways, viz to perform joint activities, as a valuable source of information and social support and influencing daily as well as life choices (Sharmeen et al., 2015a). When social networks change, they potentially bring changes in any, or all, of the ways, eventually changing am individual's activity and travel patterns.

Seminal scholar Wellman and colleagues (Wellman, Wong, Tindall, & Nazer, 1997) reported a complete turnover in the network composition to those who got married during the study period. Furthermore, they outlined that continued telephone interaction and social support positively influenced tie maintenance. In a similar attempt of a qualitative study in three waves, Bidart and Lavenu (2005) found patterns of social network maintenance due to entry to the market, a start of a romantic relationship, childbirth and geographic mobility. Social interactions and relations to activity and travel schedule, however, remained largely unexplored. Furthermore, it is not readily evident how such research can be elaborated to fit the agenda of transportation research.

Only recently empirical studies of the influence of dynamics of social networks on activity and travel behaviour were reported (Chávez, Carrasco, Tudela, 2018;

Sharmeen, 2015; van den Berg, Weijs-Perrée, & Arentze, 2017). While, on one hand, time effects were investigated (Chávez et al., 2018), theories of life-cycle events to trigger changes were employed in the other (Sharmeen & Timmermans, 2014; Sharmeen et al., 2014a, 2014b, 2015a, 2015b; Sharmeen, Chávez, Carrasco, Arentze, & Tudela, 2016; Sharmeen & Sivakumar, 2017; Sharmeen, 2015; van den Berg et al., 2017). Through a series of studies, Sharmeen and colleagues documented the effects of life-cycle events on the interaction frequencies (Sharmeen et al., 2014b), mode choice (Sharmeen & Timmermans, 2014), time allocation (Sharmeen et al., 2013) and overall activity-travel need (Sharmeen et al., 2014a). Beyond those, they constructed a model to compute the population-wide evolution of social networks (Sharmeen et al., 2015b). This model was also replicated in other countries to predict the social network evolution (Sharmeen et al., 2016). The forecasting power and predictability of social contextual variables of individuals and positive correlation with different active travel dimensions were reported (Sharmeen & Sivakumar, 2017).

Empirical evidence from all these studies suggests that personal social networks do evolve with socio-demographic status and life-cycle events. Some ties got stronger and intertwined while others fade due to either a change in priorities or changes in an individual's or a household's time budget. Individuals alter and update their choices under the social influence or at specific points in life. Those points or life-cycle events act as triggers to deliberate choice decisions. Such decisions are better understood when the broader contextual environment of individuals is incorporated. Mainly, the social network composition and geography are overall found to be primarily associated with all aspects of activity-travel scheduling (and residential location choice). Therefore, social network attributes are found to offer a better understanding and are crucial for incorporation in large-scale travel demand forecasting models not only at any one point of time and space but also during the progression and consecutive changes over the short and long term. Together with the growing leisure travel duration and location choices and the explosion of social media, social networks have become increasingly crucial in travel behaviour research. Fuelled by the demand landscape, the field has gained momentum in recent years in transportation research summarized earlier. There is much to be done to realize the full potential of social influence in travel predictions.

Social networks explanation in the transportation field is through the character of social activities, their spatial distribution and frequency, and related travel behaviour. They are the driver behind the activity-travel decisions and a key source of explanation of activity-travel generation (Carrasco & Miller, 2009; Calastri, Hess, Daly, & Carrasco, 2017). Nevertheless, research on the effects of social networks on activity-travel patterns emerged for understanding social travel, in its own right, better and for improving the performance of comprehensive activity-based models of travel demand forecasting in which the prediction of social travel was a weak link. Over the last decade, numerous empirical studies have confirmed the contention that the intensity and nature of social activity-travel behaviour are significantly influenced by the properties of personal social networks (Kim et al., 2018). Thus, while the activity-based modelling has shifted from cross-sectional

to multiple-horizon dynamic models (Sharmeen et al., 2015b; Rasouli & Timmermans, 2014), there is still space of further examination of the subject.

Social networks in economics

Social networks theory is widely applied, including economics and travel studies. There is the economic problem of scarcity in which individuals must face trade-offs while making choices. The available resources are insufficient to give satisfaction to all human wants. Choices made by persons determine the allocation of scarce resources. Social interactions and social networks are useful for examining how such decisions are made and what is the trigger off.

The social context where many economic interactions arise is taken into consideration as a significant factor and the driver of behaviours, decisions and outcomes. Social networks are used in obtaining jobs. They have an inevitable impact on decisions about buying products, education choice and so on. In the field of economic sociology, after Granovetter (1985) published a paper, the new economic sociology aroused. Thus, Granovetter has been connected to the idea of embeddedness. In such the economic relations between individuals or firms are embedded in actual social networks, as well as they do not continue in an abstract, idealized market. Granovetter (1977) stressed the importance of weak ties, more casual acquaintances outside the group of closest friends, for the job-seeking process. Montgomery (1991) digested the findings of various surveys coming up to the conclusion that about half of all currently employed workers found their job through information provided by friends or family members.

The late attentiveness of economists concerning social networks possibly comes from having pushed many economic models to their limits and found that social circumstances can help explain observed economic phenomena in ways that restricted economic models cannot. Individuals are assumed to form or maintain relationships that they find beneficial and avoid or remove themselves from relationships that are not beneficial. That is sometimes captured through equilibrium notions of network formation but is also modelled through various dynamics, as well as agent-based models where specific heuristics are specified that govern behaviour. That choice perspective traces the structure and the properties of networks back to the costs and benefits that they bestow upon their participants (Jackson, 2007).

There is a rising interest in social issues noticed in economics. The approach of new applied models can be connected with fairness criteria and existing inequalities among different groups, for example regarding employment or wages. In such conditions, the problem of market efficiency comes as the social context can change the ways the resources could be allocated. The attitude one can concern coming to the relatively new concept which is collaborative, sharing economy, in which social networks are successfully applied. The sharing economy, which is a fast-growing sector, attracts the attention of policy-makers, as sharing is related to democracy and seems to involve openness, inclusion and equality and is based on social networks and, at some point, could not exist without social media.

In the field of economics, models of network formation based on stochastic algorithms are used. Economists also model the formation of social networks as the outcome of optimization decisions. Social relations are connected with costs and benefits, and in the process of making a decision, the elements are weighted against each other. Those models make it possible to analyse the consequences of technological changes that alter the cost of communication (Mayer, 2009).

Social network analysis in economics has also been used to understand complex economics or management problems and make optimal decisions. Thus, there are some economics and management fields where social networks empirical analysis has been widely developed, such as the following:

- Operations where efficiency measures are used (Sadri, Ukkusuri, & Gladwin, 2017; Blas, Martin, & Gonzalez, 2018)
- The power of diversity in organizations regarding creativity, productivity or knowledge sharing among others (Park, Im, & Sung, 2017; Goyal, Rosenkranz, Weitzel, & Buskens, 2017; Chua, 2018)
- Organizations' collaboration (Shane & Cable, 2002; Chua, 2018; Boschet & Rambonilaza, 2018)
- Finance issues (Mollick, 2014; Huang & Knight, 2017; Polzin, Toxopeus, & Stam, 2018)
- Inversion issues (Garlaschelli, Battiston, Castri, Servedio, & Caldarelli, 2005; Wang & Wang, 2018)
- International trade flows (Barigozzi, Fagiolo, & Garlaschelli, 2010; Barigozzi, Fagiolo, & Mangioni, 2011; Lovrić, Da Re, Vidale, Pettenella, & Mavsar, 2018)

According to Boccaletti et al. (2006, p. 176), *the extensive and comparative analysis of networks from different fields has produced a series of unexpected and dramatic results.* Therefore, another theory to surplus the limitations of social network theory was needed. Thus, the authors, such as those mentioned earlier, explain the research on complex networks begun with the *effort of defining new concepts and measures to characterize the topology of real networks.*

Complex networks theory, according to some authors, provides a framework that can explain how changes in context, economics or human behaviour happen and which way the evolution could be. The theory explains that all networks share common properties. Therefore, applying complex network theory in economics makes possible identification and measure of social networks in this field and make decisions on the matter. Besides, this theory could shed light on how countries, firms or people interact and relate to themselves.

The mobility paradigm states that all places are linked into thin (or fat) networks of connections that stretch beyond each place (Sheller & Urry, 2006). In other words, nowhere can be an island (Gogia, 2006). Mobility, in a broad sense, also includes movements of income, information and images on local and global ways and one-to-one in different ways of communication, such as phones,

smartphones and social media. Empirical analysis with this theory can use qualitative and quantitative methods (Hollstein, 2011).

For example, in the case of tourism, activities are not separated from the places visited. Indeed, the places travelled used to depend, at least in part, on activities can be practised within them. Authors mentioned earlier asserted that exists a complex relationship between the mean and the way of travel and the traveller. Thus, it is necessary to examine the topology of social networks and the patterns of weak ties that could generate small-world property (Urry, 2003, 2004).

Nevertheless, a specific theory which explains complex patterns forms and changes does not exist yet. For that reason, some authors suggest that economics involves the analysis of complex systems that are neither perfectly ordered nor anarchic (Capra, 2004; Sheller & Urry, 2006).

Social network theory can help to identify "critical networks" for example, main stakeholders in a specific place. Besides, it can provide action patterns and the role of every member in the network (Timur & Getz, 2008; Scott, Baggio, & Cooper, 2008). Moreover, complex network theory provides a robust foundation for identifying critical stakeholders and their relationships inside and offside of the network (Timur & Getz, 2008).

Boccaletti et al. (2006) define a complex system that comprises a large number of components usually non-linear and operated in a non-predictive way. In a tourism destination, it is possible to find a vast number of components, both tangible and intangible (Alonso-Almeida & Celemin-Pedroche, 2016). A complex adaptive system is continuously interacting with the environment and generating dynamic adjustments on the structure and behaviour. Therefore, a theory which studies complex systems such as complex network theory will be able to predict future conditions based on past trends (Andersen & Sornette, 2005).

Complex network theory has been built from observation of the real world network properties and structures (Miguéns & Mendes, 2008). In economics, this theory has been applied to ecological networks, financial relations and companies' collaboration mainly (Caldarelli, Battiston, Garlaschelli, & Catanzaro, 2004; Tibély, Onnela, Saramäki, Kaski, & Kertész, 2006).

Statistical physics of social networks

Different systems and macroscopic, collective, phenomena require different models for the description of their dynamics. However, there is one common feature present in all many-body systems, regardless of their nature: the underlying network of interactions between elements that constitute them. The empirical analysis of data collected for different biological and sociotechnical systems has shown that complex and heterogeneous connectivity patterns, typically found in those systems, are one of the key signatures of their self-organizing dynamics (Boccaletti et al., 2006; Holme & Saramäki, 2012; Boccaletti et al., 2014). In the past few decades, physicists have put much effort into understanding the mutual influence between the structure of the interaction network and system dynamics. They

developed a set of quantitative methods and measures that allowed them to study this relationship in detail and better understand the emergence of various phenomena in social systems (Sen & Chakrabarti, 2014; Boccaletti et al., 2006; Holme & Saramäki, 2012; Boccaletti et al., 2014; Castellano, Fortunato, & Loreto, 2009).

Quantitative analysis of the structure of different social networks has shown that several universal features characterize all of them. Famous Milgram experiment has shown that an average number of connections between any two humans in the United States equals six, and that world is, in general, small (Travers & Milgram, 1967). A recent analysis of the shortest paths in Facebook has indicated that the world is even smaller than it was expected (Backstrom, Boldi, Rosa, Ugander, & Vigna, 2012). In a mathematical sense, small-world property means that the average shortest path in the network grows logarithmically with its size (Sen & Chakrabarti, 2014; Boccaletti et al., 2006). Many real-world networks are small world. Social networks are heterogeneous, on both the local and mesoscopic scales. On the local scale, that heterogeneity is manifested in long-tail degree distribution. Individuals differ in the number of friends and acquaintances they have: many of them have only a few friends, while still there are some that have a large number of friends and acquaintances. Societies and social groups are not homogeneous; that is they always consist of smaller groups of people who have more connections with each other than with the rest of the network. Social networks exhibit various types of community structure, from well-defined and separated communities to overlapping communities and communities which can be nested and thus form a hierarchical structure (Fortunato, 2010). Communities can grow, shrink and disappear (Palla, Barabási, & Vicsek, 2007). The size of the community has a crucial role for its dynamics and survival: smaller communities have longer survival time, and their membership is very stable over the long period, while the only way for a large community to survive is to change its members regularly.

Social networks exhibit positive degree–degree correlations. While friendship, collaborative and online social networks are assortative (Newman, 2002), bipartite networks, which represent social dynamics in different techno-social networks, are characterized with weak disassortative mixing (Mitrović & Tadić, 2012). A high value for the clustering coefficient is also a prominent feature of social networks (Boccaletti et al., 2006). The knowledge of how the basic properties emerge in real social networks thus is essential for understanding the evolution and dynamics of related social system and/or phenomena.

Theoretical models of complex networks have proved to be an invaluable tool for studying the evolution of social and complex systems. In parallel with the quantitative exploration of the structure of real social networks, physicists have worked on constructing theoretical models of evolving networks (Boccaletti et al., 2006; Holme & Saramäki, 2012; Boccaletti et al., 2014). Those models, although simplistic, can mimic the properties of real networks. These models enabled a detailed analysis of the emergence of different topological properties in networks and understanding of the primary mechanism that underlie social network evolution. Besides, models enable a comprehensive study on how and to which extent the network structure influences dynamical processes and the emergence

of macroscopic phenomena. We provide a brief description of some of the most fundamental models.

The first and basic model in complex network theory is Erdős-Rényi (ER) random graph model (Erdős & Rényi, 1960). The degree distribution of ER graphs is Poasonian, and they belong to a class of uncorrelated graphs. The value of their clustering coefficient is very low and tends to zero as network size tends to infinity. The ER model is instrumental and irreplaceable as a null model in testing hypothesis related to complex networks. Watts–Strogatz (WS) model, or small-world, model was the first model that successfully reproduced a small diameter and large clustering coefficient of social networks (Watts & Strogatz, 1998). The properties of networks strongly depend on the value of the single parameter that controls the percentage of rewired links. Network without rewired links is a regular graph, while the graph obtained after rewiring of all links is equivalent to ER graph. Graphs with more than 0% and less than 100% of rewired links have a small-world property and a high value of clustering coefficients. However, WS networks do not have a broad degree distribution, one of the most prominent features of many real-world networks, but instead their degree distribution exhibits exponential decay. In order to reproduce the property, Barabasi and Alber (BA) proposed an evolving model of networks (Barabási & Albert, 1999). In the BA model, networks grow by following the preferential attachment rule. In the preferential attachment mechanism, the probability for an old node to be chosen as a target of a new link is proportional to the number of its previous connections. The obtained network has broad, power-law degree distribution, with the exponent equal to three in the limit of infinite size network. Although the average shortest path of these network grows logarithmically with the network size, the clustering coefficient tends towards zero when network size tends to infinity. Other models have been proposed, including different modifications of the BA model, in order to reproduce other properties of real-world social networks. Some of these models are based on the fact that many social networks are embedded in Euclidean space (Barthélemy, 2011), while others take into account different temporal factors, for instance, ageing (Hajra & Sen, 2005).

Network structure has a crucial role in the emergence of different collective states in social systems. The real-world and model-generated networks have been used for studying dynamical processes and their connection with network structure. Opinion dynamics, a process of collective opinion emerging in the social systems, in the convergence time and possibility to reach the consensus in the social group, depend on the topological properties of its underlying social network (Sen & Chakrabarti, 2014). Local heterogeneity of social networks plays a vital role in how and to which extent the disease spread through it. The quantitative methods of complex networks have proved very useful for identifying the individuals that play a crucial role in disease spreading and create the most effective immunization plans (Pastor-Satorras, Castellano, Van Mieghem, & Vespignani, 2015). The structure of social contacts influences long-distance travel of humans (Cho, Myers, & Leskovec, 2011) and can be used for predicting and modelling mobility patterns (Palchykov, Mitrović, Jo, Saramäki, & Pan, 2014). Social

networks are not fixed. They evolve and are influenced by dynamics and activity in social systems. The exchange of emotions has been found to be an essential factor in life and death of online social communities (Mitrović & Tadić, 2012).

Conclusion

Social interactions shape and determine the evolution of human society. In that sense, social network theory, developed in parallel by different scientific disciplines, provided necessary tools for understanding the emergence of various social phenomena. This chapter provides an overview of the most important methods and results related to social networks in transport, economy and physics. Results from transportation studies show that social networks have a significant influence on every aspect of human transportation and travel behaviour. It has been shown that the length and duration of the trips of individual depend on the structure of its social contacts. However, the question of mutual dependence between the social network and travelling behaviour still stays an open question. Transportation and travel behaviour are inevitably significant and are one of the many activities that are influenced by social media and networks. Economic studies have shown that social contacts have a crucial role in the socio-economic development of any human society and that social networks are an essential component of both empirical and theoretical studies in the economy. The concept of complex networks, a broader class of studied systems that includes social networks, has originated from physics. A large number of empirical studies from physics has shown that social networks can be characterized with a relatively small set of properties: they are heterogeneous at the local and mesoscopic scales, assortative, cluster and have small-world property. Empirical studies, in combination with theoretical modelling, have shown that social networks determine the emergence of many collective phenomena in social systems, including travelling behaviour.

References

Alonso-Almeida, M. M., & Celemin-Pedroche, M. S. (2016). Competitiveness and sustainability of tourist destinations. *ESIC Market, 47*(2), 275–289.

Andersen, J. V., & Sornette, D. (2005). A mechanism for pockets of predictability in complex adaptive systems. *EPL (Europhysics Letters), 70*(5), 697.

Arentze, T., & Timmermans, H. (2008). Social networks, social interactions, and activity-travel behavior: A framework for microsimulation. *Environment and Planning B: Planning and Design, 35*(6), 1012–1027.

Axhausen, K. W. (2008). Social networks, mobility biographies, and travel: Survey challenges. *Environment and Planning B: Planning and Design, 35*(6), 981–996.

Axhausen, K. W., & Hackney, J. K. (2006). An agent model of social network and travel behavior interdependence. *Arbeitsbericht Verkehrs-und Raumplanung, 380*.

Backstrom, L., Boldi, P., Rosa, M., Ugander, J., & Vigna, S. (2012, June). Four degrees of separation. In *Proceedings of the 4th Annual ACM Web Science Conference*, Evanston, IL, (pp. 33–42). ACM.

Barabási, A. L., & Albert, R. (1999). Emergence of scaling in random networks. *Science, 286*(5439), 509–512.

Barigozzi, M., Fagiolo, G., & Garlaschelli, D. (2010). Multinetwork of international trade: A commodity-specific analysis. *Physical Review E, 81*(4), 46–104.

Barigozzi, M., Fagiolo, G., & Mangioni, G. (2011). Identifying the community structure of the international-trade multi-network. *Physica A: Statistical Mechanics and Its Applications, 390*(11), 2051–2066.

Barthélemy, M. (2011). Spatial networks. *Physics Reports, 499*(1–3), 1–101.

Bidart, C., & Lavenu, D. (2005). Evolutions of personal networks and life events. *Social Networks, 27*(4), 359–376.

Blas, C. S., Martin, J. S., & Gonzalez, D. G. (2018). Combined social networks and data envelopment analysis for ranking. *European Journal of Operational Research, 266*(3), 990–999.

Boccaletti, S., Bianconi, G., Criado, R., Del Genio, C. I., Gómez-Gardenes, J., Romance, M. . . . Zanin, M. (2014). The structure and dynamics of multilayer networks. *Physics Reports, 544*(1), 1–122.

Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., & Hwang, D. U. (2006). Complex networks: Structure and dynamics. *Physics Reports, 424*(4–5), 175–308.

Boschet, C., & Rambonilaza, T. (2018). Collaborative environmental governance and transaction costs in partnerships: Evidence from a social network approach to water management in France. *Journal of Environmental Planning and Management, 61*(1), 105–123.

Calastri, C., Hess, S., Daly, A., & Carrasco, J. A. (2017). Does the social context help with understanding and predicting the choice of activity type and duration? An application of the Multiple Discrete-Continuous Nested Extreme Value model to activity diary data. *Transportation Research Part A: Policy and Practice, 104*, 1–20.

Caldarelli, G., Battiston, S., Garlaschelli, D., & Catanzaro, M. (2004). Emergence of complexity in financial networks. In *Complex networks* (pp. 399–423). Berlin; Heidelberg: Springer.

Capra, F. (2004). *The hidden connections: A science for sustainable living.* New York, NY: Anchor.

Carrasco, J. A., Hogan, B., Wellman, B., & Miller, E. J. (2008). Agency in social activity interactions: The role of social networks in time and space. *Tijdschrift voor economische en sociale geografie, 99*(5), 562–583.

Carrasco, J. A., & Miller, E. J. (2005). Socializing with people and not places: Modelling social activities explicitly incorporation social networks. Paper presented at the Conference on Computers in Urban Planning and Urban Management, London.

Carrasco, J. A., & Miller, E. J. (2009). The social dimension in action: A multilevel, personal networks model of social activity frequency between individuals. *Transportation Research Part A: Policy and Practice, 43*(1), 90–104.

Carrasco, J. A., Miller, E. J., & Wellman, B. (2008). How far and with whom do people socialize? Empirical evidence about distance between social network members. *Transportation Research Record, 2076*(1), 114–122.

Castellano, C., Fortunato, S., & Loreto, V. (2009). Statistical physics of social dynamics. *Reviews of Modern Physics, 81*(2), 591.

Chávez, Ó., Carrasco, J. A., & Tudela, A. (2018). Social activity-travel dynamics with core contacts: Evidence from a two-wave personal network data. *Transportation Letters, 10*(6), 333–342.

Cho, E., Myers, S. A., & Leskovec, J. (2011, August). Friendship and mobility: User movement in location-based social networks. In *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Diego, CA (pp. 1082–1090). ACM.

Chua, R. Y. (2018). Innovating at cultural crossroads: How multicultural social networks promote idea flow and creativity. *Journal of Management, 44*(3), 1119–1146.

Dugundji, E. R., & Gulyás, L. (2008). Sociodynamic discrete choice on networks in space: Impacts of agent heterogeneity on emergent outcomes. *Environment and Planning B: Planning and Design, 35*(6), 1028–1054.

Erdős, P., & Rényi, A. (1960). On the evolution of random graphs. *Publication of the Mathematical Institute of the Hungarian Academy of Sciences, 5*(1), 17–60.

Ettema, D., & Kwan, M. P. (2010, July). The influence of social ties on social and recreational activity participation of ethnic groups in the Netherlands. Paper presented at 12th WCTR, July 11–15, 2010, Lisbon, Portugal.

Farber, S., Neutens, T., Carrasco, J. A., & Rojas, C. (2014). Social interaction potential and the spatial distribution of face-to-face social interactions. *Environment and Planning B: Planning and Design, 41*(6), 960–976.

Fortunato, S. (2010). Community detection in graphs. *Physics Reports, 486*(3–5), 75–174.

Foster, C. C., Rapoport, A., & Orwant, C. J. (1963). A study of a large sociogram II: Elimination of free parameters. *Behavioral Science, 8*(1), 56–65.

Freeman, L. C. (1984). Turning a profit from mathematics: The case of social networks. *Journal of Mathematical Sociology, 10*(3–4), 343–360.

Frei, A., & Axhausen, K. W. (2008). Modelling the frequency of contacts in a shrunken world. *Arbeitsberichte Verkehrs-und Raumplanung, 532*.

Garlaschelli, D., Battiston, S., Castri, M., Servedio, V. D., & Caldarelli, G. (2005). The scale-free topology of market investments. *Physica A: Statistical Mechanics and Its Applications, 350*(2–4), 491–499.

Gogia, N. (2006). Unpacking corporeal mobilities: The global voyages of labour and leisure. *Environment and Planning A, 38*(2), 359–375.

Goyal, S., Rosenkranz, S., Weitzel, U., & Buskens, V. (2017). Information acquisition and exchange in social networks. *The Economic Journal, 127*(606), 2302–2331.

Granovetter, M. S. (1977). The strength of weak ties. *Social Networks*, 347–367.

Granovetter, M. S. (1985). Economic action and social structure: The problem of embeddedness. *American Journal of Sociology, 91*(3), 481–510.

Hackney, J. K. (2009). *Integration of social networks in a large-scale travel behavior microsimulation*. Dissertation, Eidgenössische Technische Hochschule ETH Zürich, Nr. 18723.

Hajra, K. B., & Sen, P. (2005). Aging in citation networks. *Physica A: Statistical Mechanics and Its Applications, 346*(1–2), 44–48.

Han, Q., Arentze, T., Timmermans, H., Janssens, D., & Wets, G. (2011). The effects of social networks on choice set dynamics: Results of numerical simulations using an agent-based approach. *Transportation Research Part A: Policy and Practice, 45*(4), 310–322.

Hollstein, B. (2011). Qualitative approaches. *The SAGE Handbook of Social Network Analysis*, 404–416.

Holme, P., & Saramäki, J. (2012). Temporal networks. *Physics Reports, 519*(3), 97–125.

Huang, L., & Knight, A. P. (2017). Resources and relationships in entrepreneurship: An exchange theory of the development and effects of the entrepreneur-investor relationship. *Academy of Management Review, 42*(1), 80–102.

Illenberger, J., Flöteröd, G., Kowald, M., & Nagel, K. (2009). A model for spatially embedded social networks. Paper presented at the 12th International Conference on Travel Behavior Research (IATBR), Jaipur.

Jackson, M. O. (2007). The study of social networks in economics. In J. E. Rauch (Ed.), *The missing links: Formation and decay of economic networks*. New York: Russell Sage Foundation.

Kim, J., Rasouli, S., & Timmermans, H. J. P. (2014a). Expanding scope of hybrid choice models allowing for mixture of social influences and latent attitudes: Application to intended purchase of electric cars. *Transportation Research Part A: Policy and Practice, 69*, 71–85.

Kim, J., Rasouli, S., & Timmermans, H. J. P. (2014b). Hybrid choice models: Principles and recent progress incorporating social influence and nonlinear utility functions. *Procedia Environmental Sciences, 22*, 20–34.

Kim, J., Rasouli, S., & Timmermans, H. J. P. (2018). Social networks, social influence and activity-travel behaviour: A review of models and empirical evidence. *Transport Reviews, 38*(4), 499–523.

Kowald, M., & Axhausen, K. W. (2012). Focusing on connected personal leisure networks: Selected results from a snowball sample. *Environment and Planning-Part A, 44*(5), 1085.

Kowald, M., & Axhausen, K. W. (Eds.). (2015). *Social networks and travel behaviour*. Surrey: Ashgate Publishing Ltd.

Kowald, M., van den Berg, P., Frei, A., Carrasco, J. A., Arentze, T., Axhausen, K. . . . Wellman, B. (2013). Distance patterns of personal networks in four countries: A comparative study. *Journal of Transport Geography, 31*, 236–248.

Lovrić, M., Da Re, R., Vidale, E., Pettenella, D., & Mavsar, R. (2018). Submission of an original research paper: Social network analysis as a tool for the analysis of international trade of wood and non-wood forest products. *Forest Policy and Economics, 86*, 45–66.

Mayer, A. (2009). Online social networks in economics. *Decision Support Systems, 47*(3), 169–184.

McCowan, B., Beisner, B. A., Capitanio, J. P., Jackson, M. E., Cameron, A. N., Seil, S. . . . Fushing, H. (2011). Network stability is a balancing act of personality, power, and conflict dynamics in rhesus macaque societies. *PLoS One, 6*(8), e22350.

Miguéns, J. I. L., & Mendes, J. F. F. (2008). Travel and tourism: Into a complex network. *Physica A: Statistical Mechanics and Its Applications, 387*(12), 2963–2971.

Mitrović, M., & Tadić, B. (2012). Emergence and structure of cybercommunities. In *Handbook of optimization in complex networks* (pp. 209–227). Boston, MA: Springer.

Molin, E., Arentze, T., & Timmermans, H. J. P. (2008). Social activities and travel demand: Model-based analysis of social network data. *Transportation Research Record: Journal of the Transportation Research Board, 2082*, 168–175.

Mollick, E. (2014). The dynamics of crowdfunding: An exploratory study. *Journal of Business Venturing, 29*(1), 1–16.

Montgomery, J. D. (1991). Social networks and labor-market outcomes: Toward an economic analysis. *The American Economic Review, 81*(5), 1408–1418.

Newman, M. E. (2002). Assortative mixing in networks. *Physical Review Letters, 89*(20), 208701.

Orsini, C., Dankulov, M. M., Colomer-de-Simón, P., Jamakovic, A., Mahadevan, P., Vahdat, A. . . . Krioukov, D. (2015). Quantifying randomness in real networks. *Nature Communications, 6*, 8627.

Páez, A., & Scott, D. M. (2007). Social influence on travel behavior: A simulation example of the decision to telecommute. *Environment and Planning A, 39*(3), 647–665.

Páez, A., Scott, D. M., & Volz, E. (2008). A discrete-choice approach to modeling social influence on individual decision making. *Environment and Planning B: Planning and Design, 35*(6), 1055–1069.

Palchykov, V., Mitrović, M., Jo, H. H., Saramäki, J., & Pan, R. K. (2014). Inferring human mobility using communication patterns. *Scientific Reports, 4*, 6174.

Palla, G., Barabási, A. L., & Vicsek, T. (2007). Quantifying social group evolution. *Nature, 446*(7136), 664.

Park, J. Y., Im, I., & Sung, C. S. (2017). Is social networking a waste of time? The impact of social network and knowledge characteristics on job performance. *Knowledge Management Research & Practice, 15*(4), 560–571.

Pastor-Satorras, R., Castellano, C., Van Mieghem, P., & Vespignani, A. (2015). Epidemic processes in complex networks. *Reviews of Modern Physics, 87*(3), 925.

Polzin, F., Toxopeus, H., & Stam, E. (2018). The wisdom of the crowd in funding: Information heterogeneity and social networks of crowdfunders. *Small Business Economics, 50*(2), 251–273.

Rasouli, S., & Timmermans, H. (2014). Activity-based models of travel demand: Promises, progress and prospects. *International Journal of Urban Sciences, 18*(1), 31–60.

Ronald, N., Dignum, V., Jonker, C., Arentze, T., & Timmermans, H. J. P. (2012). On the engineering of agent-based simulations of social activities with social networks. *Information and Software Technology, 54*(6), 625–638.

Sadri, A. M., Ukkusuri, S. V., & Gladwin, H. (2017). Modeling joint evacuation decisions in social networks: The case of Hurricane Sandy. *Journal of Choice Modelling, 25*, 50–60.

Scott, N., Baggio, R., & Cooper, C. (2008). *Network analysis and tourism: From theory to practice.* Clevedon: Channel View Publications.

Sen, P., & Chakrabarti, B. K. (2014). *Sociophysics: An introduction.* Oxford: Oxford University Press.

Shane, S., & Cable, D. (2002). Network ties, reputation, and the financing of new ventures. *Management Science, 48*(3), 364–381.

Sharmeen, F. (2015). *Dynamics of social networks and activity travel behaviour.* Doctoral Dissertation, Technische Universiteit Eindhoven, Eindhoven.

Sharmeen, F., Arentze, T., & Timmermans, H. J. P. (2013). Incorporating time dynamics in activity travel behavior model: A path analysis of changes in activity and travel time allocation in response to life-cycle events. *Transportation Research Record: Journal of the Transportation Research Board, 2382*, 54–62.

Sharmeen, F., Arentze, T., & Timmermans, H. J. P. (2014a). An analysis of the dynamics of activity and travel needs in response to social network evolution and life-cycle events: A structural equation model. *Transportation Research Part A: Policy and Practice, 59*, 159–171.

Sharmeen, F., Arentze, T., & Timmermans, H. J. P. (2014b). Dynamics of face-to-face social interaction frequency: Role of accessibility, urbanization, changes in geographical distance and path dependence. *Journal of Transport Geography, 34*, 211–220.

Sharmeen, F., Arentze, T., & Timmermans, H. J. P. (2015a). Dynamic social networks and travel. In M. Kowald & K. Axhausen (Eds.), *Social networks and travel behaviour* (pp. 203–218). London: Routledge.

Sharmeen, F., Arentze, T., & Timmermans, H. J. P. (2015b). Predicting the evolution of social networks with life cycle events. *Transportation, 42*(5), 733–751.

Sharmeen, F., Chávez, Ó., Carrasco, J. A., Arentze, T., & Tudela, A. (2016). A modelling population-wide personal network dynamics using a two-wave data collection method

and an origin-destination survey. In *TRB 95th Annual Meeting Compendium of Papers.* Washington, DC: Transportation Research Board.

Sharmeen, F., & Sivakumar, A. (2017). Why care about social networks in travel demand forecasting? Testing the predictive power of social attributes in modeling discretionary trip frequencies. In *TRB 96th Annual Meeting Compendium of Papers.* Washington, DC: Transportation Research Board.

Sharmeen, F., & Timmermans, H. J. P. (2014). Walking down the habitual lane: Analyzing path dependence effects of mode choice for social trips. *Journal of Transport Geography, 39*, 222–227.

Sheller, M., & Urry, J. (2006). The new mobilities paradigm. *Environment and Planning A, 38*(2), 207–226.

Silvis, J., Niemeier, D., & D'Souza, R. (2006). Social networks and travel behavior: Report from an integrated travel diary. Paper presented at the 11th International Conference on Travel Behavior Research, Kyoto.

Tibély, G., Onnela, J. P., Saramäki, J., Kaski, K., & Kertész, J. (2006). Spectrum, intensity and coherence in weighted networks of a financial market. *Physica A: Statistical Mechanics and Its Applications, 370*(1), 145–150.

Timur, S., & Getz, D. (2008). A network perspective on managing stakeholders for sustainable urban tourism. *International Journal of Contemporary Hospitality Management, 20*(4), 445–461.

Travers, J., & Milgram, S. (1967). The small world problem. *Psychology Today, 1*(1), 61–67.

Urry, J. (2003). Social networks, travel and talk. *The British Journal of Sociology, 54*(2), 155–175.

Urry, J. (2004). Small worlds and the new 'social physics'. *Global Networks, 4*(2), 109–130.

van den Berg, P., Arentze, T., & Timmermans, H. J. P. (2008). Social networks, ICT use and activity-travel patterns. In *Data collection and first analyses.* Paper presented at the 9th International Conference on Design & Decision Support Systems in Architecture and Urban Planning, The Netherlands.

van den Berg, P., Arentze, T., & Timmermans, H. J. P. (2009). Size and composition of ego-centered social networks and their effect on geographic distance and contact frequency. *Transportation Research Record: Journal of the Transportation Research Board, 2135*, 1–9.

van den Berg, P., Arentze, T., & Timmermans, H. J. P. (2010). Factors influencing the planning of social activities: Empirical analysis of data from social interaction diaries. *Transportation Research Record: Journal of the Transportation Research Board, 2157*, 63–70.

van den Berg, P., Arentze, T., & Timmermans, H. J. P. (2011). Estimating social travel demand of senior citizens in the Netherlands. *Journal of Transport Geography, 19*(2), 323–331.

van den Berg, P., Arentze, T., & Timmermans, H. J. P. (2012). Involvement in clubs or voluntary associations, social networks and activity generation: A path analysis. *Transportation, 39*(4), 843–856.

van den Berg, P., Arentze, T., & Timmermans, H. J. P. (2013). A path analysis of social networks, telecommunication and social activity – travel patterns. *Transportation Research Part C: Emerging Technologies, 26*, 256–268.

van den Berg, P., Weijs-Perrée, M., & Arentze, T. (2017). Dynamics in social activity-travel patterns: Analyzing the role of life-cycle events and path dependence in face-to-face and ICT-mediated social interactions. *Research in Transportation Economics, 68*, 29–37.

Wang, G., & Wang, Y. (2018). Herding, social network and volatility. *Economic Modelling*, *68*, 74–81.

Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications* (Vol. 8). Cambridge: Cambridge University Press.

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature*, *393*(6684), 440.

Wellman, B., Wong, R. Y.-l., Tindall, D., & Nazer, N. (1997). A decade of network change: Turnover, persistence and stability in personal communities. *Social Networks*, *19*(1), 27–50.

2   How to define social network in the context of mobilities

Bridgette Wessels, Sven Kesselring and Pnina O. Plaut

Introduction

Social activities steaming from social relations are one of the main motivators for the use of transport systems. A social network is a social structure, based upon group members and the relations among them. Social network studies are concerned with the structure of sociocultural systems, such as the number of contacts and the levels of communication among different members of a certain social group. The past decade has witnessed rapid communication developments, which had major social impacts. The use of new Information and Communication Technologies accelerated the shift from social groups that were defined through a specific location (e.g. residential neighbourhood or workplace) to individually based social networks. This shift, coined by Wellman (2001, 2002) as "networked individualism", is a stage in which mobile, high-speed telecommunication allows for personalized networks and "person-to-person" social ties. These new social networks are associated with several changes, when comparing to the past 50 years, including people having a larger set of active contacts today than in the past, with wider spatial distribution, the contacts spread across more social networks than in the past and typical social networks less coherent; that is fewer people share multiple affiliations today than in the past, and leisure travel is increasing (Axhausen, **2005**). All of the preceding have an impact on travel and mobility within the urban realm.

In this chapter, we present social networks in the context of mobilities. We argue that traditional approaches to travel and travel behaviour based on ego-centred perspectives do not fully address the way in which travel is embedded within social networks and connections. We argue that senses of mobility are created from the ways in which people meaningfully organize their lives. Mobility is generated through social networks, both personal and impersonal or as Urry (2003) puts it "social networks, travel and talk" are closely and indissolubly intertwined.

Egocentred approaches to the analysis of social networks and travel behaviour

Research on the impact of Information and Communication Technology (ICT) social networks on travel behaviour is still in its infancy (about a decade old).

Mechanisms of self-organized criticality in social processes of knowledge creation

Bosiljka Tadić,[1] Marija Mitrović Dankulov,[2] and Roderick Melnik[3]

[1]*Department of Theoretical Physics, Jožef Stefan Institute, Jamova 39, 1000 Ljubljana, Slovenia*

[2]*Scientific Computing Laboratory, Center for the Study of Complex Systems, Institute of Physics Belgrade,*
University of Belgrade, Pregrevica 118, 11080 Belgrade, Serbia

[3]*MS2Discovery Interdisciplinary Research Institute, M2NeT Laboratory and Department of Mathematics,*
Wilfrid Laurier University, Waterloo, Ontario, Canada, N2L 3C5

In online social dynamics, a robust scale invariance appears as a key feature of collaborative efforts that lead to new social value. The underlying empirical data thus offers a unique opportunity to study the origin of self-organized criticality (SOC) in social systems. In contrast to physical systems in the laboratory, various human attributes of the actors play an essential role in the process along with the contents (cognitive, emotional) of the communicated artifacts. As a prototypical example, we consider the social endeavor of knowledge creation via Questions and Answers (Q&A). Using a large empirical data set from one of such Q&A sites and theoretical modeling, we reveal fundamental characteristics of SOC by investigating the temporal correlations at *all scales* and the role of cognitive contents to the avalanches of the knowledge-creation process. Our analysis shows that the universal social dynamics with power-law inhomogeneities of the actions and delay times provides the primary mechanism for self-tuning towards the critical state; it leads to the long-range correlations and the event clustering in response to the external driving by the arrival of new users. In addition, the involved cognitive contents (systematically annotated in the data and observed in the model) exert important constraints that identify unique classes of the knowledge-creation avalanches. Specifically, besides determining a fine structure of the developing knowledge networks, they affect the values of scaling exponents and the geometry of large avalanches and shape the multifractal spectrum. Furthermore, we find that the level of the activity of the communities that share the knowledge correlates with the fluctuations of the innovation rate, implying that the increase of innovation may serve as the active principle of self-organization. To identify relevant parameters and unravel the role of the network evolution underlying the process in the social system under consideration, we compare the social avalanches to the avalanche sequences occurring in the field-driven physical model of disordered solids, where the factors contributing to the collective dynamics are better understood.

I. INTRODUCTION

In recent years, the self-organized criticality (SOC) is considered as one of the principal mechanisms responsible for the emergence of new features at a larger scale in various complex systems. The transition from the microscopic interactions to the collective behavior involves nonlinear dynamical phenomena when the system is driven out of equilibrium (for an overview of physical systems exhibiting SOC, see recent reviews in [1–3] and the references there). In this context, SOC refers to the dynamical self-organization among the interacting units in response to repeatedly applied infinitesimal driving; the system's adaptation to the driving force leads to robust metastable states with system-wide correlations, fractal dynamics, and avalanches as the key signatures of criticality [2–6]. In this context, an avalanche is recognized as a mesoscopic dynamical structure consisting of a sequence of connected elementary events (a precise definition is given in Sec. II). It has been newly pointed out that SOC plays a role in the functioning of biological [7] and diverse other complex systems from neuronal dynamics [8,9] to animal behavior [10] and human history [11]. For instance, the analysis of vast amounts of the available brain imaging data and theoretical modelings provided the evidence that supports SOC as an underlying mechanism of the brain functional stability [8,12].

Although the avalanching behavior and other signatures of criticality are readily observable in the empirical data of online social dynamics [13–19], much less attention has been devoted to understanding the origin and the precise role of SOC in social systems. The key open question is whether the social avalanches represent a unique class of self-organized phenomena or, otherwise, they can be reduced to standard models of physical SOC systems, describing the transition from the microscopic interactions to the observed complex spatiotemporal patterns. Another interesting aspect of the problem concerns the interplay of the coevolving network structure and the social dynamics that it supports. The question whether the SOC process shapes the structure, or the network evolution enables the self-tuning towards the criticality remains open. Here, we address these issues by analysis of the empirical data of knowledge-creation social endeavors and using theoretical modeling.

In physics, striking examples of the multiscale dynamics characterizing SOC are observed in the turbulent flow [2,20–22] and the kinetics of earthquakes [23,24]. The signatures of SOC are also found in experiments with stressed granular materials [25,26], driven disordered systems at a hysteresis loop [27–30], and porous shape-memory alloys [31,32]. Furthermore, the avalanching dynamics is characteristic of the conduction in the assembled networks of living neurons in a solvent [33] and nanoparticle films with single-electron tunnelings conduction [34], as well as to the motion of topological objects, such as vortices [35,36] and domain walls [37–39]. The theoretical concepts were developed to describe the emergence of collective behaviors

from the microscopic interactions among many constitutive elements both in the classical and quantum systems [40]. In this regard, often paradigmatic models were used as the instruments for investigations. Further, the use of scaling and renormalization group ideas provided a better understanding of the role of different scales in the dynamical systems driven away from the equilibrium [41–43]. The precise description of the interactions in these physical systems allows investigating the microscopic mechanisms responsible for the triggering and propagation of an avalanche. The standard feature of all SOC systems is the accumulation of free energy, which then dissipates through the avalanches. While the energy source and triggering mechanisms are physics specific, common to all SOC systems is that the avalanching response is not in proportion to the forcing. Consequently, the power-law distributions of the avalanches appear as one of the key features of SOC states. It has been recognized that the propagation of avalanches involves three phases [2]: the initial growth phase is supported by multiplicative chain reactions until the maximum dissipation is reached in the peak period, after which the activity is reduced and eventually diminishes, in the stopping phase. Often, the essence of SOC can be captured by the elementary dynamics of sand-pile automata [4,5] on a two-dimensional lattice. In real systems, however, the presence of many physical parameters that can influence the dynamics makes it difficult to distinguish the potential SOC states (attractor with a large basin of attraction) from the dynamical phase transition, which occurs by fine tuning of a relevant parameter.

In this work, we investigate the nature of avalanches in the prototypical human collaborative endeavor of knowledge creation [44]. In this process, the knowledge and expertise of individual actors are transferred into a social value [45,46]—the collective knowledge, which is shared by all participants in the process. In contrast to the physical systems in the laboratory, the human cooperation, as well as the new collective states, are evident [44,47]. Therefore, the empirical data on these social systems represent a valuable source to investigate the origin of self-organized criticality. On the other hand, certain attributes of the human participants are crucial to the social cooperation; they remain elusive to the accurate theoretical modeling of the interactions, which underlie the avalanche formation. In particular, the process of knowledge creation requires the appropriate expertise of the participating actors among other human attributes. Thus, the subdynamics representing the use of the communicated cognitive contents tends to constrain the social process itself. In this regard, it remains unclear how these different aspects (social and cognitive) of the dynamics contribute to the appearance and propagation of the avalanches. To address this question, we combine the analysis of the empirical data from Questions and Answers (Q&A) site Mathematics Stack Exchange (https://math.stackexchange.com/) with the agent-directed modeling; the cognitive contents of each artifact are systematically encoded in the considered data sets. The agent's attributes are statistically similar to the users in the data, while their expertise is varied. The system is driven by the arrival of new agents. Our analysis reveals that the occurrence of avalanches is a robust social phenomenon, whereas their fine structure, geometrical, and fractal characteristics are

affected by the distribution of the expertise over the actors. Furthermore, the interplay between the social and knowledge processes is fueled by the constant tendency towards the expansion of innovation. For a deeper understanding of the potential mechanisms, we use a comparison with the better controlled avalanching dynamics in physical systems. For this purpose, we analyze the model of a disordered system of the interacting spins, where the critical states at the hysteresis loop appear in the interplay between the driving by the applied magnetic field and the domain-wall pinning along the implanted magnetic (soft) and structural (hard) defects. Although an analogy between spin alignments can be extended to knowledge matching among the social subjects, our objective here is different. We compare the fractal features of the avalanche sequences in both systems, which appear to be similar in a particular range of parameters of the physical system. These comparisons permit us to identify certain factors of the social dynamics that are essential to the appearance of the collective state and can motivate further research towards a viable modeling of the social self-organization.

In the following Sec. II, we introduce the essential characteristics of the processes of Q&A and describe details of the agent-based model and the structure of the bipartite network that coevolves with the social interaction. Then Sec. III presents a detailed analysis of the knowledge-bearing avalanches both from the empirical data and simulations. The simulations and analysis of the avalanches in the driven spin system are given in Sec. IV. A summary of the results and the discussion are given in Sec. V.

II. THE STOCHASTIC PROCESS OF KNOWLEDGE CREATION AND THE COEVOLVING NETWORKS

The knowledge creation via Questions & Answers is a collaborative social endeavor, in which the knowledge of each participant is shared with others. By its nature, the interaction between these participants is indirect, mediated by questions and answers, in a way similar to user interactions via posted texts on blogs [13,48,49]. Thus, the environment of the knowledge sharing can be represented as a coevolving *bipartite network* with the *actors* (users, agents) as one partition, and the *artifacts* (questions, answers) as the other partition [44]. In epistemology, to create a common value (knowledge), *meaningful social interactions* are required, in which the actor's response is adequate to the needs of others [45,46]. Specifically, the actor possessing an expertise can meaningfully act on the artifact where this particular expertise is required. Thus, the essence of the dynamics is the contents-matching rule, as schematically illustrated in Fig. 1(a). A part of the evolving network extracted from the empirical data is also shown in Fig. 1(b). Hence, the knowledge-creation process consists of two mutually interconnected factors: the social dynamics and the constrained use of the cognitive contents. The strict use of the available expertise in the knowledge-creation processes is in marked contrast with the informal social communications on blogs and similar systems, where the user's natural interests and emotions drive the activity [13].

In the data from the analyzed Q&A site, the cognitive content of each artifact is encoded by up to five tags, according
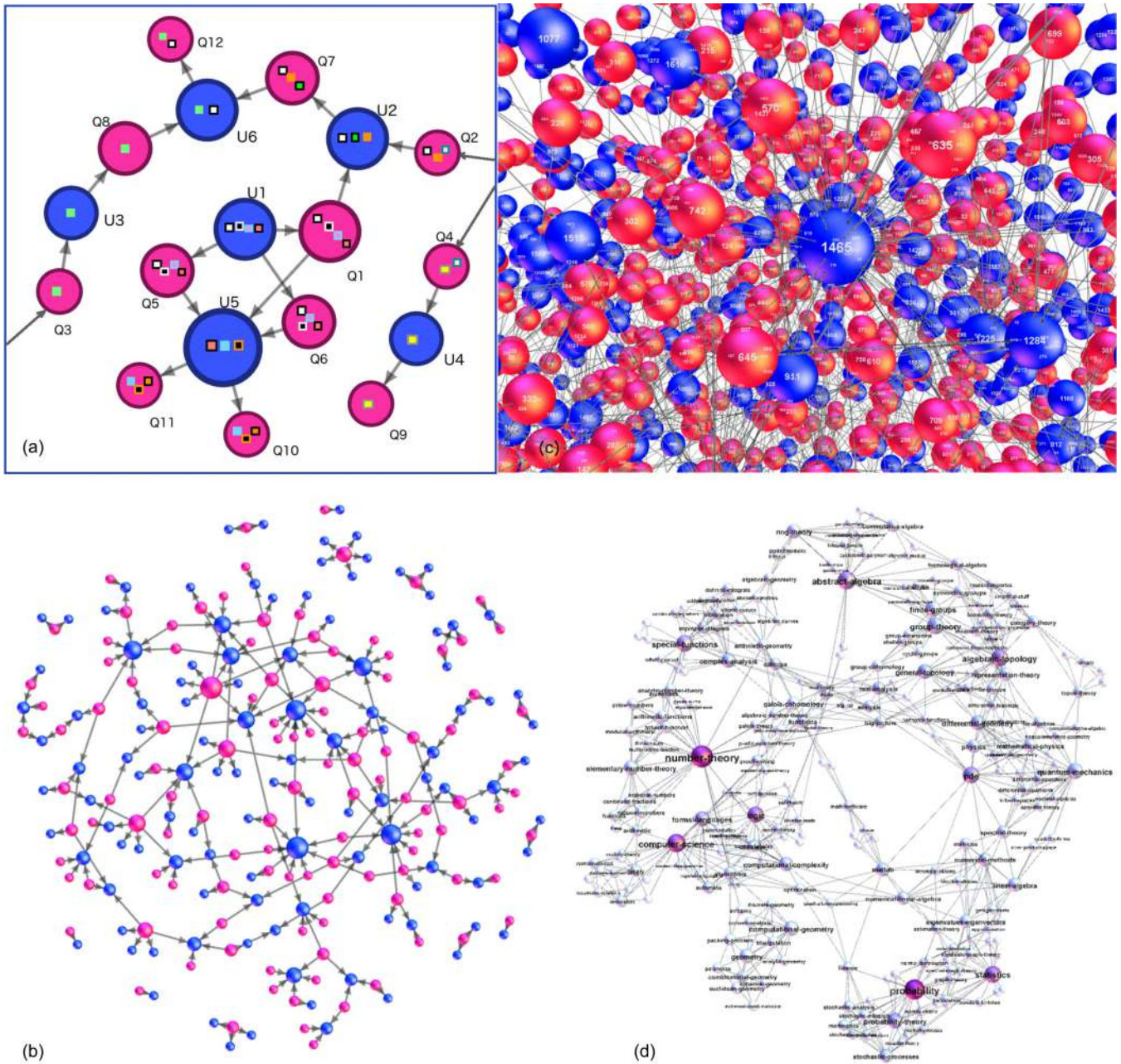
FIG. 1. (a) Schematic view of the knowledge exchange via Q&A on a bipartite network of actors (blue nodes) and questions and answers (red nodes); boxes of different colors represent the cognitive contents (tags). A directed link from question to actor node indicates activity on that question, where the matching (at least in one tag) between the actor's expertise and the contents of the question is required; the outgoing link represents posting the question or answer, whose content contains the actor's knowledge. Panels (b),(c) show two bipartite networks extracted from the empirical data; the user nodes (blue) connect to question nodes (red) which compress all existing answers to that question, and the direction of the link indicates the question on which the user was active either by posting or answering it. Specifically, (b) the innovation layer in the evolving bipartite network extracted at the end of year one, consists of the recently active questions and the users whose activity on these questions occurred in the last $T_0 = 100$ min and the nodes to which they were connected within the time depth of $6T_0$. (c) A close up of the compressed bipartite network of users and the questions filtered such to contain the tag "Linear Algebra" among other tags; the network represents the activity within the first two months of the considered empirical data. (d) The explicit-knowledge network containing the innovation tags attached to the tags of year one.

to the standard mathematical classification scheme (MCS), for example, "Graph Theory," "Probability," "Stochastic Processes," "Linear Algebra," "Algebraic Topology," "Differential Geometry," and others. Whereas, the information about the user's expertise participating in the process can be inferred

statistically as described in methods of [44]; see the inset to Fig. 2(a). This figure suggests a rather broad distribution of the expertise (i.e., different number of tags 2^{E_i}) over the users in the native system. The main Fig. 2(a) shows another key characteristic of the experimental system: the
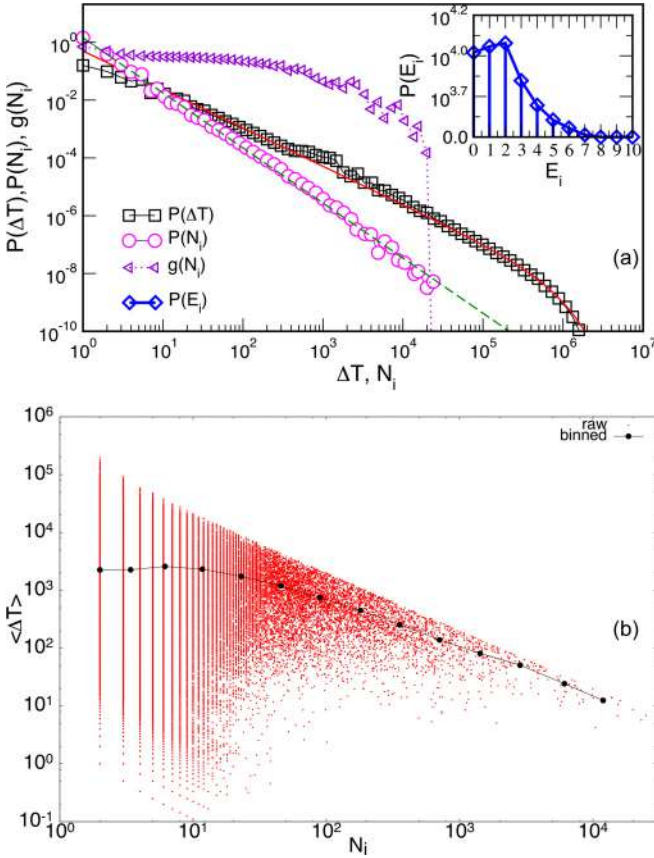
FIG. 2. (a) The distribution $P(N_i)$ of the number of actions N_i per user, the dependence $g(N_i)$, and the delay time $P(\Delta T)$ of the users in the empirical data set. Inset: The distribution of entropy $P(E_i)$ based on the probability of the users' expertise; the data are extracted from the empirical set and used for designing the attributes of the agents. (b) Scatter plot of the user-averaged delay times $\langle \Delta T \rangle$ against the number of actions N_i.

user's heterogeneity in the number of actions N_i and a broad range of their interactivity times ΔT. While the scaling exponents indicated by these histograms are system specific, the prominent power-law decay of both quantities manifests the universally observed characteristics of the human behavior online. Further patterns of the user's activity can be determined from the time stamp. Notably, a significant part of the actions of the present users is directed towards the issues posted by new arrivals. Occasionally, an active user looks for an older question with currently searched contents and brings it to the view of others. The ratio of the posted versus answered questions $g(N_i)$ was found to depend on the number of actions of a user; cf. Fig. 2(a). The very active users are in the minority; they often respond to the questions. On the other side, the majority of the least active users are engaged in posting questions and, by getting the satisfactory response, they disappear for a longer period. Note that a similar dependence of the delay times on the number of actions applies to the user-averaged delays; cf. Fig. 2(b), but not to each particular user. Certain regularities may exist for groups of users of a similar activity level [16], but they are not of interest in the present context. The arrival of new users (referred to the beginning of the data set) captured by the *time series* $p(t)$

represents a stochastic process, which depends on the user's off-line life. Here, we adopt the interval of 10 min as a suitable time step for the pace of the activity in the system and use the empirical time series $p(t)$ and its randomized version, as described in the following text, to create the number of agents per time step in the simulations (on average less than one agent per step). Then the reference time depth is $T_0 = 10$.

Observing the requirements for the minimal agent-based model (ABM) of the Web users [50], we have introduced a model where the action rules and the attributes of the agents are taken from these empirical distributions, whereas their expertise can be varied [44]. Here, we briefly describe the main features of the model which is used for the simulations in this work. In particular, in each time step:

(i) *Agents are created*. The number $p(t)$ of new agents is created; their profiles are defined by the number of actions $N_i \in P(N_i)$, the ratio of the posted vs answered questions $g(N_i)$, and the expertise (according to the selected distribution); the new agents are placed on the active list;

(ii) *The agent's action performed*. Each agent from the active list either posts a new question or selects one from the list of recently considered artifacts to act on it. The artifacts that are connected to the agent's network neighborhood are looked at first; with a fixed probability (0.5) the agent also finds a related question in the whole network, thus bringing it to the currently active context. In each case, the expertise matching rule applies.

(iii) *The active questions and network are updated*. The list of active questions within time depth T_0 is maintained, and the network connections are updated according to the executed actions; the agents linked to the questions on which the activity occurred within the previous three steps are prompted for a new action.

(iv) *New delay times are determined*. An agent gets a new delay time $\Delta t \in P(\Delta t)$ after every completed action or after being prompted for a new action.

(v) *The status of each agent is updated*. The number of actions of all agents is updated according to their activity, and the agents, whose number of steps reached the predefined N_i are removed; the delay times of each remaining agent is updated, and each agent whose delay since the previous event expired is placed on the active agents' list.

The expertise of an agent is a set of tags taken randomly from the list of 32 tags. The considered distributions of the expertise are $Exp1$ and $Exp3$, corresponding to single-tag and three-tag expertises, respectively, and a broad range of the expertise $ExpS$, according to the empirical distribution shown in the inset in Fig. 2. For a comparison, we also consider a situation (μ process) where, instead of the actions described above in step (ii), an agent finds an artifact in the entire system and acts on it with a fixed probability (0.25) while disregarding the expertise matching rule. Note that, in this case, the expertise matching can occur by chance; the agent's expertise is taken from the distribution $ExpS$.

Considering the evolution of the system in [44], we have shown that the process is characterized by the innovation growth with the number of events. In this context, the innovation is suitably defined as the number of new combinations of tags. The innovation is introduced into the system by new arrivals and the actions of the other agents through adding their

expertise to the accumulating contents on different artifacts. The innovation growth was observed both in the empirical data as well as in the simulations [44]. It applies to various distributions of the expertise, excluding the case where all actors possess a strictly single-tag expertise. In this limiting case, the tag-matching rule prevents interlinking with other contents, thus leading to isolated communities that share the same tag. In contrast, the innovation grows whenever at least a single actor possesses a combination of different contents in its expertise. The illustration in Fig. 1(a), for instance, indicates how the green tag of the artifact Q3 remains detached until the actor's U6 expertise combines it with the contents of Q7 and the actor U2 who posted it. The observed pace of the innovation growth depends on the distribution of the expertise (by the fixed activity patterns of the actors) [44]. It is important to stress that, in the empirical data, the combination of tags in the expertise of each user already possesses a logical structure of mathematical knowledge. Hence, through these meaningful interactions that structure is preserved during the process. Consequently, the developing network of the used contents also exhibits the logical structure, as shown in [47] by the community detection in the corresponding networks of tags; an example of such an explicit knowledge network extracted from the same data is shown in Fig. 1(d).

A. Growth of the bipartite networks by adding the innovation layers

The interplay between the network structure and the stochastic processes taking part on it consists of the central problem in understanding the networks evolution and their applications in various fields [51–53]. Typically, the graph architecture represents geometrical constraints that shape the diffusionlike processes, likely to cause an anomalous diffusion. For example, the superdiffusion of the information packets [54] occurs on the correlated scale-free network when the traffic rules appropriately utilize the underlying structure. Some unusual situations arise when the structure evolves at the same pace as the SOC avalanching process on it. The random rewiring during the steps of the SOC dynamics has been shown to reduce the avalanche cutoffs [55], thus preventing a catastrophic event to occur. When the rewirings are strictly confined to the current avalanche area, the network appears to have the scale-free degree distribution, where the scaling exponent coincides with one of the avalanche size distributions [56]. Other growth models that apply thresholdlike constraints inspired by SOC dynamics may yield the nonextensive features and scale freeness [57]. In bipartite networks, however, each partition plays a different role in the process, which leads to a more complicated structure–dynamics interplay. As mentioned above, these types of networks often appear in the social dynamics on websites which maintain indirect communication between the users. The appropriate analysis of the artifacts mediating the users revealed [13,44,48,49] how their emotional or cognitive contents affect the network from the node's degree to the mesoscopic community structure.

In the Q&A data set that we consider here, the network growth as well as the pattern of activity of each user and the targeted questions can be extracted from the time stamp in the data. Moreover, to visualize the bipartite networks, we
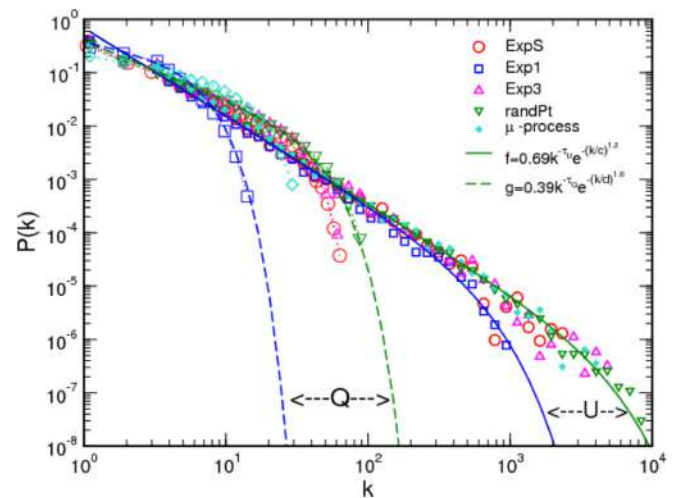

FIG. 3. The degree distribution of the actors and question nodes in the bipartite networks for different expertise and driving indicated in the legend. The power law fits with the exponent $\tau_U = 1.58 \pm 0.07$ and the stretching cutoffs in the range $c \in [508,3920]$ between two solid lines marked by $\leftarrow U \rightarrow$ are of the actor's degree. The distributions for the corresponding question nodes (indicated by the same type of symbol) have smaller cutoffs $d \in [5.8,42]$ between two dashed lines marked by $\leftarrow Q \rightarrow$ and the exponent $\tau_Q = 1.09 \pm 0.1$. In the case of $Exp1$ the exponents $\tau_Q \sim 0.5 \pm 0.16$ apply in a very narrow range, while the exponential distribution fits the data for the μ process.

introduce a compressed node which includes the question and all answers related to that question. A particular example of such compressed bipartite network from the empirical data is shown in Fig. 1(c). The corresponding networks from the simulated data exhibit the mesoscopic structure. The structure of communities sharing the emerging knowledge crucially depends on how the expertise is distributed over the involved participants, as it was shown in [44]. Here, we are interested in the statistical properties of nodes in these bipartite networks. The results of the degree distributions are shown in Fig. 3. They can be fitted by the power law with stretched-exponential cutoffs, which allows a comparison of the user properties in Fig. 2(a). Statistically, the degree distribution of the agent's nodes in these networks follows the slope of the predefined number of actions $P(N_i)$, as expected, while their cutoffs depend on the expertise of the agents. It is also interesting to point out the constraints due to the power-law decay of the delay time distribution; according to some recent studies [58], it can contribute to the convergence towards scale invariance in the growing systems. In the case of the question partition, the cases where the expertise-matching dominates exhibit a similar law but in a reduced range. In the meantime, the power-law behavior is reduced and the cutoffs dominate in the case $Exp1$ and μ process (see Fig. 3).

For the purpose of this work, it is interesting to recognize the *innovation growth layer*, see Fig. 1(b), as the segment of the growing bipartite network where the most recent activity occurs. As stated above, the new arrivals potentially bring the new combinations of the knowledge contents in their expertise, which is expressed in the questions and answers. The set of currently active artifacts are posted or answered within a
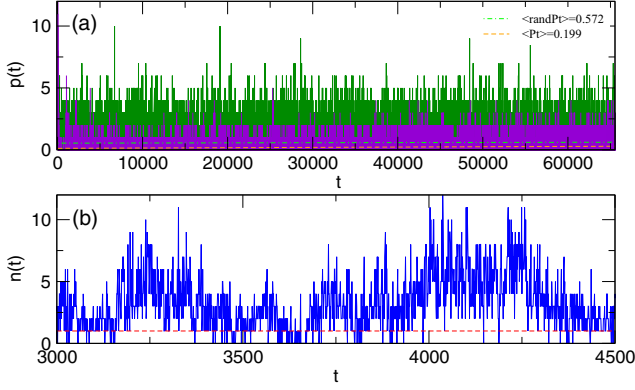
FIG. 4. (a) The driving signals that are used in simulations: $p(t)$, front curve, and its randomized version *randPt*, back curve. (b) Part of the activity time series from the empirical data, exhibiting some large and small avalanches above the baseline (dashed line).

relatively small time window (T_0). These nodes often occur at the outer layer of the network; see Fig. 1(b). Then the currently active users connect to these new artifacts while obeying the expertise matching rule; thus, they connect the new contents to the issues of their previous activity and further through the network of the connected users and their artifacts in a given time depth. Therefore, the recently added artifacts connect the in-depth network via the active users and their previously established connections. Considering a particular time depth, for instance, $6T_0$, with the window $T_0 = 100$ min, we focus on the currently active layer of the network. This layered growth of the bipartite network is fundamentally conditioned by the nature of online human communications, where the latest posts appear on the top. Besides, at each event, an artifact older than the considered time depth is searched with a small probability and connected to the currently active matching contents. The updated active layer then serves as an environment where the next arrivals often attach to, and so on. In the context of open dynamical systems, the addition of new users (or agents) and their artifacts can be seen as the *driving mode* of this bipartite networked system.

B. Extracting the time series and avalanches from the sequence of events

Our focus in this work is on the avalanching behavior which occurs as the network's response to the driving. Therefore, from the sequence of events in the empirical data or the simulated events, we first construct the corresponding time series that capture the fluctuations of the system's activity over time. The possible occurrence of the clustering of events along these time series is a signature of the avalanching dynamics. For the illustration, an example of the time series with the avalanches is shown in Fig. 4(b). Specifically, an avalanche is identified as a segment of the time series consisting of the data points $n(t)$ between two consecutive drops of the signal to the baseline, which is set above a zero or the noise level. To define the baseline, we use a standard approach, as described in [27] for the experimental Barkhausen noise signal. When the signal contains an (extrinsic) noise, the baseline is first put as a horizontal line with a maximum number of intersects
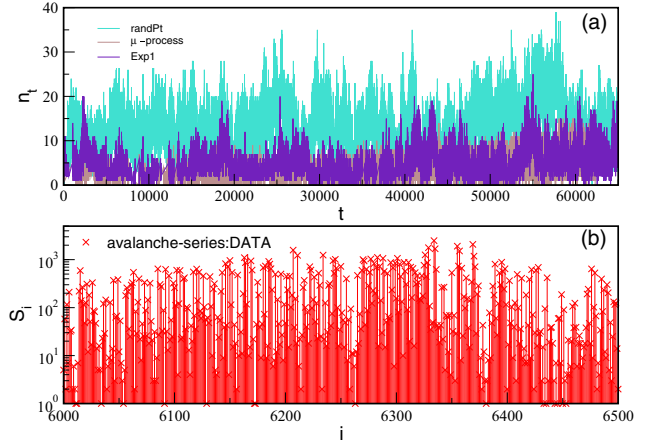
FIG. 5. (a) Examples of time series of the activity simulated by the agent-based model for the different expertise of the agents and μ processes. (b) Sequence of the avalanches determined in the empirical data set—a closeup in the segments with large activity.

with the signal, then the part below the line is considered as noise. The standard deviation σ of the noise is then computed, and the baseline is shifted upward by the distance σ. In the simulated quasistatic driving, Sec. IV, the signal drops to the zero level before next driving event occurs, making the zero level a natural baseline (see also the discussion below). The two intersections of the signal with the baseline are recognized as the beginning t_b and the end t_e time of the avalanche. Then the avalanche size s is given by the sum of all data points between the marked beginning and the end of the avalanche while the distance between these two points along the time axis defines the avalanche duration T, i.e.,

$$s = \sum_{t=t_b}^{t_e} n(t); \quad T = t_e - t_b. \quad (1)$$

The representative examples of the time series studied in this work are shown in Figs. 4 and 5(a). As the example in Fig. 4 shows, the avalanches in the considered stochastic process of knowledge creation differ in size, duration, and shape, closely reflecting the way that the activity propagates in the network. Moreover, a massive avalanche may follow immediately after a small one and vice versa. See also a closeup of the avalanche sequence derived from the empirical data in Fig. 5(b).

It is important to notice that the time series studied in Sec. III are fractal; hence, the avalanches defined through Eq. (1) possess self-affinity. This implies that, by changing a linear scale ℓ, both the segment of the baseline along the x axis, i.e., the avalanche duration T, and the corresponding area above the baseline in the y direction, representing the avalanche size s, scale with *different exponents* such that the respective dimensionless quantities s/ℓ^D and T/ℓ^z remain unchanged. Consequently, the distributions of the avalanche sizes and durations then obey the following scale invariance [1,2]

$$P(s,\ell) = s^{\tau_s} \mathcal{S}(s/\ell^D); \quad P(T,\ell) = T^{\tau_T} \mathcal{T}(T/\ell^z), \quad (2)$$

where τ_s and τ_T are the scaling exponents of the avalanche size and duration, respectively, and D and z are the corresponding

fractal exponents. Furthermore, for such self-affine objects, the following relation holds between the size of avalanches which have a given duration T and the duration T

$$\langle s \rangle_T \equiv \sum_{t=t_b}^{t_b+T} n(t) \sim T^{\gamma_{ST}}, \qquad (3)$$

with the scaling relation $\gamma_{ST} = (\tau_T - 1)/(\tau_s - 1)$. This property of the analyzed avalanches, which we prove in Sec. III, further implies that the scaling exponents are not sensitive to the exact position of the baseline, which discriminates the noise level. Note that we use the initial part of the empirical time series comprising of 65 536 data points, where these methods can be safely applied. In the next section, we quantitatively study these features of the avalanches by considering the time series from both the empirical data and the simulated events.

Figure 5(a) panel shows several examples of the time series of the number of event $n(t)$. Specifically, the time series indicated by $Exp1$ corresponds to the case when each agent possesses a single-tag expertise and these agents are added with the pace $p(t)$ as the new users appear in the real system. The time series marked by $randPt$ is the system of agents with the distribution of the expertise $ExpS$ taken from the empirical data and driven by the randomly shuffled signal $p(t)$ signal. Apart from a few high values at the start, the signal $p(t)$ exhibits an increasing trend, which induces larger activity at later times. By randomizing the time series, however, these larger values may occur randomly along the time axis; consequently, the initial part of the signal which is used for simulations appears to be higher than the original $p(t)$; cf. Fig. 4(a). Hence, the underlying network grows faster when the system is driven by $randPt$; on average, the number of added agents putting their artifacts per 1000 steps is 572, compared to 199 in the case of the original signal $p(t)$. In this way, this accelerated network growth mimics a larger driving rate in the context of SOC systems. The simulated data consist of 65 536 steps, corresponding to the first 15 months of the real system time, where we find that 13 045 users were active, posting 21 998 questions and 179 537 answers. The bottom panel of the same figure shows the avalanche sequence determined from the empirical data set.

III. THE STRUCTURE OF AVALANCHES IN KNOWLEDGE-CREATION PROCESSES

The use of fractal geometry and nonlinear analysis of time series has advanced the understanding of complex systems. Here, we employ the detrended multifractal analysis to study the time series of events as well as the sequences of the avalanches (clustered events) for various model parameters and the empirical data of the knowledge-creation processes. By investigating these complex time series at all scales, we aim to reveal a fine structure of the underlying SOC states of the system.

A. Temporal correlations and avalanche sizes

The occurrence of avalanches in composite signals is not accidental but built on the temporal correlations at a larger scale. These correlations are manifested in the corresponding power spectrum as a power-law decay $W(\nu) \sim \nu^{-\phi}$, for a broad
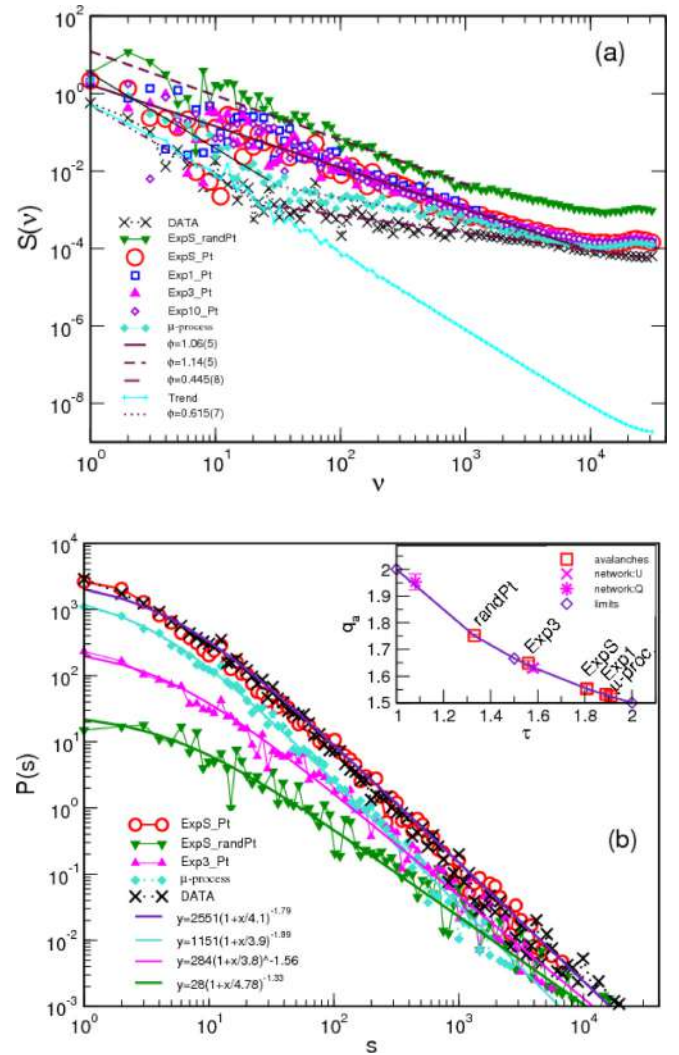
FIG. 6. The (a) power spectra and (b) avalanche size distributions in the empirical data and ABM for different expertise. Inset: The nonextensivity parameter q_a vs τ_s for the considered cases.

range of frequencies ν. In Fig. 6(a), we show the results for the power spectrum of the time series of events in the empirical data and the simulated signals for different agent's expertise and two driving modes. The corresponding distributions of the avalanche sizes obtained from these time series are shown in Fig. 6(b). These figures indicate that an extended scaling range occurs over several orders of magnitude in the power spectrum as well as in the avalanche sizes. However, the driving mode and the actor's expertise affect the scale invariance in a different manner. Specifically, the increasing trend in the driving signal $p(t)$ is pronounced in the power spectrum of the empirical data and μ process. Whereas, the increased activity in the innovation layer is balanced by the strict expertise-matching, resulting in the correlations of the flicker-noise type (middle curves). When the driving rate is elevated, i.e., $randPt$ case, the slope ϕ increases in the region of low frequencies and decreases in the high-frequency region (top curve).

While the shape of the driving signal is essential for the scaling in the power spectrum, the scale invariance of the avalanches is equally sensitive to the actual expertise of
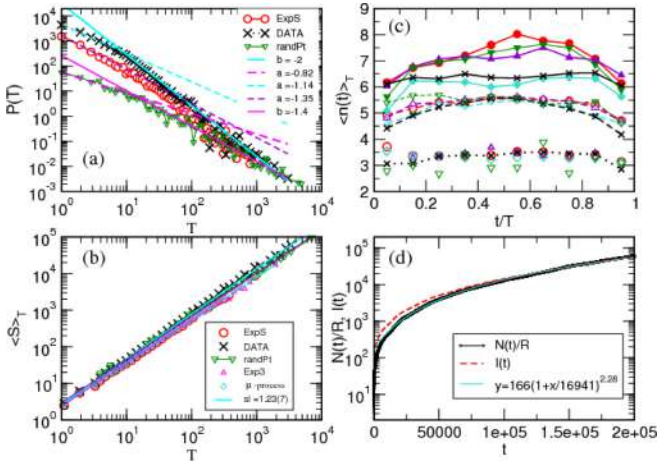
FIG. 7. (a) The distribution of avalanche duration and (b) the average size $\langle S \rangle_T$ of the avalanches of a fixed duration T plotted against T for the empirical data and simulations with varied expertise and driving rate, as indicated. (c) The average activity $\langle n(t) \rangle_T$ within an avalanche of the duration T averaged over the avalanches in three ranges: $T \leqslant 10$ (half-filled symbols, dotted line), $T \in (10, 100]$ (open symbols, dashed lines), and $T > 100$ (filled symbols, full lines). The symbol shape and color correspond to the legend in the panel (b). (d) Extracted from the empirical data, the innovation $I(t)$ increase with time and the integrated activity time series $N(t) = \sum_t n(t)$, normalized by the ratio $R = \langle n(t) \rangle / \langle dI(t)/dt \rangle = 32.5$.

the agents. The slopes of the distribution of the avalanche sizes τ_s are found in the range from 1.33 to 1.93, depending on the expertise and driving; note that the degree distributions of user and question partitions in Fig. 3 are within this range (see also Discussion). However, the mathematical expressions that fit these distributions are different. In particular, the distribution of avalanche sizes, shown in Fig. 6(b) can be fitted by the q_a-exponential function

$$P(S) = A[1 - (1 - q_a)S/S_0]^{1/1-q_a} , \qquad (4)$$

where the parameter $q_a > 1$ measures the degree of nonextensivity in the underlying stochastic process [40,59–62]. The values of the scaling exponents of the avalanche size distribution and the corresponding values of the nonextensivity parameter q_a are shown in the inset in Fig. 6(b). Notably, the distribution of the avalanche sizes obtained from the empirical data and simulations with the expertise of the agents $ExpS$ taken from the empirical distribution are similar and close to the case of μ process. The probability of large avalanches increases with the increased average expertise; the case $Exp3$ is shown. It is important to stress that the same type of distribution with a power-law tail is also obtained in the case of $randPt$, representing an increased driving rate with the accelerated network growth, as mentioned above. Moreover, the exponent of the distribution of avalanche sizes is smaller than when the lower driving rate pertinent to $p(t)$ signal is used; see also Fig. 7(a) for the distribution of duration. The decrease of the scaling exponents provides an observable measure of the effects of avalanche merging, which occurs more often at the elevated driving rate, to the critical state. These findings are in agreement with the studies of SOC in cellular

automata [63] and physical systems [64] under various driving rates.

B. Propagation and geometry of avalanches

Statistics of the avalanche sizes with power-law tails, as shown in Fig. 6(b), is compatible with the occurrence of self-organized dynamics. In the following, we show that the propagation of avalanches and their shapes further confirm these features of the underlying dynamics. The time-dependent characteristics of the avalanches evaluated for the above-studied sets of parameters are demonstrated in Figs. 7(a)–7(c). Specifically, the distribution of the duration T of avalanches, cf. Fig. 7(a), exhibits different scaling behavior for small avalanches with the duration $T < T_x \sim 10$ as compared to the asymptotic scaling law $P(T) \sim T^b$. In the asymptotic region, we find that (within the numerical error bars) $b = -2$ for the avalanche durations in the empirical data and a close value $b = -1.96$ for the simulated system with $ExpS$. Gradually smaller slopes are found for the cases $Exp3$ (not shown) and $randPt$. Similarly, the scaling exponents in the short-avalanche region vary with the expertise and the driving rate from $a = -0.82$ for $randPt$ to $a = -1.35$ for $ExpS$. The corresponding range for the empirical data is even shorter; see Fig. 7(a). Given different shapes, the size of the avalanches of a fixed duration can vary; cf. Fig 4(b) and Eq. (1). Nevertheless, in the SOC systems, the average size $\langle S \rangle_T$ of the avalanches of a given duration T scales with T according to Eq. (3). Here, the tail exponent b of the duration distribution appears, i.e., $\langle S \rangle_T \sim T^{\gamma_{ST}}$, where the exponent $\gamma_{ST} = (b-1)/(\tau_s - 1)$. Figure 7(b) exhibits the plots $\langle S \rangle_T$ against T corresponding to the avalanches studied in this work. The apparent power-law dependence in this plot suggests that the scale invariance of the avalanches can not be affected by a reasonable shift of the baseline, for instance, by 2σ. The average exponent $\gamma_{ST} = 1.23 \pm 0.07$ suggests rather narrow avalanches; apart from the duration range, which varies with the simulation parameters, the variations of the exponent are rather small.

The precise shape of the avalanche of the duration T is given by the sequence of the elementary pulses $n(t)$ over time. In Fig. 7(c), we show the average height $\langle n(t) \rangle_T$ belonging to the avalanche of a given duration T evaluated in bins of the reduced time t/T. Three groups of avalanches are distinguished, in particular, the small avalanches of the durations $T \leqslant 10$, medium-duration $10 < T \leqslant 100$, and large avalanches for the durations $T > 100$. As the Fig. 7(c) shows, the shape of the small avalanches is practically independent of the system's parameters. The same conclusion applies for the medium-size avalanches in the peak region, whereas they slightly differ in the decaying phase and even more in the raising phase. In the case of the large avalanches, however, the major differences occur in the peak phase. Moreover, the peak shifts towards later times when the total expertise is increased, or a larger driving rate applied.

In the panel (d) of Fig. 7, we show how the total innovation $I(t)$ increases over time. The innovation, which is precisely defined as the number of unique combinations of tags, is obviously related to the activity but its fluctuations over time is a more subtle feature of the process, which depends on the available expertise of the agents, as shown in [44]. Here,
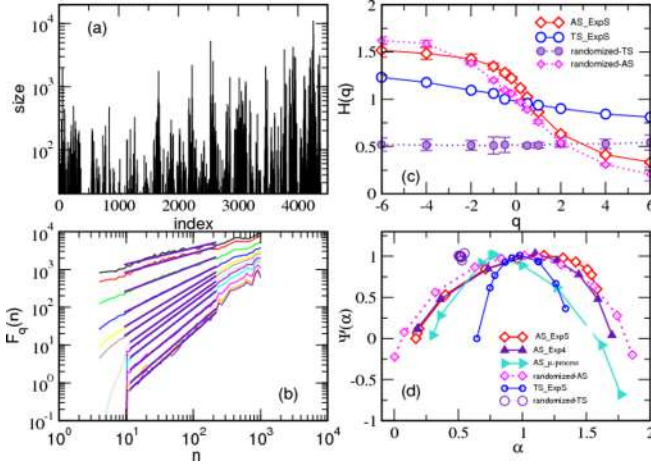
FIG. 8. For the case $ExpS$, (a) the avalanche sequence, (b) the fluctuation function of the avalanche size series, and (c) the generalized Hurst exponent $H(q)$ for the avalanche series and the underlying time series and their randomized versions. The panel (d) shows the singularity multifractal spectrum for the avalanche sequences determined for the cases with the varied expertise and driving, as indicated in the legend.

we compute the temporal dependence of the innovation from the empirical data. The total activity (scaled by a constant factor) as a function of time is also shown in Fig. 7(d). These results suggest that, apart from the excess of the innovation in the initial period, the asymptotic law is the same for the total activity and the innovation growth with time. While the average activity is 32.5 times larger, we can conclude that the activity is driven by the fluctuations of the innovation rate. As mentioned earlier, the innovation is brought by the expertise of the actors. Hence, the increase of the innovation can be considered as a driving force for the knowledge-creation processes.

C. Avalanche sequences and multifractality

Compared to the time series of the number of events $n(t)$, the avalanches are the objects occurring at a mesoscopic scale. Each avalanche consists of a certain number of the elementary pulses, cf. (1); these pulses are combined in a different way to embody the growth, peak, and relaxation phase of the avalanche. Thus, the sequence of avalanches in time contains additional information about the nature of the underlying stochastic process. For instance, the avalanche return [61,62] in many different complex systems reveals a non-Gaussian relaxation. For the knowledge-creation processes, we have demonstrated [44] that the avalanche first-return statistics obeys a q_r-Gaussian distribution with a large parameter $q_r \approx 2.45$. Here, we apply multifractal analysis to examine another signature of the complexity of these avalanche sequences. In particular, we analyze the temporal sequences of the avalanche sizes S_k, where $k = 1, 2, \ldots, K_{max}$ is the index of the avalanche and K_{max} stands for the total number of avalanches that occur in a particular time series. We consider the avalanche sequences obtained for the combinations of the parameters studied in the preceding sections and the empirical data; two examples of such avalanche sequences are displayed in Fig. 5 (bottom) and Fig. 8(a).

To determine the multifractal spectrum $\Psi(\alpha)$ for the sequences S_k, we apply the detrended multifractal analysis (DMFA); we use the approach which was applied to different types of complex signals, as described in Refs. [65–68]. According to the standard procedure, the profile of the signal is first constructed by the integration

$$Y(i) = \sum_{k=1}^{i} (S_k - \langle S \rangle). \tag{5}$$

The profile is then divided into $N_s = int(K_{max}/n)$ nonoverlapping segments of equal length n. Then for each segment $\mu = 1, 2 \ldots, N_s$, the local trend $y_\mu(i)$ is determined and the standard deviation around the local trend

$$F^2(\mu, n) = \frac{1}{n} \sum_{i=1}^{n} \{Y[(\mu - 1)n + i] - y_\mu(i)\}^2 \tag{6}$$

is found. Similarly, the procedure is repeated starting from the end of the signal, resulting in $F^2(\mu, n) = \frac{1}{n} \sum_{i=1}^{n} \{Y[N - (\mu - N_s)n + i] - y_\mu(i)\}^2$ for $\mu = N_s + 1, \ldots, 2N_s$. Combining the deviations at all segments, the qth order fluctuation function $F_q(n)$ is obtained according to

$$F_q(n) = \left\{ \frac{1}{2N_s} \sum_{\mu=1}^{2N_s} [F^2(\mu, n)]^{q/2} \right\}^{1/q} \sim n^{H(q)}, \tag{7}$$

and plotted against the varied segment length $n \in [2, int(K_{max}/4)]$. The scale invariance of $F_q(n)$ against the segment length n is examined to determine the corresponding scaling exponent $H(q)$. Here, the distortion parameter q takes a range of real values. The main idea is that the segments of the signal with potentially different fractal features will be suitably enhanced by a particular q value to become self-similar to the full signal and the corresponding scaling exponent $H(q)$ as a function of q is measured. Notably, different *small fluctuation segments* are enhanced by the negative values of q, and the segments with *large fluctuations* dominate the fluctuation function for the positive values of q. In the limiting case of monofractal, $H(q) = H(q = 2)$ is the standard Hurst exponent. Using the scaling relation $\tau(q) = qH(q) - 1$, the exponent $\tau(q)$ of the box probability, defined in the partition function method [65], is computed. Thus, the generalized Hurst exponents $H(q)$ can be related with the singularity multifractal spectrum via the Legendre transform $\Psi(\alpha) = q\alpha - \tau(q)$, where $\alpha = d\tau/dq$ is the singularity strength.

Figures 8(a)–8(c) shows the results for the case $ExpS$, which incorporates the features of the empirical data and the expertise matching in the simulations. Specifically, the avalanche series, the corresponding fluctuation function, and the generalized Hurst exponent are shown to demonstrate the procedure. Also, we show the results of the DMFA applied to the underlying time series for the same parameters. For the comparison, the analysis is performed for the randomized signals; the corresponding scaling exponents $H(q)$ are also depicted in Fig. 8(c). Notably, both the time series of the activity and the related avalanche series exhibit multifractal features. The span of the generalized Hurst exponent is much larger in the avalanche series. In the randomized case, the avalanche series exhibit almost unchanged multifractality while the time series of pulses becomes a monofractal with the properties of white noise [$H(q) = 0.5$ within the numerical

error bars]. These findings indicate that the origin of the multifractality in the time series may be found in the temporal correlations, which are of the $1/\nu$ type; see Fig. 6(a). While, for the avalanche series, the scale invariance of the distribution of sizes can be the sole reason for the observed multifractality. In this regard, it is interesting to analyze the width of the singularity spectrum $\Psi(\alpha)$ for different parameters of the model. These results are shown in Fig. 8(d). In agreement with the values of $H(q)$ for the case $ExpS$ in the panel (c), the spectra corresponding to the avalanche series and the randomized avalanche series are wide. In contrast, the multifractality of the original time series results in the narrow range; further, the spectrum is reduced to a close vicinity of the point $\Psi(\alpha = 0.5) = 1$, corresponding to the monofractal randomized time series. For the varied expertise of the actors, we obtain wide spectra of the avalanche series, where the singularity strength $\alpha \in [0.2, 1.7]$; cf. Fig. 8(d). Notably, the various parameters of the model mostly affect the right side of the spectrum, corresponding to $q < 0$ region, i.e., small fluctuation segments in the avalanche series. While, the variations are less pronounced for the large fluctuation segments, appearing in the left end of the spectrum $\Psi(\alpha)$. In this end, the empirical data and model simulations lead to similar results. These findings indicate how the sequences of small and large avalanches are affected by the available expertise; the occurrence of the large fluctuations in the avalanche sizes might be chiefly conditioned by the number of the actors involved.

IV. THE IMPACT OF VACANCIES ON MULTIFRACTAL SPECTRUM: A COMPARISON TO SPIN-ALIGNMENT AVALANCHES

As mentioned in the Introduction, the complexity of the social interactions prevents the exact description of the mechanisms at the elementary scale, calling for a comparison to better-understood physics models. In this regard, the interacting spin system described by the random-field Ising model at zero temperature and driven along the hysteresis loop represents a paradigm of complex dynamical behavior far from the equilibrium [69]. In this example, the spin alignment along the slowly increasing external field is balanced by spin-spin interactions and the local constraints due to the random-field disorder. The dynamics of spin flips under the disorder induced constraints and interactions was often employed to model opinion formation [70], processes driven by social balance [71] and other cases. For the purpose of this work, we aim to explore the impact of vacancies in the underlying network onto the multifractal spectrum of the avalanche sequences. We consider spin-alignment avalanches in the zero-temperature random-field Ising model (ZTRFIMc) with the two-state spin site $S_i(t) = \pm 1$ at each lattice site $i = 1, 2, \ldots, N$ with a fraction $c > 0$ of defect sites where the spin is absent. The energy $\mathcal{H} = -\sum_i \tilde{h}_i(t) S_i(t)$ is minimized by the spin alignment along the current value of the local field $\tilde{h}_i(t)$, where

$$\tilde{h}_i(t) = \sum_{j \in nn} J_{ij} S_j(t) + h_i + B(t); \quad S_i(t+1) = \text{sgn}[\tilde{h}_i(t)]$$

(8)

Here, h_i is the local quenched random field, which is described by the Gaussian distribution of zero mean and the width Δ. Starting from a large negative value, the system is driven quasistatically, i.e., by the slowly increasing external field $B(t+T) = B(t) + \delta B$ after an avalanche stops. The ferromagnetic interaction J_{ij} among the pair of spins at the adjacent sites i, j has the positive mean $\langle J_{ij} \rangle = J$ and the second cumulant $Jc(1-c)$, where $c > 0$ is the probability that the spin is absent at a randomly selected site. The dimensionless parameters $f \equiv \Delta/J$ and $r \equiv \delta B/J$ characterize the pinning strength and the driving rate, respectively.

Given the domain structure in these disordered systems, the magnetization reversal occurs in a series of jumps by the slow field ramping along the hysteresis loop. These magnetization changes are directly related to the motion of the domain walls, accompanying the expansion of the domains which are oriented parallel to the field. The size of the magnetization changes thus occur in the interplay between the driving by the external field and pinning of the domain walls at by the local random fields, oriented opposite. These magnetization changes in time represent the data points in the Barkhausen noise, a complex time series from which then the avalanches can be determined. The scale-invariant behavior of the Barkhausen avalanches and their dependence on the strength of the random-field disorder has been well understood [27,69,72–74]. Recently, it has been shown [68] that the Barkhausen noise exhibits multifractal structure. Moreover, the dynamical regime in the central part of the hysteresis loop, where large avalanches can occur for the weak disorder, has a significantly different spectrum from the dynamics at the beginning of the hysteresis loop. In contrast, the presence of hard defects in ZTRFIMc has been much less investigated. Specifically, even in the weak random-field pinning that allows system-wide avalanches, the presence of hard defects induces a characteristic length, which affects the cutoff size of the avalanches and also the universality of the scaling exponents [75,76]. Here, we are interested in the dynamics of the avalanches in the presence of site defects.

We consider a small concentration of the randomly dispersed site defects $c = 0.05$ on top of the weak random field disorder and slow driving; thus we use a representative set of parameters in this regime [68]: $f = 2.3 < f_c$ and $r = 0.02$ in the three-dimensional cubic lattice of 100^3 spins. In the absence of the site defects ($c = 0$), the system of this dimension would undergo a domain-wall depinning via large avalanches in the central part of the hysteresis loop [68]. However, the small percentage of site defects suffices to perturb this critical behavior by the pinning of the domain walls at a distance $\ell \sim 1/c$; whereas, at the distances $x \lesssim \ell$ the domain wall motion is accelerated by the external field, corresponding to the regime of the weak random-field pinning. Consequently, the small avalanches are similar as in the case of weak pinning without site defects, while the propagation of large avalanches is considerably hindered. These effects are reflected in the multifractal spectrum of the avalanche sequences, in particular, by increasing the difference in the generalized Hurst exponent for $q > 0$ and $q < 0$ values. The resulting singularity multifractal spectra are shown in the inset to Fig. 9. Although the avalanches tend to be larger in the central part of the hysteresis loop, the presence of site defects
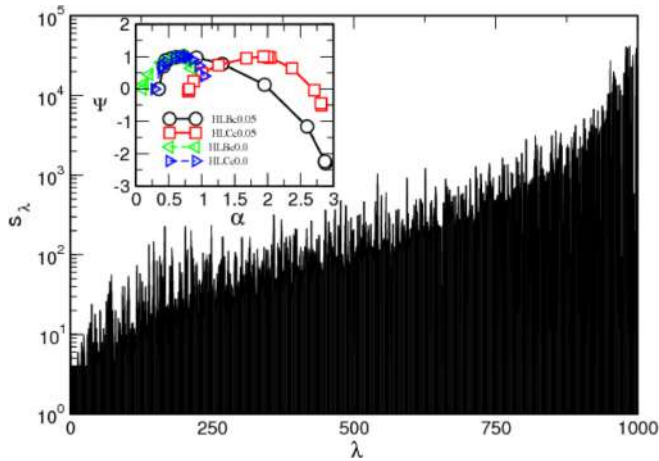
FIG. 9. The avalanche sequence at the beginning of the hysteresis loop (HLB) of the ZTRFIMc for $c = 0.05$ and weak random field pinning. Inset: The singularity multifractal spectra in the hysteresis loop center (HLC) and the loop beginning (HLB) for $c = 0.05$ and $c = 0.0$.

induces a mixture of small and large events, which results in the substantial broadening of the spectrum in comparison to the case of $c = 0$.

It is interesting to point the similarity between the spectra in the avalanche series that are affected by site defects in Fig. 9 with the spectra of social avalanches of knowledge creation, Fig. 8. In this case, the avalanche propagation is conditioned by the actor's activity patterns and the (non)possession of the required expertise. According to Fig. 2(a), a large number of users (agents) stops to be active after a given number of actions, while their artifacts are still available and can be the subject of interest to others. Thus, each avalanche becomes pinned by reaching the network node of such a user, in the manner that a site defect pins the moving domain wall. Also, the users whose delay time is typically larger than the avalanche duration may have a similar effect on the current avalanche propagation. On the other hand, the extremely active users [a small fraction, represented by the end of the distribution in Fig. 2(a)] accelerate the propagation through their numerous connections and also by being active at more than one question within a short time interval. This type of the actor's heterogeneity results in a typical mixture of small and large events, as seen in the analysis in Sec. III. Beyond the shape of the multifractal spectra, representing a combination of small and large events, the avalanche distributions differ in the random-site ferromagnetic model. Among other reasons, the evolving bipartite networks of the actors and their artifacts are identified as those being of the key importance.

V. DISCUSSION AND CONCLUSIONS

Considering the large data set from Q&A site Mathematics Stack Exchange and the agent-directed modeling, we have analyzed the avalanching behavior and the bipartite network that underlies the creation of collective knowledge. Given the complexity of modeling the human actors, we have kept the agent's properties statistically similar to the features of the users, detectable from the same empirical data. In

particular, the agent's activity pattern is designed by the distributions of the number of actions per user $P(N_i)$ and the interactivity time $P(\Delta t)$ as well as the arrival rate $p(t)$, which are mutually interconnected and characterize the human dynamics in the considered empirical system. We have varied the agent's expertise, as the most relevant feature to the knowledge creation. To evaluate its impact, the expertise matching to the contents of the artifact has been clearly observed in the simulations. (Other potential extensions of the model, e.g., neglecting or altering the above distributions independently, that goes beyond human dynamics [50], are not discussed in this work.)

The analysis indicates that the knowledge-creation process represents a particular class of social dynamics, in which the self-tuning towards the criticality is controlled by the use of knowledge in the *meaningful* actions and the coevolution of the underlying network. Our key findings are here discussed.

The self-organized criticality. The robustly observed temporal correlations, avalanching and multifractality, as well as the scale-invariance dependence on the driving rate, indicate that the criticality might occur in these stochastic processes in a self-tuned manner. The social dynamics, driven by the arrival of new actors and innovation that they bring, represents the main source of the avalanching behaviors. Whereas, the altered degree of branching (e.g., in the μ process) and, more importantly, the strict use of the expertise imposes the constraints to the social dynamics, which affects the avalanche propagation. These constraints then manifest in the nonuniversal scaling exponents of the avalanche sizes and durations. Moreover, the relative fraction of the small and large avalanches, which appear to be mixed in the course of the process, depends on the available expertise of the actors, thus affecting the width of the multifractal spectrum. Nonextensivity ($q_a > 1$) is another remarkable feature of the knowledge-creation process, where the q_a-exponential distribution applies to the avalanche sizes and the q_r-Gaussian distribution to the avalanche returns. Although the observed multifractality of the avalanche series is compatible with these distributions, more theoretical work is needed to unravel their origin.

The structure of network partitions. In this process, the growth of the bipartite network occurs by the addition of layers, which contain new arrivals and other active agents, and their artifacts. How the network will grow is strongly related to the expertise-conditioned linking. After a sufficient time, the network exhibits a broad distribution of the degree of nodes in both partitions. The scaling exponent of the agent's degree distribution in each case is close to the introduced distribution of the activity $P(N_i)$. However, the cutoffs of the distributions, indicating the actual size of the network, vary with the considered apportionment of the expertise. Notably, the network size increases when the agents possess a larger expertise in the average. In this case, the probability of an agent to connect to a suitable artifact is elevated. Similarly, the network growth is accelerated by the addition of a larger number of actors in the initial stages of the process, which results in a greater number of the available artifacts. Thus, in the social process, the scale freeness of the user partition is determined by the actual activity profiles alone. Where the corresponding edges will appear in the network, and,

consequently the network's community structure [44], depends on the available expertise and how it is distributed over the actors. On the other hand, the question-partition degree distribution strongly depends on the expertise. Note that in both cases the determined scaling exponents are smaller than 2. These results suggest that the effective mechanisms might be different from the popular "preferential" attachment-and-rewiring rule that leads to the exponents larger than 2.

The self-tuning factors. The heterogeneity of the actor's profiles and the activity patterns, with the extended range of delay times and the number of actions, is certainly an important factor for the appearance of the avalanches in the studied social dynamics. Apart from these "human factors," it is stipulated that some other ingredients of the knowledge creation may contribute to the paths towards the criticality. Specifically, the importance and the use of expertise in the human collaborative endeavor makes the social process substantially different from, for instance, the activity on popular blogs, where the negative emotional charge of comments may lead to a supercritical avalanche [13,14]. Here, the required expertise of the actors needs to match the contents of the question, resulting in a balanced activity (resembling the energy balance in the driven physical system). Consequently, the activity stops when sufficient knowledge is built through the answers on a particular question, depending on the available expertise of the actors. The same artifacts may become a focus in the later stages when new necessary knowledge becomes available, i.e., by the arrival of new players. Thus, in contrast to the standard social dynamics, a kind of optimization of the available expertise applies, which is also compatible with the nonextensive dynamics mentioned above. Note that the optimization of the system's efficiency is often associated with the functioning of biological systems [7] and the avalanching process in neuronal assemblies [8,9] and the brain [12], which are still not well understood. Furthermore, the underlying network evolution by the addition of the innovative contents in the active layer can be seen as another decisive factor to provide a particular type of critical behavior. Theoretically, changing the random environment for the self-organizing process affects the universality of the critical behavior that can be achieved. The renormalization group study of the critical sandpile model in the presence of quenched [43] or annealed [42] random currents has demonstrated that a new stable fixed point appears, which is controlled by the variance of the random variable. For the knowledge creation in the online Q&A communities, it is relevant to mention that the innovation expansion builds

the network of contents with a logical structure, which originates in the participant's individual knowledge. For the same empirical data, this aspect of knowledge creation was demonstrated in [47] by analysis of the knowledge network (containing the cognitive contents that are used in all questions and answers, and encoded by the standard mathematical classification scheme). Further research is needed to disclose the potential importance of avalanches of knowledge creation in the off-line social communities, where knowledge sharing can lead to the creation of a common opinion and other collective behavior [13,19].

In summary, the creation of collective knowledge through questions and answers is a self-organized critical process where the mechanisms of self-tuning are provided by the interplay of the social and cognitive layer. The observed SOC is robust to the increased driving rate within limits pertinent to the considered experimental system. Our study suggests some questions for further theoretical considerations, in particular: the formal differences between knowledge creation and common social dynamics; the potential similarity between the knowledge creation and the brain avalanching dynamics; the origin of the nonextensivity (although the nonextensive character of the dynamics is intuitive in the context of knowledge, the formal origin of the q-Gaussian fluctuations is not understood), and other issues. The presented results, based on the empirical data and the agent-directed model, which is almost equally complex as the empirical system itself, reveal many factors that act in unison and contribute to the observed SOC. The presented comparison to the driven spin system with site defects suggests that apart from the expertise, the heterogeneity of the actor's activity patterns is an essential factor that prevents the appearance of the supercritical avalanches. Our findings may help design formal theoretical models of SOC, e.g., of the cellular automata type or the continuous models suitable for the renormalization group analysis, which may be capable of describing the unique role of each of these factors.

[1] D. Marković and C. Gros, Power laws and self-organized criticality in theory and nature, Phys. Rep. **536**, 41 (2014).

[2] M. J. Aschwanden, *Self-organized criticality systems* (Open Academic Press, Berlin, 2013).

[3] *Avalanches in Functional Materials and Geophysics*, edited by E. K. H. Salje, A. Saxena, and A. Planes (Springer International Publishing AG, Cham, Switzerland, 2017).

[4] P. Bak, C. Tang, and K. Wiesenfeld, Self-Organized Criticality: An Explanation of the 1/f Noise, Phys. Rev. Lett. **59**, 381 (1987).

[5] D. Dhar, Self-Organized Critical State of Sandpile Automaton Models, Phys. Rev. Lett. **64**, 1613 (1990).

[6] B. Tadić and D. Dhar, Emergent Spatial Structures in Critical Sandpiles, Phys. Rev. Lett. **79**, 1519 (1997).

[7] T. Mora and W. Bialek, Are biological systems poised at criticality? J. Stat. Phys. **144**, 268 (2011).

[8] A. Levina, J. M. Herrmann, and T. Geisel, Dynamical synapses causing self-organized criticality in neural networks, Nat. Phys. **3**, 857 (2007).

[9] J. G. Orlandi, J. Soriano, E. Alvarez-Lacalle, S. Teller, and J. Casademunt, Noise focusing and the emergence of coherent activity in neuronal cultures, Nat. Phys. **9**, 582 (2013).

[10] A. Cavagna, A. Cimarelli, I. Giardina, G. Parisi, R. Santagati, F. Stefanini, and M. Viale, Scale-free correlations in starling flocks, Proc. Natl. Acad. Sci. USA **107**, 11865 (2010).

[11] D. S. Zhukov, V. V. Kanishchev, and S. K. Lyamin, Application of the theory of self-organized criticality to the investigation of historical processes, SAGE Open **6**, 2158244016683216 (2016).

[12] J. Hesse and T. Gross, Self-organized criticality as a fundamental property of neural systems, Front. Syst. Neurosci. **8**, 166 (2014).

[13] B. Tadić, V. Gligorijević, M. Mitrović, and M. Šuvakov, Co-evolutionary mechanisms of emotional bursts in online social dynamics and networks, Entropy **15**, 5084 (2013).

[14] M. Mitrović and B. Tadić, Bloggers behavior and emergent communities in blog space, Eur. Phys. J. B **73**, 293 (2010).

[15] M. Mitrović, G. Paltoglou, and B. Tadić, Quantitative analysis of bloggers' collective behavior powered by emotions, J. Stat. Mech. (2011) P02005.

[16] M. Šuvakov, M. Mitrović, V. Gligorijević, and B. Tadić, How the online social networks are used: Dialogues-based structure of MySpace, J. R. Soc., Interface **10**, 20120819 (2012).

[17] M. Šuvakov and B. Tadić, Collective emotion dynamics in chats with agents, moderators and bots, Condens. Matter Phys. **17**, 33801 (2014).

[18] Yu. Holovatch, O. Mrygold, M. Szell, and S. Thurner, Analysis of a virtual world, in *Math Meets Myths: Quantitative Approaches to Ancient Narratives*, edited by R. Kenna *et al.*, Understanding Complex Systems (Springer International Publishing Switzerland, 2017).

[19] G. Kou, Y. Zhao, Y. Peng, and Y. Shi, Multi-level opinion dynamics under bounded confidence, PLoS One **7**, e43507 (2012).

[20] M. Malkov, A. Das, I. Gruzinov, M. N. Rosenbluth, C. Holland, B. Wecht, A. I. Smolyakov, F. L. Hinton, Z. Lin, T. S. Hahm, P. H. Diamond, and S. Champeaux, Secondary instability in drift wave turbulence as a mechanism for zonal flow and avalanche formation, Nucl. Fusion **41**, 1067 (2001).

[21] Y. Idomura, H. Urano, N. Aiba, and S. Tokuda, Study of ion turbulent transport and profile formations using global gyrokinetic full-f Vlasov simulation, Nucl. Fusion **49**, 065029 (2009).

[22] K. P. Zybin and V. A. Sirota, Multifractal structure of fully developed turbulence, Phys. Rev. E **88**, 043017 (2013).

[23] Á. Corral, Long-Term Clustering, Scaling, and Universality in the Temporal Occurrence of Earthquakes, Phys. Rev. Lett. **92**, 108501 (2004).

[24] C. Narteau, S. Byrdina, P. Shebalin, and D. Schorlemmer, Common dependence on stress for the two fundamental laws of statistical seismology, Nature (London) **462**, 642 (2009).

[25] F. Dalton and D. Corcoran, Basin of attraction of a bounded self-organized critical state, Phys. Rev. E **65**, 031310 (2002).

[26] T. Hatano, C. Narteau, and P. Shebalin, Common dependence on stress for the statistics of granular avalanches and earthquakes, Sci. Rep. **5**, 12280 (2015).

[27] D. Spasojević, S. Bukvić, S. Milošević, and H. E. Stanley, Barkhausen noise: Elementary signals, power laws, and scaling relations, Phys. Rev. E **54**, 2531 (1996).

[28] A. Berger, A. Inomata, J. S. Jiang, J. E. Pearson, and S. D. Bader, Experimental Observation of Disorder-Driven Hysteresis-Loop Criticality, Phys. Rev. Lett. **85**, 4176 (2000).

[29] H.-S. Lee, K.-S. Ryu, I.-S. Kang, and S.-C. Shin, Universal Barkhausen critical scaling behavior observed in Ni_xFe_{1-x} ($x = $ 0–0.5) films, J. Appl. Phys. **109**, 07E101 (2011).

[30] L. Carrillo and J. Ortín, Avalanches in the growth of stress-induced martensites, Phys. Rev. B **56**, 11508 (1997).

[31] D. Soto-Parra, X. Zhang, S. Cao, E. Vives, E. K. H. Salje, and A. Planes, Avalanches in compressed Ti-Ni shape-memory porous alloys: An acoustic emission study, Phys. Rev. E **91**, 060401 (2015).

[32] G. F. Nataf, P. O. Castillo-Villa, J. Baró, X. Illa, E. Vives, A. Planes, and E. K. H. Salje, Avalanches in compressed porous SiO_2-based materials, Phys. Rev. E **90**, 022405 (2014).

[33] N. Friedman, S. Ito, B. A. W. Brinkman, M. Shimono, R. E. Lee DeVille, K. A. Dahmen, J. M. Beggs, and T. C. Butler, Universal Critical Dynamics in High Resolution Neuronal Avalanche Data, Phys. Rev. Lett. **108**, 208102 (2012).

[34] M. Šuvakov and B. Tadić, Modeling collective charge transport in nanoparticle assemblies, J. Phys.: Condens. Matter **22**, 163201 (2010).

[35] E. Altshuler, T. H. Johansen, Y. Paltiel, P. Jin, K. E. Bassler, O. Ramos, Q. Y. Chen, G. F. Reiter, E. Zeldov, and C. W. Chu, Vortex avalanches with robust statistics observed in superconducting niobium, Phys. Rev. B **70**, 140505 (2004).

[36] M.-Y. Im, K.-S. Lee, A. Vogel, J.-I. Hong, G. Meier, and P. Fischer, Stochastic formation of magnetic vortex structures in asymmetric disks triggered by chaotic dynamics, Nat. Commun. **5**, 5620 (2014).

[37] X. Jiang, L. Thomas, R. Moriya, M. Hayashi, B. Bergman, C. Rettner, and S. S. P. Parkin, Enhanced stochasticity of domain wall motion in magnetic racetracks due to dynamic pinning, Nat. Commun. **1**, 25 (2010).

[38] V. D. Nguyen, W. Savero Torres, P. Laczkowski, A. Marty, M. Jamet, C. Beigné, L. Notin, L. Vila, and J. P. Attané, Elementary depinning processes of magnetic domain walls under fields and currents, Sci. Rep. **4**, 6509 (2014).

[39] B. Tadić, Dynamic criticality in driven disordered systems: Role of depinning and driving rate in Barkhausen noise, Physica A **270**, 125 (1999).

[40] *Nonextensive Entropy—Interdisciplinary Applications*, edited by M. Gell-Mann and C. Tsallis, Santa Fe Institute Studies in the Science of Complexity (Oxford University Press, Oxford, 2004).

[41] U. C. Täuber, Phase transitions and scaling in systems far from equilibrium, Annu. Rev. Condens. Matter Phys. **8**, 185 (2017).

[42] N. V. Antonov, and P. I. Kakin, Effects of random environment on a self-organized critical system: Renormalization group analysis of a continuous model, EPJ Web Conf. **108**, 02009 (2016).

[43] B. Tadić, Disorder-induced critical behavior in driven diffusive systems, Phys. Rev. E **58**, 168 (1998).

[44] M. M. Dankulov, R. Melnik, and B. Tadić, The dynamics of meaningful social interactions and the emergence of collective knowledge, Sci. Rep. **5**, 12197 (2015).

[45] J. Kimmerle, U. Kress, and C. Held, The interplay between individual and collective knowledge: technologies for organisational learning, Knowledge Management Research Practice **8**, 33 (2010).

[46] *Social Interactions and the Development of Knowledge*, edited by J. I. M. Carpendale and U. Müller (Lawrence Erlbaum Associates, Inc., Mahwah, NJ, 2013).

[47] M. Andjelković, B. Tadić, M. M. Dankulov, M. Rajković, and R. Melnik, Topology of innovation spaces in the knowledge networks emerging through questions-and-answers, PLoS ONE **11**, e0154655 (2016).

[48] M. Mitrović, G. Paltoglou, and B. Tadić, Networks and emotion-driven user communities at popular blogs, Eur. Phys. J. B **77**, 597 (2010).

[49] M. Mitrović and B. Tadić, Dynamics of bloggers' communities: Bipartite networks from empirical data and agent-based modeling, Physica A **391**, 5264 (2012).

[50] B. Tadić, Modeling behavior of Web users as agents with reason and sentiment, in *Advances in Computational Modeling Research: Theory, Developments and Applications*, edited by A. B. Kora (Nova Publishing, New York, 2013), pp. 177–186.

[51] S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes, Critical phenomena in complex networks, Rev. Mod. Phys. **80**, 1275 (2008).

[52] *Complex Networks and their Applications*, edited by H. Cherifi (Cambridge Scholar Publishing, Newcastle upon Tyne, 2014).

[53] T. Nishikawa, J. Sun, and A. E. Motter, Sensitive dependence of network dynamics on network structure, arXiv:1611.01164.

[54] B. Tadić and S. Thurner, Information super-diffusion on structured networks, Physica A **332**, 566 (2004).

[55] N. E. Savitskaya, Stabilization of avalanche processes on dynamical networks, JETP Lett. **103**, 206 (2016).

[56] P. Fronczak, A. Fronczak, and J. A. Hołyst, Self-organized criticality and coevolution of network structure and dynamics, Phys. Rev. E **73**, 046117 (2006).

[57] D. J. B. Soares, C. Tsallis, A. M. Mariz, and L. R. Da Silva, Preferential attachment growth model and nonextensive statistical mechanics, Europhys. Lett. **70**, 70 (2005).

[58] L. Hébert-Dufresne, A. Allard, J.-G. Young, and L. J. Dubé, Constrained growth of complex scale-independent systems, Phys. Rev. E **93**, 032304 (2016).

[59] C. Tsallis and D. J. Bukman, Anomalous diffusion in the presence of external forces: Exact time-dependent solutions and their thermostatistical basis, Phys. Rev. E **54**, R2197 (1996).

[60] C. Tsallis, The nonadditive entropy Sq and its applications in physics and elsewhere: Some remarks, Entropy **13**, 1765 (2011).

[61] G. P. Pavlos, L. P. Karakatsanis, M. N. Xenakis, E. G. Pavlos, A. C. Iliopoulos, and D. V. Sarafopoulos, Universality of nonextensive Tsallis statistics and time series analysis: Theory and applications, Physica A **395**, 58 (2014).

[62] F. Caruso, A. Pluchino, V. Latora, S. Vinciguerra, and A. Rapisarda, Analysis of self-organized criticality in the Olami-Feder-Christensen model and in real earthquakes, Phys. Rev. E **75**, 055101 (2007).

[63] Á. Corral and M. Paczuski, Avalanche Merging and Continuous Flow in a Sandpile Model, Phys. Rev. Lett. **83**, 572 (1999).

[64] F. J. Pérez-Reche, B. Tadić, L. Mañosa, A. Planes, and E. Vives, Driving Rate Effects in Avalanche-Mediated First-Order Phase Transitions, Phys. Rev. Lett. **93**, 195701 (2004).

[65] J. W. Kantelhardt, S. A. Zschiegner, E. Koscielny-Bunde, S. Havlin, A. Bunde, and H. E. Stanley, Multifractal detrended fluctuation analysis of nonstationary time series, Physica A **316**, 87 (2002).

[66] A. N. Pavlov and V. S. Anishchenko, Multifractal analysis of complex signals, Phys. Usp. **50**, 819 (2007).

[67] M. S. Movahed, G. R. Jafari, F. Ghasemi, S. Rahvar, and M. R. R. Tabar, Multifractal detrended fluctuation analysis of sunspot time series, J. Stat. Mech. (2006) P02003.

[68] B. Tadić, Multifractal analysis of Barkhausen noise reveals the dynamic nature of criticality at hysteresis loop, J. Stat. Mech. (2016) 063305.

[69] T. P. Handford, F. J. Perez-Reche, and S. Taraskin, Mechanisms of evolution of avalanches in regular graphs, Phys. Rev. E **87**, 062122 (2013).

[70] *Biology, Sociology, Geology by Computational Physicists*, edited by D. Staufer, S. Moss De Oliveira, P. de Oliveira, and J. de Sá Martins (Elsevier, Amsterdam, 2006), Vol. 1.

[71] Q.-K. Meng, Self-organized criticality in small-world networks based on the social balance dynamics, Chin. Phys. Lett. **28**, 118901 (2011).

[72] F. J. Pérez-Reche and E. Vives, Finite-size scaling analysis of the avalanches in the three-dimensional Gaussian random-field Ising model with metastable dynamics, Phys. Rev. B **67**, 134421 (2003).

[73] F. J. Pérez-Reche and E. Vives, Spanning avalanches in the three-dimensional Gaussian random-field Ising model with metastable dynamics: Field dependence and geometrical properties, Phys. Rev. B **70**, 214422 (2004).

[74] F. J. Perez-Reche, C. Triguero, G. Zanzotto, and L. Truskinovsky, Origin of scale-free intermittency in structural first-order phase transitions, Phys. Rev. B **94**, 144102 (2016).

[75] B. Tadić, Nonuniversal Scaling Behavior of Barkhausen Noise, Phys. Rev. Lett. **77**, 3843 (1996).

[76] B. Tadić, Switching current noise and relaxation of ferroelectric domains, Eur. Phys. J. B **28**, 81 (2002).

Spectral properties of hyperbolic nanonetworks with tunable aggregation of simplexes

Marija Mitrović Dankulov,[1,2] Bosiljka Tadić,[2,3] and Roderick Melnik[4,5]

[1]*Scientific Computing Laboratory, Center for the Study of Complex Systems, Institute of Physics Belgrade,*
University of Belgrade, Pregrevica 118, 11080 Belgrade, Serbia
[2]*Department of Theoretical Physics, Jožef Stefan Institute, Jamova 39, 1000 Ljubljana, Slovenia*
[3]*Complexity Science Hub Vienna, Josephstadterstrasse 39, 1080 Vienna, Austria*
[4]*MS2Discovery Interdisciplinary Research Institute, M2NeT Laboratory and Department of Mathematics,*
Wilfrid Laurier University, 75 University Ave W, Waterloo, Ontario, Canada N2L 3C5
[5]*BCAM–Basque Center for Applied Mathematics, Alameda de Mazarredo 14, E-48009 Bilbao, Spain*

Cooperative self-assembly is a ubiquitous phenomenon found in natural systems which is used for designing nanostructured materials with new functional features. Its origin and mechanisms, leading to improved functionality of the assembly, have attracted much attention from researchers in many branches of science and engineering. These complex structures often come with hyperbolic geometry; however, the relation between the hyperbolicity and their spectral and dynamical properties remains unclear. Using the model of aggregation of simplexes introduced by Šuvakov *et al.* [Sci. Rep. **8**, 1987 (2018)], here we study topological and spectral properties of a large class of self-assembled structures or nanonetworks consisting of monodisperse building blocks (cliques of size $n = 3, 4, 5, 6$) which self-assemble via sharing the geometrical shapes of a lower order. The size of the shared substructure is tuned by varying the chemical affinity ν such that for significant positive ν sharing the largest face is the most probable, while for $\nu < 0$, attaching via a single node dominates. Our results reveal that, while the parameter of hyperbolicity remains $\delta_{\max} = 1$ across the assemblies, their structure and spectral dimension d_s vary with the size of cliques n and the affinity when $\nu \geqslant 0$. In this range, we find that $d_s > 4$ can be reached for $n \geqslant 5$ and sufficiently large ν. For the aggregates of triangles and tetrahedra, the spectral dimension remains in the range $d_s \in [2, 4)$, as well as for the higher cliques at vanishing affinity. On the other end, for $\nu < 0$, we find $d_s \approx 1.57$ independently on n. Moreover, the spectral distribution of the normalized Laplacian eigenvalues has a characteristic shape with peaks and a pronounced minimum, representing the hierarchical architecture of the simplicial complexes. These findings show how the structures compatible with complex dynamical properties can be assembled by controlling the higher-order connectivity among the building blocks.

I. INTRODUCTION

Controlled self-assembly of nanoparticles with various properties has enabled the engineering of wide classes of materials with new functional features [1]. Among others, the possibilities of designing and assembling three-dimensional (3D) structures of colloidal particles have increased significantly by the discovery of methods for the synthesis of colloids with controlled symmetries and directional interactions [2]. Further possibilities are opened with cooperative self-assembly, where the groups of nanoparticles forming small clusters can join the growing structure [3–7]. These processes utilize a variety of interparticle forces [8], as well as geometry-guided self-assembly [9–11]. By varying the building blocks in different self-assembly processes, the impact of the system's architecture on the emergent functionality in nanostructured materials has been evidenced by experimental investigation, e.g., by the charge transport or spin diffusion, resulting in the enhanced collective dynamics of the assembly [1,4–7,9,12]. On the other side, theoretical investigations of the structure-function interdependence have been greatly facilitated by mapping the assembly onto mathematical graphs or nanonetworks [13]. In this representation, nodes can indicate nanoparticles with their geometrical, physical,

and chemical properties, and edges specify the interparticle interaction or chemical bonding often enabled by their physical proximity. This representation allows the use of graph-theory methods to quantify topology and facilitates numerical modeling, as was done, for example, in the study of charge transport by single-electron tunnelings in nanoparticle films [12,14–16], carbon nanotube fillers [17], and others.

On a more global scale, the interplay between the structure and dynamics is captured by spectral properties of networks [18,19]. More specifically, spectral analysis of the adjacency matrix or the Laplacian operator related to the adjacency matrix [20] revealed Fiedler spectral partitioning of the graph and detection of functional modules or mesoscopic communities [21,22], hierarchical organization and homeostatic response [23], the structural changes at the percolation threshold [24], or the occurrence of assortative correlations between nodes [25] and the origin and implications of the degeneracy in network spectra [26,27]. A direct relation between the Laplacian eigenspectrum and the diffusion processes on that network revealed the role of the small-degree nodes and features of the return time of random walks [21,28], as well as the universality of dynamical phase transitions [29] and a deeper understanding of synchronization on complex networks [30]. In this context, the key quantity that relates the structure

to the diffusion and synchronization on a network is the *spectral dimension* [31–33], which can be determined from the properties of the Laplacian spectrum.

The complex functional systems often exhibit a hierarchical architecture and the related hyperbolic geometry. The underlying higher-order connectivity in these structures can be modeled with *simplicial complexes* and describe it with mathematical techniques of the algebraic topology of graphs [34–38]. In this context, simplexes are cliques of different orders $q = 0, 1, 2, 3, \ldots$ representing, respectively, nodes, edges, triangles, tetrahedra, and so on, which are joined into larger-scale structures. Note that a clique of the order q contains cliques of the lower orders up to $q - 1$ as its *faces*. The assemblies of cliques can be regarded as topological spaces represented by simplicial complexes. Formally, it holds that in a simplicial complex $\mathcal{K}$, every face of a clique $\sigma \in \mathcal{K}$ also belongs to $\mathcal{K}$, and that a nonempty intersection of two simplexes $\sigma_1, \sigma_2 \in \mathcal{K}$ is a face of both of them. In the context of simplicial complexes, the 1-skeleton consists of nodes and edges, which is the topological graph that is accessible to analysis using standard methods of graph theory. The idea of hierarchical architecture is a center piece in the development of many modern innovative technologies such as 3D printing [39]. In the materials science that motivates our work, such structures are grown by cooperative self-assembly [1,4,6–8,11,40]. Recently, these processes have been modeled by attachments of preformatted objects or simplexes [10,41] under geometric constraints and suitably specified binding rules and parameters. We also draw attention to several other contemporary studies [42–47] that show the importance of simplexes in modeling interactions of higher orders and complex geometry in various physical and biological systems. Whereas in real complex systems whose structure is detectable from experimental data, the corresponding structure can be decomposed into simplicial complexes. For example, in the case of human connectome studied in Refs. [48,49], these simplicial complexes comprise the inner structure of brain anatomical modules. The presence of cliques leading to a hierarchical organization was also found in social networking dynamics [50–52], problems related to traffic dynamics [53], protein-protein interaction networks [23], and so on.

As mentioned above, the hierarchically organized networks possess emergent hyperbolicity or negative curvature in the shortest-path metric, that is, they are Gromov hyperbolic graphs [54–58]. Recently the graphs with a small hyperbolicity parameter δ have been in the focus of the scientific community for their ubiquity in real systems and applications, as well as due to their mathematically interesting structure [55–57,59]. Namely, the upper bound of a small hyperbolicity parameter can be determined from a subjacent smaller graph of a given structure. Generally, it is assumed that both naturally evolving, biological, physical, and social systems develop a negative curvature to optimize their dynamics [42,49,52,60,61]. However, the precise relationship between the hyperbolicity of a network and its spectral and dynamical features remains mostly unexplored.

In this paper, we tackle these issues by systematically analyzing the spectral properties of a class of Gromov 1-hyperbolic graphs, which represent nanonetworks with different architectures of simplicial complexes. Based on

the model for the cooperative self-assembly of simplexes introduced in Ref. [41], here we grow several classes of nanonetworks and analyze their topology and spectral properties; the monodisperse building blocks are cliques of the order $n = 3, 4, 5, 6$ while the geometrical compatibility tunes their assembly in the interplay with the varying chemical affinity ν of the growing structure towards the binding group. Specifically, for the negative values of the parameter ν, the effective repelling interaction between the simplexes occurs, while it is gradually attractive for the positive ν. At $\nu = 0$ purely geometrical factors play a role. Our results show that while the hyperbolicity parameter remains constant $\delta = 1$ for all classes, their spectral dimension varies with the chemical affinity ν and the size of the elementary building blocks n. Moreover, these networks exhibit a community structure when the parameter $\nu \geqslant 0$. The inner structure of these communities consists of simplicial complexes with a hierarchical architecture, which manifests itself in the characteristic spectral properties of the Laplacian of the network.

In Sec. II we present details of the model and parameters, while in Sec. III we study different topology features of the considered networks. In Sec. IV we analyze in detail spectral properties of all classes of these networks for varied parameters ν and the size of elementary blocks. Section V is devoted to discussion of the results.

II. SELF-ASSEMBLY OF SIMPLEXES AND THE TYPE OF EMERGENT STRUCTURES

For the growth of different nanonetworks, we use the model rules for the cooperative self-assembly [41,62] with the chemical aggregation of simplexes. Preformatted groups of particles are described by simplexes (full graphs, cliques) of different size $n \equiv q_{\max} + 1$, where $q_{\max}$ indicates the order of the clique. Starting from an initial simplex, at each step, a new simplex is added and attached to the growing network by *docking* along one of its faces, which are recognized as simplexes of the lower order $q = 0, 1, 2, \ldots, q_{\max} - 1$; see online demo [63]. For example, a tetrahedron can be attached by sharing a single node, i.e., a simplex of the order $q = 0$ with the existing network, or sharing an edge, $q = 1$, or a triangle, $q = 2$, with an already existing simplex in the network. The attaching probability depends both on the geometrical compatibility of the q-face of the adding simplex with the current structure as well as on the parameter ν that describes the chemical affinity of that structure towards the addition of new n_a nodes, where $n_a = q_{\max} - q$. More precisely, we have [41]

$$p(q_{\max}, q; t) = \frac{c_q(t) e^{-\nu(q_{\max} - q)}}{\sum_{q=0}^{q_{\max}-1} c_q(t) e^{-\nu(q_{\max} - q)}} \qquad (1)$$

for the normalized probability that a clique of the order $q_{\max}$ attaches along its face of the order q. Here $c_q(t)$ is the number of the geometrically similar docking sites of the order q at the evolution time t. Eventually, one of them is selected randomly. By varying the parameter ν from large negative to large positive values, the probability of docking along with a particular face is considerably changed. For example, for the negative values of ν, the growing system "likes" new vertices; consequently, a simplex preferably attaches along a shared vertex
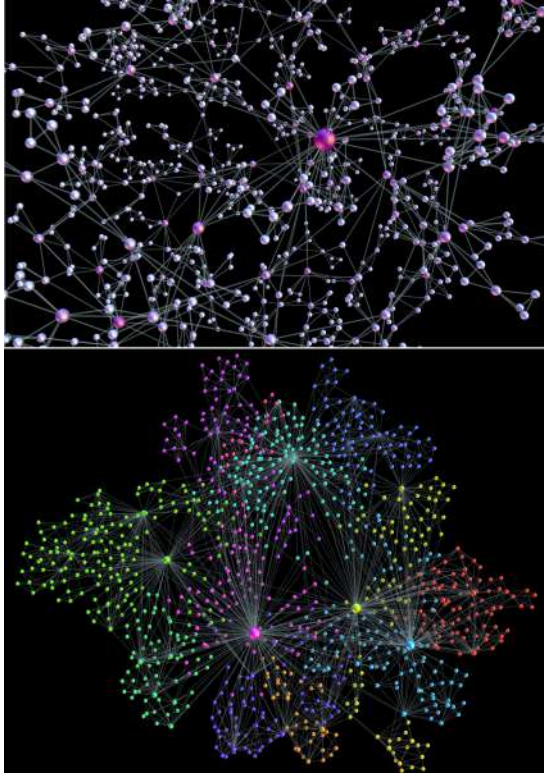
FIG. 1. Aggregates of tetrahedra with strong repulsion, a segment is shown in the top panel, and the case with strong attraction resulting in the network with communities indicated by different colors is shown in the bottom panel.

rather than a larger structure. Effectively, a repulsion between simplexes occurs; see Fig. 1 top. In the other limit, for a large positive ν, the most probable docking is along the potentially largest face, such that an added simplex of the size n shares the maximum number $n-1$ of vertices with the existing structure; see bottom panel in Figs. 1 and 2. Here the simplexes in question experience a strong attraction, which gradually decreases with decreasing ν. For the neutral case $\nu = 0$, the assembly is regulated by strictly geometrical compatibility factors $C_q(t)$, which change over time as the network grows.

In the original model [41], the size of the incoming simplexes is taken from a distribution, whose parameters can be varied. To reveal the impact of the size of these building blocks on the spectral properties of the new structure, here we focus on the networks with *monodisperse* cliques; in particular, we investigate separately the structures grown by aggregation of cliques of the size $n = 3$, 4, 5, and 6 for continuously varied affinity ν. For comparison, we also consider the case with a distribution of simplexes in the range $n \in [3, 6]$. As the examples in Fig. 1 and Fig. 2 show, the structure of the assembly varies considerably with both the size of simplexes and the level of attraction between them. Notice that in the case $n = 2$ the simplex consists of two vertices with an edge between them resulting in a random tree graph. Here $q_{max} = 1$ and all docking faces are single-vertex sites ($q = 0$). Therefore, the probability $p(1, 0; t) = 1$ is independent of the value of the parameter ν. In this work, we consider networks of different number of vertices $N = 1000$, 5000, and 10 000.
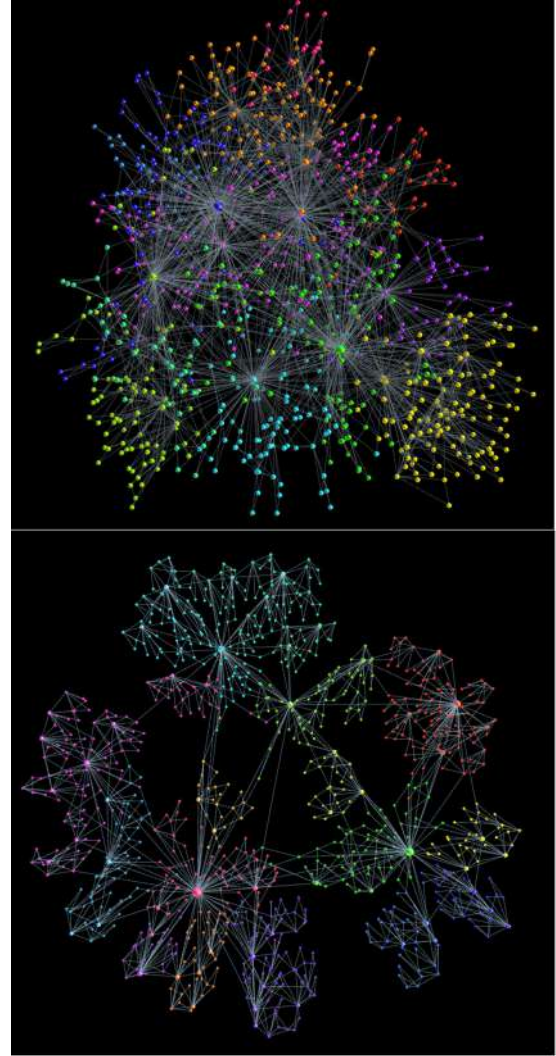

FIG. 2. The networks of the aggregated cliques of mixed sizes $n \in [3, 6]$ distributed according to $\propto n^{-2}$ for $\nu = 5$ (top), and aggregates of triangles for $\nu = 9$ (bottom). The community structure is indicated by different colors of nodes.

III. TOPOLOGICAL PROPERTIES OF THE ASSEMBLED NANONETWORKS

The structure of the assemblies strongly depends on the chemical affinity ν and the size n of the building blocks. For example, a strong repulsion between cliques enables sharing a single node, thus minimizing the geometrical compatibility factor and resulting in a sparse graph (a tree-of-cliques). An example with the tetrahedra as building blocks at $\nu = -9$ is shown in the top panel of Fig. 1. However, for extremally attractive cliques, e.g., for $\nu = 9$, the same building blocks attach mostly via sharing their largest subgraphs (in this case, triangles); thus the geometrical constraints play an important role. This situation results in a dense nanonetwork with a nontrivial community structure, determined by the modularity optimization method [64], as shown in the bottom panel of Figs. 1 and 2. Meanwhile, the modules in the sparse structure can be recognized as the elementary cliques. Notably, the presence of a large clique increases the efficiency of building

TABLE I. Graph measures of the assemblies of cliques of the size n with $N \approx 1000$ vertices for three representative values of the affinity parameter ν: The number of edges E, average degree $\langle k \rangle$, and clustering coefficient $\langle Cc \rangle$, graph's modularity mod, diameter D, and ratio of the hyperbolicity parameter $\delta_{\max}$ to $D/2$. Two bottom rows are for mixed clique sizes $n \in [3, 6]$ distributed according to $\propto n^{-\alpha}$.

bb	ν	E	$\langle k \rangle$	$\langle Cc \rangle$	$\langle \ell \rangle$	mod	D	$\delta_{\max}/D/2$
$n = 3$	−5	1501	2.999	0.766	9.789	0.928	22	1/11
	0	1734	3.465	0.741	7.265	0.902	16	1/8
	+5	1991	3.982	0.735	4.958	0.861	10	1/5
$n = 4$	−5	2009	4.61	0.847	8.718	0.927	19	2/19
	0	2426	4.852	0.808	6.023	0.895	12	1/6
	+5	2984	5.968	0.813	3.23	0.715	8	1/4
$n = 5$	−5	2514	5.013	0.878	8.89	0.921	19	2/19
	0	3182	6.351	0.829	5.01	0.856	11	2/11
	+5	3997	7.958	0.850	2.703	0.850	5	2/5
$\alpha = 2$	+5	2905	5.810	0.820	3.172	0.620	7	2/7
$\alpha = 0$	+5	3464	6.298	0.844	2.857	0.569	6	1/3

a nontrivial structure, even for a small attractive potential; cf. Fig. 2 top. We will further discuss the community structure of these networks in connection with their spectral properties in Sec. IV. As explained in the Introduction, we analyze the standard Laplacian operator, which is related to the adjacency matrix of the graph, i.e., a 1-skeleton of the simplicial complex. (A study of combinatorial Laplacians associated with higher-order structures remains out of the scope of this work.) Therefore, we examine the graph's properties that can be related to the Laplacian spectra. In Table I we summarize different graph measures of some monodisperse assemblies whose spectral properties are studied in Sec. IV. We note that the self-assembly process of cliques can result in a broad range of the degree of vertices. Depending on the size of cliques $n \geqslant 3$, several hubs and a power-law tail can appear at the sufficiently strong attraction between them [41]. For illustration, Fig. 3(a) shows the ranking distribution of the degree for several monodisperse assemblies in the case of intense attraction. To get an insight into the structure of the simplicial complexes of these assemblies, we show in Fig. 3(b), how the population f_q of cliques and faces along different topological levels q varies with the size of the building block n. For comparison, we also display f_q in the case of the size $n \in [3, 6]$ distributed as $\sim n^{-\alpha}$ with a small number of larger cliques ($\alpha = 2$) and the statistically similar number of cliques of all sizes ($\alpha = 0$).

As mentioned above, the assemblies of cliques possess a negative curvature in the graph metric space, which implies that they fulfill the Gromov four-point hyperbolicity criterion [54]. More precisely, the graph G is hyperbolic *iff* there is a constant $\delta(G)$ such that for any four vertices (a, b, c, d), the relation $d(a, b) + d(c, d) \leqslant d(a, d) + d(b, c) \leqslant d(a, c) + d(b, d)$ implies that

$$\delta(a, b, c, d) = \frac{d(a,c) + d(b,d) - d(a,d) - d(b,c)}{2} \leqslant \delta(G), \tag{2}$$

where $d(u, v)$ indicates the shortest path distance. Note that the difference in (2) is bounded from above by the minimum distance in the smallest sum $d_{\min} \equiv \min\{d(a, b), d(c, d)\}$.

Thus, by plotting $\delta(a, b, c, d)$ against $d_{\min}$ for a large number of 4-tuples, we numerically estimate $\delta(G) \equiv \delta_{\max}$ as the maximum value observed in the entire graph.

As described in Sec. II, the cliques aggregate by sharing their faces, i.e., cliques of a lower order, which leads to some specific properties of the grown structures [41]. In particular, the order of the simplicial complex cannot exceed the order of the largest attaching clique. Moreover, theoretical investigations of these types of structures predict [55–58] that
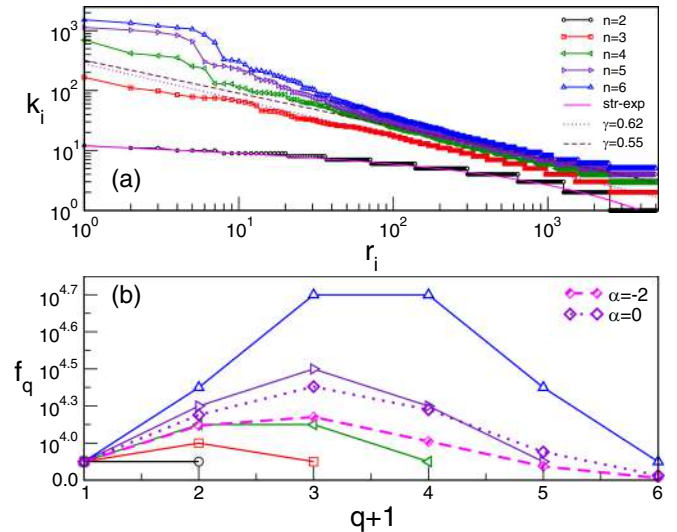
FIG. 3. (a) Ranking distribution of nodes $i = 1, 2, \ldots, 5000$ according to the decreasing degree. The degree k_i of the vertex i is plotted against its rank r_i for different assemblies of cliques of size n, indicated in the legend, and $\nu = 9$. Stretched exponential curve approximates the data for the random tree ($n = 2$), while the asymptotic power-law decay with the exponent γ is appropriate for $n \geqslant 3$. (b) The number of simplexes and faces f_q at different topology level q is plotted against $q + 1$ for the same monodisperse assemblies of the size n as in the top panel (the same legend applies). The additional dotted and dashed lines with diamonds are for the mixed sizes $n \in [3, 6]$ with the distribution $\sim n^{-\alpha}$ and two values of α given in the legend of panel (b).

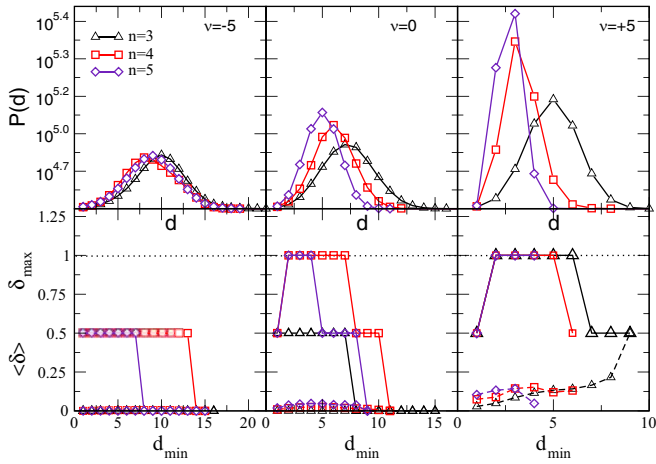FIG. 4. For the assemblies of simplexes of the size n given in the legend, and three indicated values of the chemical affinity parameter ν. Top panels show the shortest-path distance distributions $P(d)$ vs the distance d. The corresponding bottom panels display the hyperbolicity parameter $\delta_{\max}$ (upper curves, full lines) and $\langle\delta\rangle$ (lower curves, dashed lines) against $d_{\min}$. Thin dotted line indicates the level $\delta_{\max} = 1$.
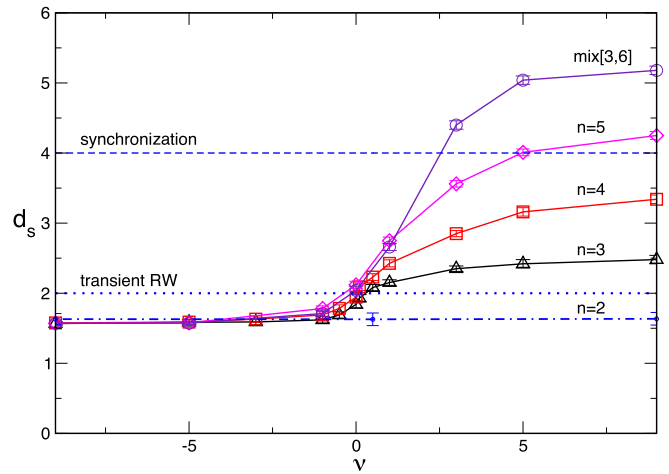

FIG. 5. The lines with different symbols represent the spectral dimension d_s plotted against chemical affinity ν for the aggregates of monodisperse cliques of sizes $n = 3, 4, 5$ and a mixture of cliques of different sizes in the range $n \in [3, 6]$. The bottom line corresponds to the random tree case $n = 2$.

the upper bound of the hyperbolicity parameter of the graph differs from the hyperbolicity of the "atoms" of the structure by at most one unit, that is, $\delta_{\max} = \delta_a + 1$. Given that a clique is ideally hyperbolic (i.e., treelike in the shortest path metric space), we have $\delta_a = 0$, which gives $\delta_{\max} = 1$ for all clique complexes grown by the rules of our model. By sampling up to 10^9 4-tuples of vertices and computing the graph hyperbolicity parameter $\delta(G)$ in Eq. (2), we demonstrate that the hyperbolicity parameter remains $\delta(G) \leqslant 1$ for all studied assemblies. More precisely, while the structure of different assemblies, as well as their distribution of the shortest-path distances, varies with the chemical affinity ν, the upper bound of their hyperbolicity parameter remains fixed in agreement with the theoretical prediction. In Fig. 4 we show the results of the numerical analysis for three representative sets of the assemblies of cliques of different sizes. See also Table I.

IV. SPECTRAL ANALYSIS OF MONODISPERSE ASSEMBLIES

Spectral dimension d_s of a graph, which is defined via $\lim_{t\to\infty} \frac{\log P_{ii}(t)}{\log t} = -\frac{d_s}{2}$, characterizes the distribution of return time $P_{ii}(t)$ of a random walk on that graph [31,65–67]. The diffusion type of processes on network is described by Laplacian operators [21,28]. More precisely, for the undirected graph of N vertices, two diffusion operators are defined, i.e., the Laplacian operator with the components

$$L_{ij} = k_i \delta_{ij} - A_{ij}, \quad (3)$$

and the symmetric normalized Laplacian [68]

$$L_{ij}^n = \delta_{ij} - \frac{A_{ij}}{\sqrt{k_i k_j}}. \quad (4)$$

Here A_{ij} are the matrix elements of the adjacency matrix, k_i is the degree of the node i, and δ_{ij} is the Kroneker symbol.

The operators defined with Eqs. (3) and (4) are symmetric and have real non-negative eigenvalues. Both operators have the eigenvalue $\lambda = 0$ with the degeneracy that is equal to the number of connected components in the network. For the networks that have a finite spectral dimension, spectral densities of both Laplacians scale as $P(\lambda) \simeq \lambda^{\frac{d_s}{2}-1}$ for small values of λ. Therefore, the corresponding cumulative distribution $P_c(\lambda)$ scales as

$$P_c(\lambda) \simeq \lambda^{\frac{d_s}{2}}, \quad \lambda \ll 1, \quad (5)$$

and it is suitable [33] for estimating the spectral dimension d_s of the network. Here we analyze the spectral properties of both Laplacian operators (3) and (4) for the networks grown with different building blocks and varied chemical affinity ν; see Figs. 5–7.

We analyze the cumulative spectral density $P_c(\lambda)$ for the Laplacian defined by the expressions (3) and (4) to determine the spectral dimension of the graphs with the adjacency matrix A_{ij}. Note that the spectrum is bounded from below, i.e., $0 \leqslant \lambda$ for all eigenvalues λ. According to Eq. (5), we estimate d_s for each sample by fitting the data of $P_c(\lambda)$ for the values in the range $\lambda \lesssim 0.3$, as illustrated in Fig. 6. The error bars are determined by taking the average from different samples of networks that have 1000 and 5000 nodes. The results summarized in Fig. 5 show how the spectral dimension of the corresponding graphs varies with the chemical affinity ν depending on the size of the elementary building blocks.

As Fig. 5 shows, the impact of the size of the cliques strongly depends on the way that they aggregate, which is controlled by the chemical affinity ν. Precisely, for the sparse structures grown under the considerable repulsion between cliques when $\nu < 0$, we find that the spectral dimension is practically independent of the size of cliques until the repulsion becomes vanishingly weak. In contrast, when $\nu \geqslant 0$ the spectral dimension increases with the size of the elementary cliques. Here the attaching cliques can share their larger faces, thus increasing the impact of the
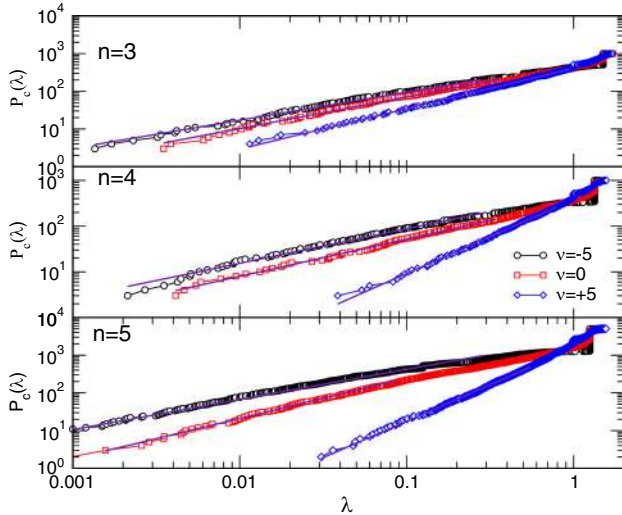
FIG. 6. Several examples of the cumulative spectral density $P_c(\lambda)$ in the range of small λ for the Laplacian operator (4) for the aggregates of triangles (top), tetrahedra (middle), and 5-cliques (bottom). In each panel, three lines in the top-to-bottom order correspond to different values of the chemical affinity $\nu = -5$, 0, and $+5$, respectively. The corresponding symbol and color for the online version are indicated in the legend.

geometrical compatibility factor. Remarkably, the spectral dimension increases with the strength of the attraction between cliques, which favors sharing increasingly larger faces. These faces are limited by the size of the elementary cliques. More specifically, for all $\nu \geqslant 0$ values, d_s is systematically larger in the aggregates of tetrahedra than those of triangles. In both cases we have that d_s exceeds the limit of the transient random walk, $d_s = 2$, for relatively weak attraction between cliques $\nu \sim 1$. However, both curves remain below $d_s = 4$ for the entire range of ν values. Note that $d_s > 4$ is recognized as the full synchronization condition for the Kuramoto oscillators on network [33], whereas, in the region $d_s \in (2, 4]$, an entrained synchronization with a complex spatiotemporal pattern can be expected [33,69]. Even though a quite compact structure is grown by attaching tetrahedrons via their triangular faces (see bottom panel in Fig. 1), its spectral dimension remains limited as $d_s < 4$, enabling the complex synchronization patterns. We find that the limit $d_s = 4$ can be exceeded when the size of the clique is at least $n = 5$ and the attraction is considerably large, i.e., $\nu \geqslant 5$. In this situation, the agglomerate consists of 5-cliques sharing many tetrahedrons as their largest faces. Interestingly, it suffices to have a few cliques of a large size to grow such agglomerates that cause the spectral dimension $d_s \geqslant 4$. For example, the mixture shown in the top panel of Fig. 2 with $n \in [3, 6]$, where the population of 6-cliques is only 1/4 of the population of 3-cliques, leads to the spectral dimension shown by the top line in Fig. 5. Furthermore, Fig. 6 indicates that not only the spectral dimension but the entire spectrum changes with the size of the cliques and the chemical affinity, as we discuss in more detail in the following.

Next, we determine the spectral density of the normalized Laplacian, defined by (4), by averaging over 10 networks of

size $N \approx 1000$ generated for the same values of the model parameters. Note that the eigenvalues of the normalized Laplacian are bounded in the range [21,28] $\lambda_{LN} \in [0, 2]$. In Fig. 7 we show the spectral density of the normalized Laplacian for several representative cases, in particular, for three different aggregates of tetrahedrons corresponding to the strong repulsion, vanishing interaction, and strong attraction, respectively. Also, in panel (e), the spectral distribution is shown for the case of strong attraction $\nu = 9$ for the cliques of different sizes $n \geqslant 3$. It should be noted that iso-spectral structures are observed in the case of the significant repulsion between the cliques $\nu = -9$. In this limit, apart from a structure at small eigenvalues, there is a prominent peak at $\lambda_{LN} = n/(n-1)$, i.e., $\lambda = -1$ in the adjacency matrix, indicating the presence of minimally connected cliques. In contrast, for $\nu \geqslant 0$, the attraction between cliques and the relevance of the geometrical compatibility factors lead to the appearance of larger simplicial complexes. A peak at $\lambda_{LN} = 1$, which is absent in panel (a), starts building at $\nu = 0$, and gradually increases with the increasing ν, as shown in panels (c) and (e). The occurrence of the peak at $\lambda_{LN} = 1$, (i.e., $\lambda = 0$ in the corresponding adjacency matrix [20,26,27]) appears as a characteristic feature of these hyperbolic networks. According to previous studies of scale-free and modular networks [21,28], this peak can be related to the nodes of the lowest degrees in the network. In the present study, such nodes are found in the bottom-right corner of the ranking distribution in Fig. 3(a). Apart from the random tree case, the appearance of this peak reflects the fact that with the increased chemical affinity a broad distribution of degrees occurs with a power-law tail; cf. Fig. 3(a). Notably, the highest peak is when the building cliques are of different sizes $n \in [3, 6]$, compared to the monodisperse structure with $n = 6$. Recently, a more insightful analysis [26,27] revealed different origins of the degeneracy in the adjacency matrix that leads to these two peaks in the spectra. More specifically, this analysis suggests that the occurrence of substructures of nodes, which are equally connected to a surrounding structure in the network, increases the degeneracy of the -1 eigenvalue. Moreover, the reasons for the degeneracy of the 0-eigenvalue were found in the nodes duplication, which is known to characterize the evolution of biological networks, notably demonstrated in analysed protein-protein interaction networks [26,27]. We note that such situations often occur in our model by combining simplexes through their shared faces. The two peaks mentioned above are increasingly more prominent in the case of larger cliques n for $\nu > 0$, where the dominant docking events occur via sharing the largest subclique. Similar spectral properties can be expected for the simplicial complexes grown by different rules, for example, in the models described in Refs. [42–47].

A further exciting feature of these spectral densities is that a characteristic minimum appears between $\lambda_{LN} = 1$ and the structure above it. The results in previous investigations [23] suggest that such minimum in the spectral density is a signature of the hierarchical organization, as demonstrated by an artificial network, which also occurs in the protein-protein interaction network. In the present study, the hierarchical organization of cliques into simplicial complexes occurring at $\nu \geqslant 0$ has been demonstrated by the algebraic topology
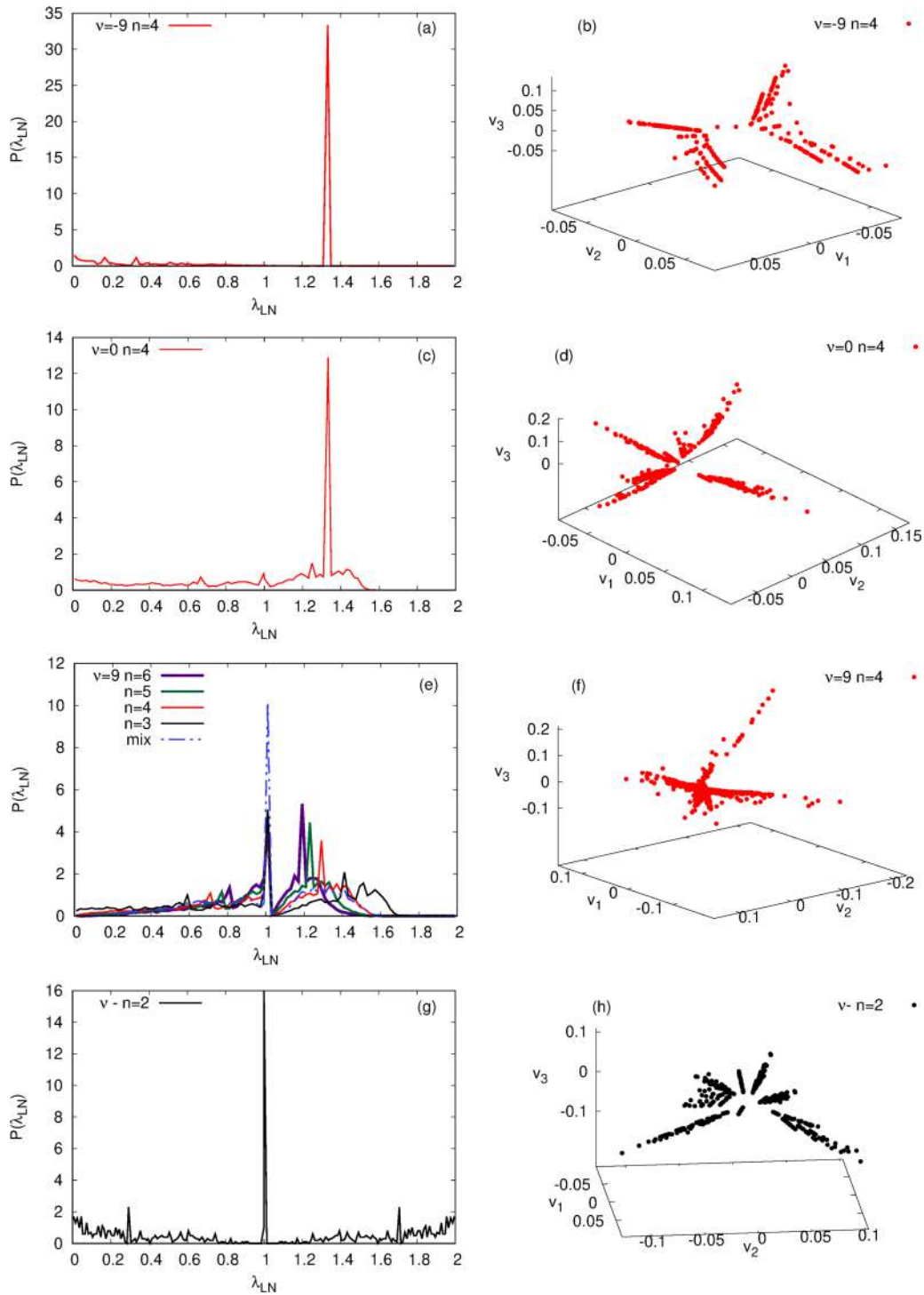
FIG. 7. Spectral distribution (left column) and the corresponding scatter plots of the eigenvectors $v1$, $v2$, $v3$ of the three lowest nonzero eigenvalues of the normalized Laplacian (right column) for the aggregates of tetrahedra $n = 4$ for three different values of the affinity parameter $v = -9$ (a, b), $v = 0$ (c), (d), and $v = +9$ (e), (f). The bottom panels (g) and (h) are for the random tree structure $n = 2$, which is independent of v. For comparison, we also show the spectra for the cliques of different sizes $n = 3$, 5, and 6, and the mixture $n \in [3, 6]$ in panel (e); the corresponding lines are explained in the legend. The orientation of each 3D plot in panels (b), (d), (f), and (h) is such that it best depicts the number and size of different branches of nonzero components of the corresponding eigenvectors.

methods in Ref. [41]. Here we show by the spectral analysis that these simplicial complexes make the inner structure of mesoscopic communities, which can be identified by the localization of the eigenvectors of the lowest nonzero eigenvalues [21]. In the right column of Fig. 7, panels (b), (d), and

(f) show the scatter plot of the three eigenvectors related to the lowest nonzero eigenvalues corresponding to the aggregates of tetrahedrons in the left column. In the limit of strong repulsion between the cliques, the modularity of the entire structure is determined by the original cliques; see Figs. 7(a) and 7(b),

whereas the larger communities with subcommunities appear for $\nu \geqslant 0$ where higher-order connections are increasingly more effective; cf. panels (d) and (f) of Fig. 7. We can expect that similar spectral properties can be found in various real-world networks with the prominent hierarchical organization mentioned in the Introduction. More precisely, the spectra of networks representing functional brain connections, protein-protein interaction networks, as well as various hyperbolic graphs with cliques emerging from the online social dynamics can have features qualitatively similar to the ones discussed above. For completeness, panels (g) and (h) of Fig. 7 show the case $n = 2$, exhibiting the spectral density of a typical random tree structure.

V. DISCUSSION AND CONCLUSIONS

We have studied topological and spectral properties of classes of hyperbolic nanonetworks grown by the cooperative self-assembly. The growth rules [41] that can be tuned by changing the parameter of chemical affinity ν enable us to investigate the role of higher-order connectivity in the properties of the emerging structure. Attaching groups of particles are parameterized by simplexes (cliques) of different sizes which share a geometrical substructure by docking along with the growing network. For the negative values $\nu < 0$, the repulsion among cliques makes them share a single node rather than an edge or a higher structure. On the other hand, $\nu \geqslant 0$ implies that the geometrical factors and the size of the attaching clique become relevant. In particular, the higher positive value of ν implies that a new clique attaches to a previously added clique by sharing its face of the larger order, thus building a more compact structure. Mathematically [58], the attachment of cliques by sharing a face (of any order) leads to simplicial complexes whose hyperbolicity parameter cannot exceed one.

Our results revealed that, while the hyperbolicity parameter remains fixed $\delta_{\max} = 1$ across different assemblies, their topological and spectral properties change with the increased chemical affinity; see Table I and Figs. 5 and 7. Remarkably, the spectral dimension of the structure of strongly repelled cliques of any size is practically indistinguishable from the one of a random tree of the same number of ver-

tices. However, the rest of the spectrum is different from the one of the tree structure; its dominant feature is the presence of cliques as the prominent network modules. On the other hand, the compelling attraction between the cliques for $\nu \gtrsim 0$ results in the spectral dimension that for all sizes $n \geqslant 3$ exceeds the limit $d_s = 2$, compatible with the transient random walk on the network. Further increase of the spectral dimension with the increased affinity parameter ν strongly depends on the size of the cliques. Our results suggest that for a strong attraction with the cliques of size $n \geqslant 5$, the spectral dimension of the network can exceed the limit $d_s = 4$, above which the synchronized phase is expected to exist [33]. However, more interesting structures are grown by smaller cliques or a mixture of different clique sizes with a weak attraction (small positive values of the parameter ν) allowing the sharing a variety of clique's faces. In these cases, we find that the spectral dimension remains in the range of $d_s \in (2, 4)$. These spectral properties are expected to be compatible with an entrained synchronization [33] or a frustrated hierarchical synchronization with intricate spatiotemporal patterns [69]. A detailed analysis of such synchronization patterns on these graphs as well as potentially superdiffusive processes [70] remains for future work. Due to their spectral properties, these structures can be interesting for modeling the complex dynamics in a variety of biological systems and for potential applications. In the framework of the cooperative self-assembly of nanoparticle groups, our analysis shows how the control of the chemical affinity can lead to complex structures with different functional properties. Furthermore, the presented results can be relevant for a deeper understanding of the functional complexity of many important structures with built-in simplicial complexes, such as human connectome [49] and other hierarchically modular networks.

ACKNOWLEDGMENTS

[1] M. A. Boles, M. Engel, and D. V. Talapin, Self-assembly of colloidal nanocrystals: From intricate structures to functional materials, Chem. Rev. **116**, 11220 (2016).

[2] Y. Wang, Yu. Wang, D. R. Breed, V. N. Manoharan, L. Feng, A. D. Hollingsworth, M. Weck, and D. J. Pine, Colloids with valence and specific directional bonding, Nature (London) **491**, 51 (2012).

[3] L. Herkert, A. Sampedro, and G. Fernández, Cooperative self-assembly of discrete metal complexes, Cryst. Eng. Commun. **18**, 8813 (2016).

[4] C. J. Meledandri, J. K. Stolarczyk, and D. F. Brougham, Hierarchical gold-decorated magnetic nanoparticle clusters with controlled size, ACS Nano **5**, 1747 (2011).

[5] D. Toulemon, M. V. Rastei, D. Schmool, J. S. Garitaonandia, L. Lezama, X. Cattoen, S. Begin-Collin, and B. P. Pichon, Enhanced collective magnetic properties induced by the controlled

assembly of iron oxide nanoparticles in chains, Adv. Funct. Mat. **26**, 2454 (2016).

[6] Y. Gu, R. Burtovyy, J. Townsend, J. R. Owens, I. Luzinov, and K. G. Kornev, Collective alignment of nanorods in thin Newtonian films, Soft Matter **9**, 8532 (2013).

[7] S. Liu and J. Yu, Cooperative self-construction and enhanced optical absorption of nanoplates-assembled hierarchical Bi2Wo6 flowers, J. Solid State Chem. **181**, 1048 (2008).

[8] D. Luo, C. Yan, and T. Wang, Interparticle forces underlying nanoparticle self-assemblies, Small **11**, 5984 (2015).

[9] A. Hirata, L. J. Kang, T. Fujita, B. Klumov, K. Matsue, M. Kotani, A. R. Yavari, and M. W. Chen, Geometric frustration of icosahedron in metallic glasses, Science **341**, 376 (2013).

[10] S. Ikeda and M. Kotani, Materials inspired by mathematics, Sci. Technology Adv. Mater. **17**, 253 (2016).

[11] B. Senyuk, Q. Liu, E. Bililign, P. D. Nystrom, and I. I. Smalyukh, Geometry-guided colloidal interactions and self-tiling of elastic dipoles formed by truncated pyramid particles in liquid crystals, Phys. Rev. E **91**, 040501(R) (2015).

[12] M. O. Blunt, M. Šuvakov, F. Pulizzi, C. P. Martin, E. Pauliac-Vaujour, A. Stannard, A. W. Rushforth, B. Tadić, and P. Moriarty, Charge transport in cellular nanoparticle networks: Meandering through nanoscale mazes, Nano Lett. **7**, 855 (2007).

[13] J. Živković and B. Tadić, Nanonetworks: The graph theory framework for modeling nanoscale systems, Math. Quantum Nanotechnol. **2**, 30 (2013).

[14] M. Šuvakov and B. Tadić, Modeling collective charge transport in nanoparticle assemblies, J. Phys.: Condens. Matter **22**, 163201 (2010).

[15] M. Šuvakov and B. Tadić, Transport processes on homogeneous planar graphs with scale-free loops, Physica A **372**, 354 (2006).

[16] B. Tadić, M. Andjelković, and M. Šuvakov, The influence of architecture of nanoparticle networks on collective charge transport revealed by the fractal time series and topology of phase space manifolds, J. Coupled Syst. Multiscale Dyn. **4**, 30 (2016).

[17] G.-M. Weng *et al.*, Layer-by-layer assembly of cross-functional semi-transparent MXene-carbon nanotubes composite films for next-generation electromagnetic interference shielding, Adv. Funct. Materials **28**, 1803360 (2018).

[18] C. Castellano and R. Pastor-Satorras, Relating Topological Determinants of Complex Networks to Their Spectral Properties: Structural and Dynamical Effects, Phys. Rev. X **7**, 041024 (2017).

[19] C. Sarkar and S. Jalan, Spectral properties of complex networks, Chaos **28**, 102101 (2018).

[20] J. F. Lutzeyer and A. T. Walden, Comparing graph spectra of adjacency and Laplacian matrices, arXiv:1712.03769.

[21] M. Mitrović and B. Tadić, Spectral and dynamical properties in classes of sparse networks with mesoscopic inhomogeneities, Phys. Rev. E **80**, 026123 (2009).

[22] E. Andreotti, D. Remondini, G. Servizi, and A. Bazzani, On the multiplicity of Laplacian eigenvalues and Fiedler partitions, Linear Algebra Appl. **544**, 206 (2018).

[23] M. A. M. de Aguiar and Y. Bar-Yam, Spectral analysis and the dynamic response of complex networks, Phys. Rev. E **71**, 016106 (2005).

[24] G. Palla and G. Vattay, Spectral transitions in networks, New J. Phys. **8**, 307 (2006).

[25] S. Jalan and A. Yadav, Assortative and disassortative mixing investigated using the spectra of graphs, Phys. Rev. E **91**, 012813 (2015).

[26] A. Yadav and S. Jalan, Origin and implications of zero degeneracy in networks spectra, Chaos **25**, 043110 (2015).

[27] L. Marrec and S. Jalan, Analysing degeneracies in networks spectra, Europhys. Lett. **117**, 48001 (2017).

[28] S. N. Dorogovtsev, A. V. Goltsev, J. F. F. Mendes, and A. N. Samukhin, Spectra of complex networks, Phys. Rev. E **68**, 046109 (2003).

[29] O. Shpielberg, T. Nemoto, and J. Caetano, Universality in dynamical phase transitions of diffusive systems, Phys. Rev. E **98**, 052116 (2018).

[30] A. Arenas, A. Díaz-Guilera, J. Kurths, Y. Moreno, and C. Zhou, Synchronization in complex networks, Phys. Rep. **469**, 93 (2008).

[31] B. Durhuus, T. Jonsson, and J. F. Wheater, The spectral dimension of generic trees, J. Stat. Phys. **128**, 1237 (2007).

[32] I. Seroussi and N. Sochen, Spectral analysis of a non-equilibrium stochastic dynamics on a general network, Sci. Rep. **8**, 14333 (2018).

[33] A. P. Millán, J. J. Torres, and G. Bianconi, Synchronization in network geometries with finite spectral dimension, Phys. Rev. E **99**, 022307 (2019).

[34] R. H. Atkin, From cohomology in physics to q-connectivity in social science, Int. J. Man-Machine Studies **4**, 139 (1972).

[35] P. Gould, Q-analysis, or a language of structure: An introduction for social scientists, geographers, and planners, Int. J. Man-Machine Studies **13**, 169 (1980).

[36] H.-J. Bandelt and V. Chepoi, Metric graph theory and geometry: A survey, in *Surveys on Discrete and Computational Geometry: Twenty Years Later*, Vol. 453, edited by J. E. Goodman, J. Pach, and R. Pollack (Providence, RI: AMS, 2008).

[37] J. Jonsson, *Simplicial Complexes of Graphs*, Lecture Notes in Mathematics Vol. 1928 (Springer-Verlag, Berlin, 2008).

[38] S. Maletić and Y. Zhao, *Simplicial Complexes in Complex Systems: The Search for Alternatives* (Harbin Institute of Technology, Harbin, Peoples Republic of China, 2017).

[39] J. U. Surjadi, L. Gao, H. Du, X. Li, X. Xiong, N. X. Fang, and Y. Lu, Mechanical metamaterials and their engineering applications, Adv. Eng. Mater. **21**, 1800864 (2019).

[40] X. Sun and Y. Q. Ma, Self-Assembly of heterogeneous nano-chain and nano-sheet through the "breaking-to-assembling" route, Ceram. Int. **44**, 23305 (2018).

[41] M. Šuvakov, M. Andjelković, and B. Tadić, Hidden geometries in networks arising from cooperative self-assembly, Sci. Rep. **8**, 1987 (2018).

[42] G. Bianconi and C. Rahmede, Emergent hyperbolic network geometry, Sci. Rep. **7**, 41974 (2017).

[43] G. Bianconi and C. Rahmede, Network geometry with flavor: From complexity to quantum geometry, Phys. Rev. E **93**, 032315 (2016).

[44] O. T. Courtney and G. Bianconi, Weighted growing simplicial complexes, Phys. Rev. E **95**, 062301 (2017).

[45] D. C. da Silva, G. Bianconi, R. A. da Costa, N. S. Dorogovtsev, and J. F. F. Mendes, Complex network view of evolving manifolds, Phys. Rev. E **97**, 032316 (2018).

[46] G. Petri and A. Barrat, Simplicial Activity Driven Model, Phys. Rev. Lett. **121**, 228301 (2018).

[47] N. Cinardi, A. Rapisarda, and G. Bianconi, Quantum statistics in network geometry with fractional flavor, arXiv:1902.10035v1.

[48] B. Tadić, M. Andjelković, B. M. Boshkoska, and Z. Levnajić, Algebraic topology of multi-brain connectivity networks reveals dissimilarity in functional patterns during spoken communications, PLoS ONE **11**, e0166787 (2016).

[49] B. Tadić, M. Andjelković, and R. Melnik, Functional geometry of human connectome and robustness of gender differences, arXiv:1904.03399 [q-bio.NC].

[50] M. Andjelković, B. Tadić, S. Maletić, and M. Rajković, Hierarchical sequencing of online social graphs, Physica A **436**, 582 (2015).

[51] M. Andjelković, B. Tadić, M. M. Dankulov, M. Rajković, and R. Melnik, Topology of innovation spaces in the knowledge networks emerging through questions-and-answers, PLoS ONE **11**, e0154655 (2016).

[52] B. Tadić, Self-organised criticality and emergent hyperbolic networks: Blueprint for complexity in social dynamics, Eur. J. Phys. **40**, 024002 (2019).

[53] M. Andjelković, N. Gupte, and B. Tadić, Hidden geometry of traffic jamming, Phys. Rev. E **91**, 052817 (2015).

[54] J. M. Rodriguez and E. Touris, Gromov hyperbolicity through decomposition of metric spaces, Acta Math. Hungar. **103**, 107 (2004).

[55] S. Bermudo, J. M. Rodríguez, J. M. Sigarreta, and J.-M. Vilaire, Gromov hyperbolic graphs, Discrete Math. **313**, 1575 (2013).

[56] S. Bermudo, J. M. Rodríguez, O. Rosario, and J. M. Sigarreta, Small values of the hyperbolicity constant in graphs, Discrete Math. **339**, 3073 (2016).

[57] A. Martínez-Pérez, Generalized chordality, vertex separators, and hyperbolicity on graphs, arXiv:1708.06153v1.

[58] N. Cohen, D. Coudert, G. Ducoffe, and A. Lancin, Applying clique-decomposition for computing Gromov hyperbolicity, Theor. Comput. Sci. **690**, 114 (2017).

[59] W. Carballosa, D. Pestana, J. M. Rodríguez, and J. M. Sigarreta, Distortion of the hyperbolicity constant in minor graphs, Electron. Notes Discr. Math. **46**, 57 (2014).

[60] W. S. Kennedy, I. Saniee, and O. Narayan, On the hyperbolicity of large-scale networks and its estimation, in *2016 IEEE International Conference on Big Data (Big Data), Washington,*

DC, December 5–8, 2016 (IEEE, Computer Society, 2016), pp. 3344–3351.

[61] V. Salnikov, D. Cassese, and R. Lambiotte, Simplicial complexes and complex systems, Eur. J. Phys. **40**, 014001 (2018).

[62] M. Šuvakov and B. Tadić, Topology of Cell-Aggregated Planar Graphs, in *International Conference on Computational Science*, edited by V. N. Alexandrov, G. D. van Albada, P. M. A. Sloot, and J. Dongarra, Lecture Notes in Computer Science Vol. 3993 (Springer, Berlin, Heidelberg, 2006), pp. 1098–1105.

[63] M. Šuvakov, M. Andjelković, and B. Tadić, Applet: Simplex aggregated growing graph, http://suki.ipb.rs/ggraph/ (2017).

[64] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, Fast unfolding of communities in large networks, J. Stat. Mech. (2008) P10008.

[65] R. Rammal and G. Toulouse, Random walks on fractal structures and percolation clusters, J. Phys. Lett. **44**, 13 (1983).

[66] R. Burioni and D. Cassi, Universal Properties of Spectral Dimension, Phys. Rev. Lett. **76**, 1091 (1996).

[67] R. Burioni and D. Cassi, Random walks on graphs: Ideas, techniques, and results, J. Phys. A **38**, R45 (2005).

[68] A. N. Samukhin, S. N. Dorogovtsev, and J. F. F. Mendes, Laplacian spectra of, and random walks on, complex networks: Are scale-free architectures really important, Phys. Rev. E **77**, 036115 (2008).

[69] P. Villegas, P. Moretti, and M. A. Muñoz, Frustrated hierarchical synchronization and emergent complexity in the human connectome network, Sci. Rep. **4**, 5990 (2014).

[70] B. Tadić and S. Thurner, Information superdiffusion on structured networks, Physica A **332**, 566 (2004).

ORIGINAL RESEARCH

Nitrogen plasma surface treatment for improving polar ink adhesion on micro/nanofibrillated cellulose films

Katarina Dimic-Misic · Mirjana Kostić · Bratislav Obradović ·
Ana Kramar · Stevan Jovanović · Dimitrije Stepanenko · Marija Mitrović-Dankulov ·
Saša Lazović · Leena-Sisko Johansson · Thad Maloney · Patrick Gane

Abstract We find that nitrogen plasma treatment of micro/nanofibrillated cellulose films increases wettability of the surface by both liquid polar water and nonpolar hexadecane. The increased wetting effect is more pronounced in the case of polar liquid, favouring the use of plasma treated micro/nanofibrillated cellulose films as substrates for a range of inkjet printing including organic-based polar-solvent inks. The films were formed from aqueous suspensions of progressively enzymatic pretreated wood-free cellulose fibres, resulting in increased removal of amorphous species producing novel nanocellulose surfaces displaying increasing crystallinity. The mechanical properties of each film are shown to be highly dependent on the enzymatic pretreatment time. The change in surface chemistry arising from exposure to nitrogen plasma is revealed using X-ray photoelectron spectroscopy. That both polar and dispersive surface energy components become increased, as measured by contact angle, is also linked to an increase in surface roughness. The change in surface free energy is exemplified to favour the trapping of photovoltaic inks.

Keywords DBD plasma · Nitrogen plasma surface treatment · Nanocellulose films · Enzymatic nanocellulose · Printing of organic-based polar inks

K. Dimic-Misic (✉) · L.-S. Johansson ·
T. Maloney · P. Gane
Department of Bioproducts and Biosystems, School of
Chemical Engineering, Aalto University,
00076 Aalto, Helsinki, Finland
e-mail: katarina.dimic.misic@aalto.fi

M. Kostić · A. Kramar
Faculty of Technology and Metallurgy, University of
Belgrade, Karnegijeva 4, Belgrade 11000, Serbia

B. Obradović
Faculty of Physics, University of Belgrade, Studentski trg
12, Belgrade 11001, Serbia

S. Jovanović · D. Stepanenko · M. Mitrović-Dankulov ·
S. Lazović
Institute of Physics Belgrade, University of Belgrade,
Pregrevica 118, Belgrade 11080, Serbia

Introduction and background

Sustainability is one of the key targets for industrial practice today. The related research aimed at new biobased materials derived from renewable sources, is relevant for the sustainable economy. In the bioproducts industry, micro/nanofibrillated cellulose (MNFC) has attracted attention in a number of potential applications (Hubbe et al. 2017a). It can be used in standard wood products, such as paper and boards. However, most of the benefits derived from MNFC stem from its wider uptake in a range of industrial value chains, such as biodegradable packaging films and laminates. MNFC has interesting intrinsic properties derived from large specific surface area and its

alternate regions of crystallinity. The hydroxylated surface chemistry is readily suitable for chemical modification. Films formed from MNFC are considered smart materials and studied for functional materials applications. Enzyme-treated fibres used to produce cellulose nanofibrils provide higher crystallinity in the resulting nanocellulose, as enzymes digest amorphous cellulose, which acts as the glue between crystalline cellulose regions. Direct hydrogen bonding of crystalline cellulose, therefore, gives a stronger material film. An example of an important application of MNFC is as a substrate for printed solar cells based on organic inks (Zhu et al. 2014). The surface properties of MNFC films, such as wettability by liquid, topography, chemistry, surface charge, the presence of hydrophobic and hydrophilic domains, density and conformation of functional groups, all play a crucial role in printability and barrier properties. Their ability to support controlled migration of solvent ink vehicle and chromatographic differentiation of ink components is important in the printing of inkjet printable (IP) inks, and especially for production of bio-based printed functionality in a wide range of applications, such as printed electronics and printed diagnostics (Hoeng et al. 2016; Jutila et al. 2018).

Solar panel IP photovoltaic (PV) inks contain a complex mix of materials, including the organic electron acceptor (p-type) and negative electron donor (n-type) suspended in solvent together with specific surfactant(s) intended to keep the p-type and n-type components de-mixed (Kumar and Chand 2012). Although drop-on-demand (DoD) inkjet printing is a very competitive candidate for printing PV inks on film substrates, there are limitations in respect to mutual compatibility between the surface of MNFC films and mixed polar-dispersive solvents constituting the PV ink (Singh et al. 2010; Yinhua et al. 2013). Electrolyte is highly polar, for example, and so sufficient wettability is needed by providing a polar surface, despite the parallel requirement for wettability by organic species (Schultz et al. 1977; Özkan et al. 2016). This complex polar-dispersive surface energy balance is, therefore, critical (Hansson et al. 2011).

Exposure to plasma is a convenient method to modify the surface properties of polymeric materials, while keeping their bulk properties intact, making a material better adapted for printing (Möller et al. 2010; Kramer et al. 2006; Catia et al. 2015). Furthermore, as we demonstrate, it is a convenient way to introduce desired groups onto the surface of materials (Mihailovic et al. 2011). Surface properties depend on parameters of plasma treatment such as applied electrical field energy, type of feed gas, pressure, exposure time, and reactor geometry (van de Vyver et al. 2011; Jun et al. 2008).

In this work, we modify enzyme pretreated fibre-derived MNFC film surfaces using nitrogen plasma to enhance their amphiphilic surface affinity to polar and non-polar IP PV inks. Measurements of the surface free energy, surface roughness (atomic force microscopy (AFM)) and material composition [X-ray photoelectron spectroscopy (XPS)] were used to characterise the MNFC film surface before and after plasma treatment. The affinity for IP PV ink was assessed visually after inkjet printing. We also identify a correlation between the observed change in free surface energy of the MNFC film, arising from the plasma treatment, with the effect of the enzymatic pretreatment. This is related to the level of residual crystallinity increasing as a function of progressive enzymatic pretreatment (Galagan et al. 2011; Cernakova et al. 2006; Pertile et al. 2010; Vanneste et al. 2017).

To meet the requirement of sufficient tensile strength of MNFC films for the application exemplified, the rheological properties of enzymatically pretreated MNFC fibrillar suspensions were compared with the mechanical properties of corresponding obtained films, so that rheology can be used as a predictor of film strength (Maloney 2015; Zhu et al. 2014).

Materials and methods

Preparation of MNFC

For the manufacture of short MNFC fibrils, the pulp was first washed to create the sodium form by adding sodium hydroxide to a 2 w/w% fibre suspension until the pH reached 10, and then re-washed with deionised water to a conductivity of 8.2 μS. The enzymatic treatment was performed with a commercial enzyme ECOPULP® R (Ecopulp Finland Oy), produced by a genetically modified strain of Trichoderma reesei fungus (Rantanen et al. 2015). The activity properties of the enzyme are reported to be 17,700 nkat cm^{-3} cellulase with a protein level of 93 mg cm^{-3}

(Willberg-Keyriläainen et al. 2019). An amount of 3 mg of enzyme per gram of pulp fibre was added to a 2.5 w/w% suspension and the temperature was increased to 57 °C at pH 5.5 during hydrolysis, whilst keeping under constant agitation. The period of digestion was increased for each subsequent sample in 30 min steps, Table 1. The enzymatic activity was terminated by adjusting the pH to 9–10 by sodium carbonate and increasing the temperature to 90 °C. After cooling the suspension overnight in cold storage, the samples were refined using an homogeniser (model M-110P, Microfluidics, USA), passing the material under a pressure of 2000 bar through a 100 μm flow gap. The solids content of the MNFC suspension after the fluidisation was 1.65 w/w%.

The enzymatic pretreatment of pulp as a route for producing low-charged MNFC resulted in the production of short fibrils, which, in the case studied here, have much lower aspect ratio than MFC and NFC produced via chemical oxidative pretreatment or mechanical refining alone, as illustrated in Fig. 1 comparing MNFC/300/and MNFC/0/suspensions (Table 1), revealing much shorter fibrils obtained upon 300 min of enzymatic hydrolysis.

MNFC film preparation

With increasing enzymatic treatment time, the resulting MNFC suspension viscosity decreased significantly, and the solid content for preparation of the respective films ranged from 0.6 to 1.9 w/w% to meet the target film grammage of 60 g m^{-2} produced under conditions of 23 °C and relative humidity (RH) 50%.

Films were made on a sheet-former according to ISO standard 5269-1, with some modification of the screen to aid fines retention. Due to the very strong water retention of MNFC, and its fine size, a polyamide monofilament open mesh fabric SEFAR NITEX® 03-1/1 with a pore size of 1 μm was placed on top of a 125 μm metal screen. The pulp suspension was poured at high viscosity onto the former without adding water or stirring the slurry. The system was pressurised to 0.3 bar and the sealing lid was used on the sheet-former. Double-sided adhesive tape, of 5 mm width, was attached to the edges of the drying plate between plate and formed film, with purpose of fixing the edge of the film to prevent it shrinking during drying (Fig. 2).

Material treatment and characterisation

Optical microscopy was used to study the fibrillar sample suspensions and films using an Olympus BX 61 microscope equipped with a DP12 camera.

Water retention the water retention value (WRV) of the MNFC was determined in accordance to the standard SCAN-C 102XE with a slight modification in that 10 w/w% suspension of the MNFC was added in various ratios to a suspension of bleached unrefined pulp. The pulp matrix helps the MNFC dewater and remain retained on the screen. The WRV of neat MNFC can be evaluated by extrapolating to zero pulp, not including the swelling of the pulp fibres (Möller et al. 2010). The experiment was performed in triplicate for each sample.

Dielectric barrier discharge (DBD) plasma operates in a thermodynamically non-equilibrium condition (so-called cold plasma) in which the ion and molecular translational temperature is much lower than the electron temperature, such that excessive gas heating can be suppressed (Kostic et al. 2009; Prysiazhnyi et al. 2013). The advantage is that the plasma can be generated at atmospheric pressure, either in open or closed environment. In an open atmosphere, the plasma discharges can be produced with a gas flow between the electrodes (Mihailovic et al. 2011; Chu et al. 2002; Jens et al. 2017).

Table 1 Materials used in this study: bleached hardwood Kraft pulp treated with enzymes under controlled conditions, with progressive increase in enzymatic digestion time by 30 min steps for each subsequent sample

Enzymatic treatment time/ min	0 (reference)	30	60	90	120	150	180	210	240	270	300
Sample label	MNFC/0/	MNFC/ 30/	MNFC/ 60/	MNFC/ 90/	MNFC/ 120/	MNFC/ 150/	MNFC/ 180/	MNFC/ 210/	MNFC/ 240/	MNFC/ 270/	MNFC/ 300/
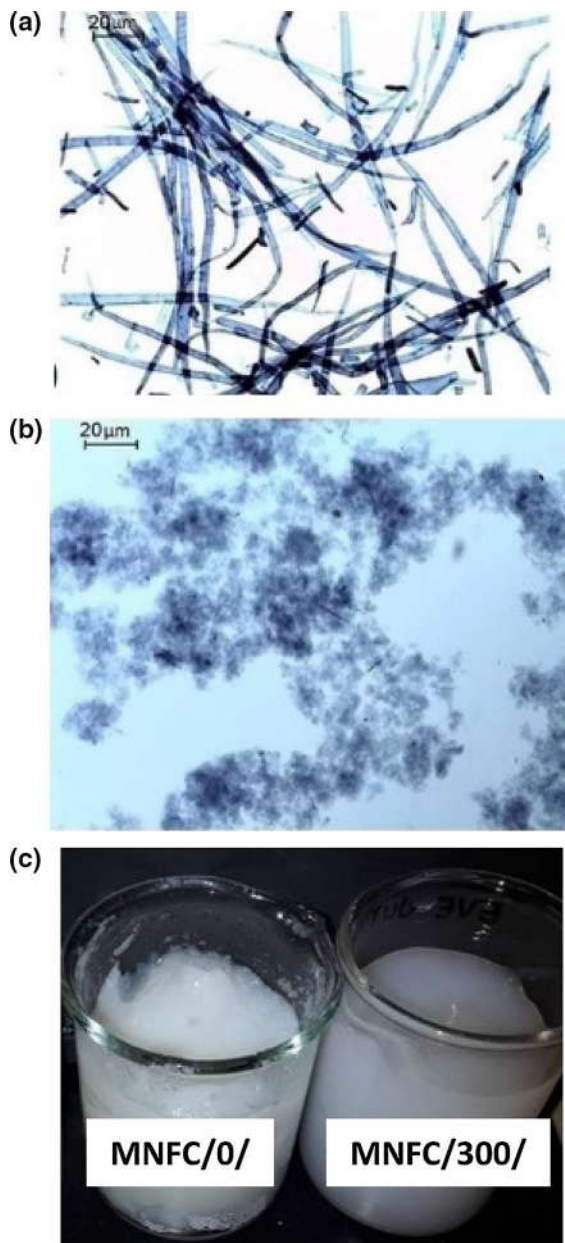
Fig. 1 Images of fibrils sample suspensions obtained with optical microscopy revealing the effect of processing conditions on the fibril size and aspect ratio: **a** without enzymatic treatment produced MNFC/0/yielding long fibrils, **b** MNFC/300/short, low aspect ratio fibrils, and **c** displaying the corresponding 2 w/w % MNFC suspensions of MFC/0/and MNFC/300/. The difference in gelation strength is due to the different size of fibrils and corresponding amount of water dispersed within the fibrillar matrix
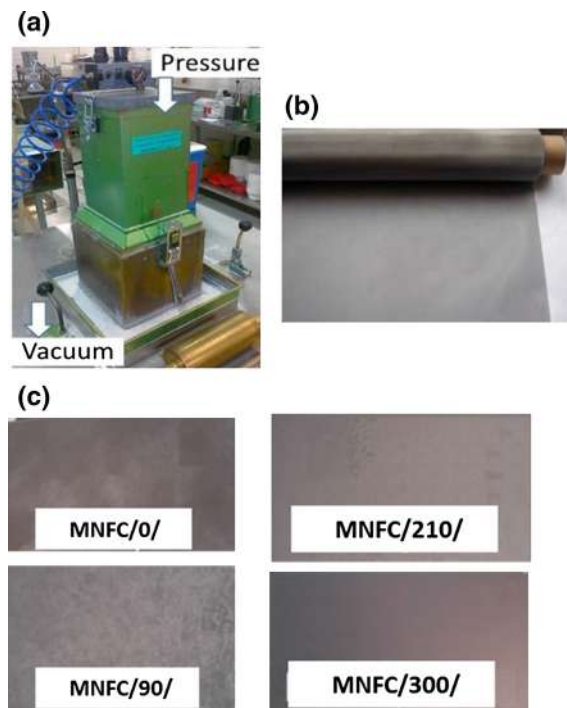
Fig. 2 MNFC film preparation: **a** sheet forming device with **b** 10 μm mesh supplemented nylon screen, and **c** samples of cut-offs (60×15 mm^2) from MNFC films produced from pulp refined with different enzymatic pretreatment time (Table 1). Transparency and uniformity of films increases with hydrolysis time

A further attractive characteristic of the DBD plasma at atmospheric pressure is that it can be used to modify or activate surfaces of a wide range of materials, from polymers, textile fibres to biological tissues, without damaging them (Kostic et al. 2009; Pertile et al. 2010; Mihailovic et al. 2011). To generate the DBD plasma we used a home-made device built at the Faculty of Physics, University Belgrade, Fig. 3. The DBD is assembled in a chamber with nitrogen gas injected into the discharge volume (6 dm^3 min^{-1}) through ten equidistant holes to ensure homogeneous gas flow. MNFC films were treated for 0 s, 30 s and 60 s, respectively. The device was operated at 6 kV DC and 300 electric field pulses per second (Hz) for the prescribed durations of time, for all the films, as a higher voltage resulted in burning of the thin MNFC films, especially for those made from pulp exposed to long enzymatic pretreatment time.

Fig. 3 DBD device with two electrodes and sample placed between them: **a** schematic illustration of DBD plasma devise, **b** plasma chamber housing the sample placed 1 mm from the upper electrode, and **c** closed plasma set up with glass lid placed above the top of the upper electrode
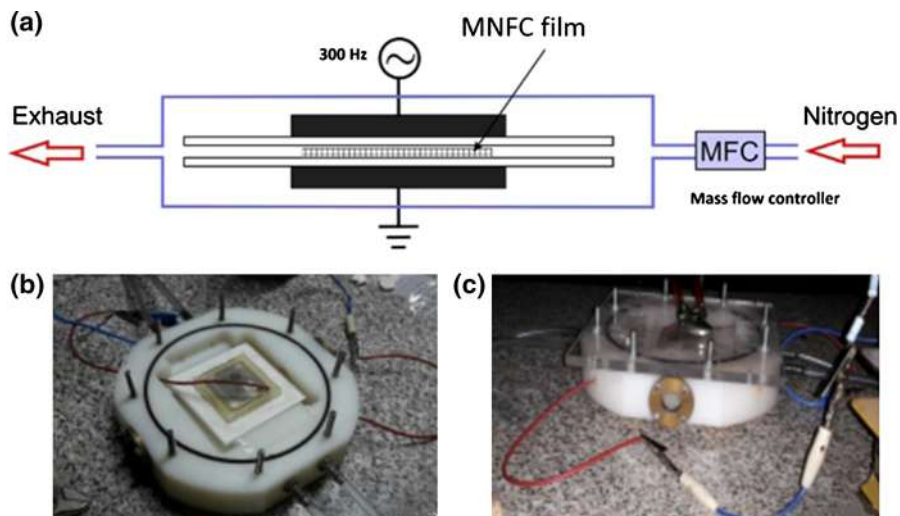
Determination of free surface energy (FSE) components

For the evaluation of any change in free surface energy of MNFC films arising from nitrogen plasma treatment, the contact angle (CA) is determined.

Most liquids are rapidly spreading on a high energy surface, and so a representative contact angle (CA) cannot be readily measured, Schultz et al. (1977) developed a method where CA can be measured by submerging the surface in one liquid and using a second liquid to measure the contact angle. In this case a hydrocarbon *n*-hexadecane is used as the submerging liquid having the purely dispersive liquid–vapour surface tension of $\gamma_{LV}^{h} = 27.4$ mJ m^{-2}, much lower than the expected surface free energy of the MNFC samples, and water as the contact angle liquid with the highly polar liquid–vapour surface tension $\gamma_{LV}^{w} = 72.8$ mJ m^{-2} (Hansson et al. 2011). A sessile drop of water is lowered into contact with the horizontal film immersed under hexadecane using a precise pipette delivering 70 µl of liquid and the progressive change in drop shape due to the change in CA recorded with a Nikon camera (D5000) in time steps of 1 ms. The CA of water is also recorded separately to represent the print challenge of a highly polar ink (Özkan et al. 2016; Dimic-Misic et al. 2015). For each given MNFC sample and given liquid data variation is within 10%. The identification of contact line geometry and evaluation of CA uses numeric software tools, as presented visually in Fig. 4. For a parallel optimal method for polar FSE determination

with water alone, the Girifalco and Good approach (1957), combined with the Neumann equation of state was used. This latter allowed the polar contribution to FSE be estimated and thus can be added to the formerly measured dispersive component. Each measurement was conducted five times. For each given MNFC sample, the relative error of measured FSE was shown to be ∼ 10%.

Surface topography

Plasma action on the film surface can lead to a degree of debonding of fibrils as well as electrostatic charging and potential for subsequent additional moisture adsorption. Such changes can lead to re-conformation of the surface, even though no mechanical forces have been applied (Kostic et al. 2009; Chu et al. 2002). The change in topography of the MNFC films was investigated by Atomic Force Microscopy (AFM) (Veeco Instruments, model Dimension V). Using a MultiMode 8 with Bruker NanoScope V controller. Each MNFC film sample was dry-cast onto a Mica support for AFM imaging. Micrographs were obtained in trapping mode under ambient conditions, using TAP 300 tips (resonant frequency 300 kHz, line force being kept constant at 40 Nm^{-1} and the AFM images were processed and analysed with the Bruker NanoScope Analysis 1.5 software.
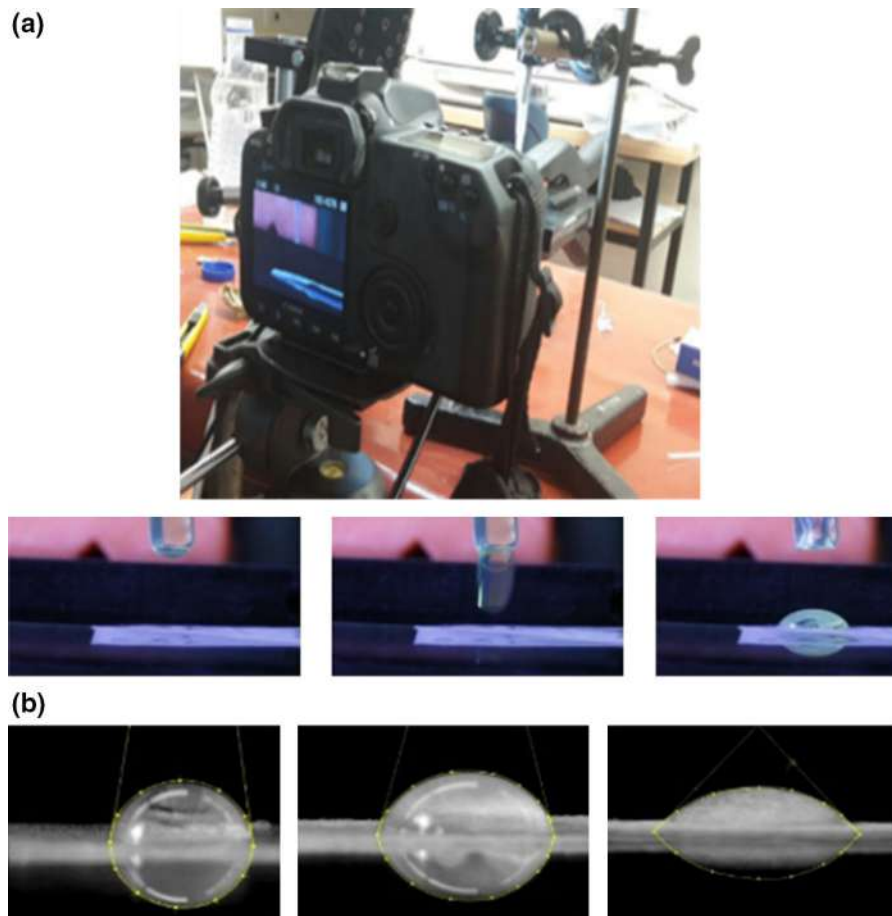
Fig. 4 Set-up for evaluating water CA under *n*-hexadecane with high speed camera (Nikon D5000): **a** images of films on camera viewfinder and **b** image processing of drop spreading (see also Fig. 8)

Mechanical properties

Mechanical properties of the MNFC films were measured by an MTS 400/M vertical tensile tester equipped with a 20 N load cell. The instrument was controlled by a TestWorks 4.02 program. Specimen strips with dimensions of 60×15 mm^2 were clipped from the MNFC films with a lab paper cutter (Afsahi et al. 2018). The thickness of the strips was separately measured with an L&W micrometer SE 250. The gauge length was 40 mm and the testing velocity was 0.5 mm min^{-1}. The results are presented as an average value obtained from five parallel specimens.

Surface chemical composition

Surface composition of the MNFC films was evaluated with X-ray photoelectron spectroscopy (XPS), using a

Kratos AXIS Ultra electron spectrometer, with monochromatic Al Kα irradiation at 100 W and under charge neutralisation. Both the untreated MNFC films and plasma treated specimens were analysed. For the preparation, samples were pre-evacuated for at least 12 h, after which wide area survey spectra (for elemental analysis) as well as high resolution regions of C1*s* and O1*s* were recorded from several locations, and an in situ reference of pure cellulose was recorded for each sample batch (Johansson and Campbell 2004). With the parameters used, XPS analysis was recorded on an area of 1 mm^2 and the analysis depth is less than 10 nm. Carbon high resolution data were fitted using CasaXPS and a four component Gaussian fit tailored for celluloses.

MNFC suspension rheology

The rheological properties of MNFC suspensions were analysed at 2 w/w% concentration at 23 °C with an Anton Paar MCR 300 shear rheometer. The dynamic viscosity (η) was determined by steady shear-flow measurements, using the bob-in-cup geometry (Mohtaschemi et al. 2014). Due to the potential for wall depletion (apparent slip) and thixotropic behaviour of MNFC suspensions, the "bob" was a four-bladed vane spindle with a diameter of 10 mm and a length of 8.8 mm, while the metal cup had a diameter of 17 mm. A pre-shear protocol was applied using constant shear at a shear rate $\dot{\gamma} = 100$ s^{-1} for 5 min, followed by a rest time of 10 min prior to recording the flow curves. Flow curves of MNFC suspensions were constructed under decreasing shear rate of $\dot{\gamma} = 1000–0.01$ s^{-1}, with a logarithmic spread of data points (Dimic-Misic et al. 2013). To distinguish the MNFC suspensions in terms of their colloidal interactions as an effect of hydrolysis time, aspect ratio, crystallinity and friction between nanofibrils during the flow (Pääkkönen et al. 2016; Dimic-Misic et al. 2018), the log–log plot flow curves were fitted to a power law according to the Oswald–de Waele empirical model, as shown in Eq. (1)

$$\eta = k\dot{\gamma}^{1-n} \tag{1}$$

where k and n are the flow index and the power-law exponent, respectively: $n = 1$ indicates a Newtonian fluid and $n > 1$ indicates pseudo-plastic (shear thinning) behaviour.

The Herschel–Bulkley equation describes the dynamic yield stress τ_d^0 as

$$\tau = \tau_d^0 + k\dot{\gamma}^n \tag{2}$$

where τ is the shear stress.

Printing

The photovoltaic (PV) inkjet printing inks (IP) contain a complex mix of materials, solvent and surfactants that keep the p-type and n-type components de-mixed (Hashmi et al. 2015; Özkan et al. 2016). A piezoelectric laboratory scale drop-on-demand (DoD) materials inkjet printer (Dimatix 2831-DMP) was used to test the printability of the plasma treated MNFC films (Dimic-Misic et al. 2015). The solvent of the IP ink is 3-methoxypropionitrile, which is highly polar and non-volatile (boiling point 164 °C), viscosity 1.2 mPa s and density 0.937 g cm^{-3}, as stated by the supplier, Sigma Aldrich. The surface tension measurement was performed on the ink with an optical tensiometer (CAM 200 from KSV instruments) in pendant drop mode, giving a value of 29.2 mN m^{-1} (mJ m^{-2}).

Results and discussion

The *rheological properties* of the MNFC suspensions are given in Table 2, showing the change in dewatering, dynamic yield point and flocculation/water trapping gel-like structure (consistency coefficient, k) and shear thinning properties (index, n, expressed as the positive difference $n - 1$) and change in fibre morphology expressed as the fines content using the dynamic drainage jar (DDJ).

It is clear to see that with increase in enzymatic hydrolysis time, dewatering decreases as fibrils become thinner and smaller, and suspensions become more gel-like rheologically (Rantanen et al. 2015). At the same time, crystallinity of fibrils increases and water trapping structure/flocculation within the matrix with contrasting increased mobility in the flow regime once the structure is broken (Pääkkönen et al. 2016). The dynamic yield point, the minimum stress needed to be induced to set the suspension into flow increases as the suspensions become more gel like, but, also, breakage of that suspension induces greater shear thinning as fibrils are smaller and more crystalline, orienting easily in the flow direction (Pääkkönen et al. 2016; Hubbe et al. 2017b).

The *mechanical and optical* properties of MNFC films are presented in Table 3, where it is evident that the sheet density of the films increases with increase in hydrolysis time, while the packing density of the smaller crystalline particles increases. The permeability of those films created with the finer nanofibrils obtained after 120 min hydrolysis in turn falls rapidly, and it was not possible to measure using air flow techniques. The light scattering coefficient decreases also as the packing density is increased and the amorphous parts of the cellulose fibres were reduced, while, due also to higher packing density, the elasticity modulus increases, showing that films had improved strength.

Table 2 Properties of MNFC suspensions

Enzymatic treatment time (min)	WRV ($cm^3 g^{-1}$)	Yield point, τ_d^0 (Pa)	Consistency coefficient, k (Pa s^{-n})	Shear thinning coefficient, $|(1 - n)|$	DDJ fines value (%)
MNFC suspension properties					
0	1.25	34.12	431.23	0.82	93.8
30	1.61	47.34	241.3	0.81	88.8
60	1.83	54.23	139.65	0.81	79.5
90	2.19	68.45	89.67	0.81	62.4
120	2.55	91.45	69.45	0.84	27.0
150	2.85	438.34	57.23	0.84	21.0
180	2.98	29.82	35.15	0.86	11.8
210	3.33	19.64	19.67	0.86	9.6
240	3.37	12.67	14.34	0.87	6.5
270	3.32	8.99	9.97	0.89	1.5
300	3.34	4.74	5.45	0.91	0.2

Table 3 Mechanical and optical properties of MNFC films

Enzymatic treatment time (min)	Film weight ($g\ m^{-2}$)	Density ($g\ cm^{-3}$)	Permeability [μm(Pa s)$^{-1}$]	Light scattering coefficient ($m^2\ kg^{-1}$)	E-Modulus (GPa)
Film properties					
0	73.91	0.637	69.86	37.43	2.53
30	76.12	0.794	9.96	22.83	4.16
60	71.35	0.910	1.06	16.12	5.12
90	72.31	1.016	NA	9.94	7.02
120	70.53	1.090	NA	6.93	8.59
150	70.81	1.127	NA	5.81	9.13
180	69.57	1.145	NA	4.48	8.95
210	71.08	1.178	NA	3.74	11.26
240	70.10	1.179	NA	3.08	9.17
270	71.18	1.226	NA	3.11	9.76
300	65.27	1.187	NA	3.31	10.03

Roughness colour contour and profile plots of the surface of MNFC/30/150/300 films before and after plasma treatment are presented in Fig. 5. Before plasma treatment, the roughness of the films is directional, being different in in the two measured directions (red and blue profile lines). The map for MFC/30/indicates that there are voids present between 1 and 2 μm wide, while in the case of MFC/300/the surface is flatter with less voids and of much smaller size. This means that the degree of enzyme hydrolysis directly increases the resulting smoothness due to the ever finer fibrillar elements produced, as the crystalline parts are separated due to breakdown of the amorphous constituent. After plasma treatment, the amorphous material containing surfaces, e.g. MNFC/30/, are also seen to become relatively rougher than the highly hydrolysed crystalline films, e.g. MNFC/300/. The action of the plasma is to increase voyage in the courser particulate systems, as previously described, due to effects of charge, fibril debonding etc. (Jun et al. 2008). In MNFC/30/, it is possible to identify irregular both small and large voids appearing after plasma
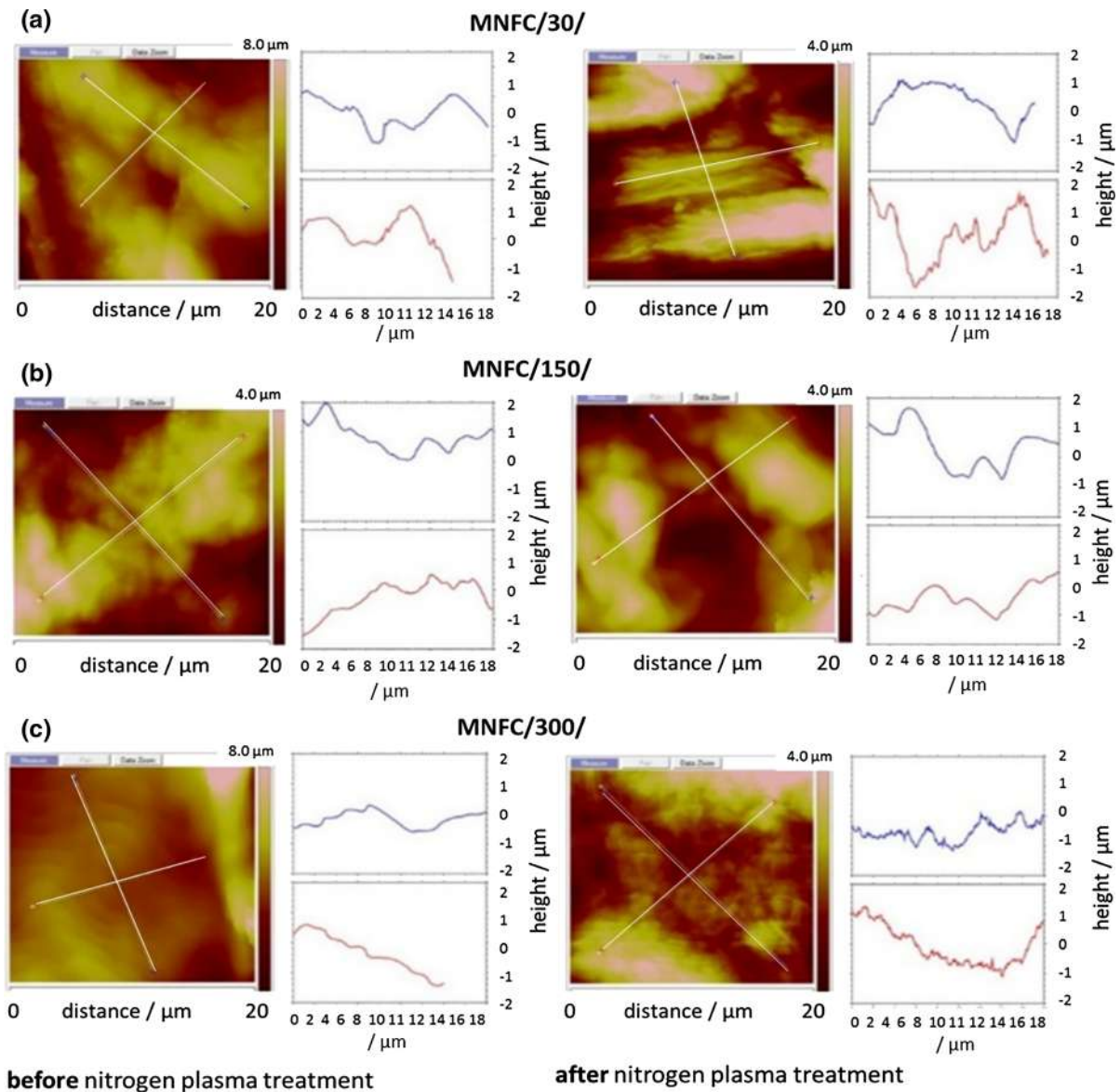
Fig. 5 Surface morphology and roughness of **a** MNFC/30/, **b** MNFC/150/ and **c** MNFC/300/before and after DBD nitrogen plasma treatment

treatment, while in MNFC/300/, the surface of the film has almost no such jagged appearance with voids only smaller than 1 μm. Nitrogen plasma treatment, thus, obviously changes the morphology of the films, on both the micro (nano) and macro level, which is likely also to have an influence on the wetting behaviour and decrease in CA due to the increased meniscus liquid–solid wetting line length (Prysiazhnyi et al. 2013; Pertile et al. 2010).

The *surface chemical species* are revealed by the XPS spectra, from which the atomic % of C–C, C–O, O–C=O and N can be derived, Fig. 6. The effect of surface modification after nitrogen plasma can be clearly seen as the level of N attachment increasing as a function of the enzymatic removal of amorphous content (Johansson and Campbell 2004). The samples with increased crystalline proportion after longer enzymatic treatment nonetheless show similar C–C bond content. Similarly, with reduction of the

Fig. 6 Surface modification obtained through XPS data showing **a** increase in N atoms at constant carbon content, and **b** change in ratio of C–O/O–C=O groups
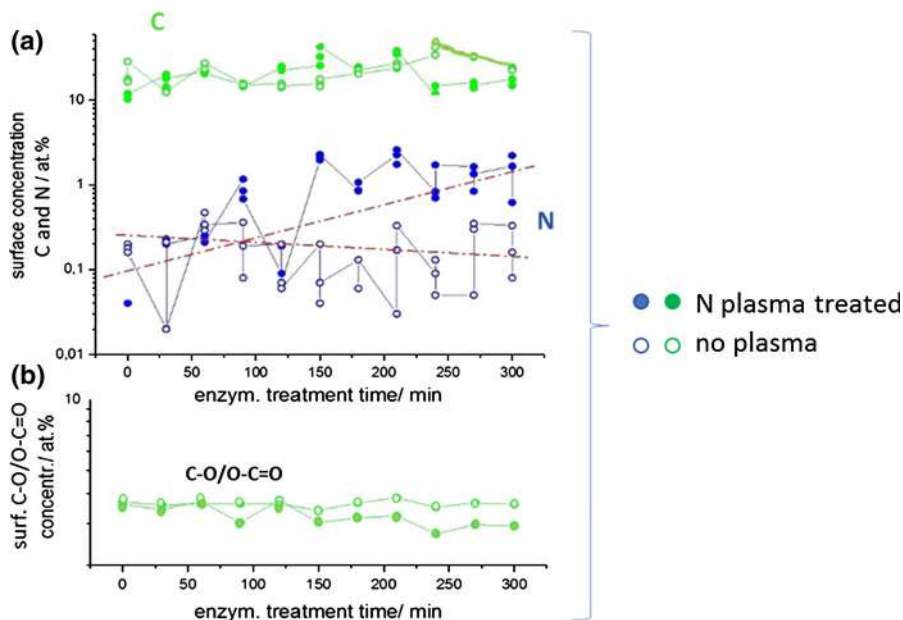

amorphous part with increased hydrolysis, the number of C–O groups decreases while C=O groups and other C and N containing groups are formed.

The results shown in Fig. 7 reveal that with the increase in enzymatic treatment of the raw material pulp there is a reduction of total FSE in the corresponding MNFC films in both polar and dispersive energy (green and blue unfilled symbols, respectively). A reversal of the decline in FSE as a function of enzymatic treatment can be observed resulting from nitrogen plasma treatment, showing compensating increases in both polar and dispersive measured

components (green and blue filled symbols, respectively). Thus, an increase in wettability for water and n-hexadecane is reflected by a decrease in CA as the plasma treatment acts on the more crystalline samples (Johansson and Campbell 2004). However, as the roughness is also seen to increase as a function of plasma treatment for the lower crystalline samples (less exposure to enzymatic breakdown), one would expect from the Wenzel model that the wettability would increase. That we see a recorded increase in n-hexadecane CA, and thus decrease in dispersive FSE, we can conclude that the action of the plasma

Fig. 7 Surface free energy (SFE) of MNFC films as a function of the treatment time (Table 1)
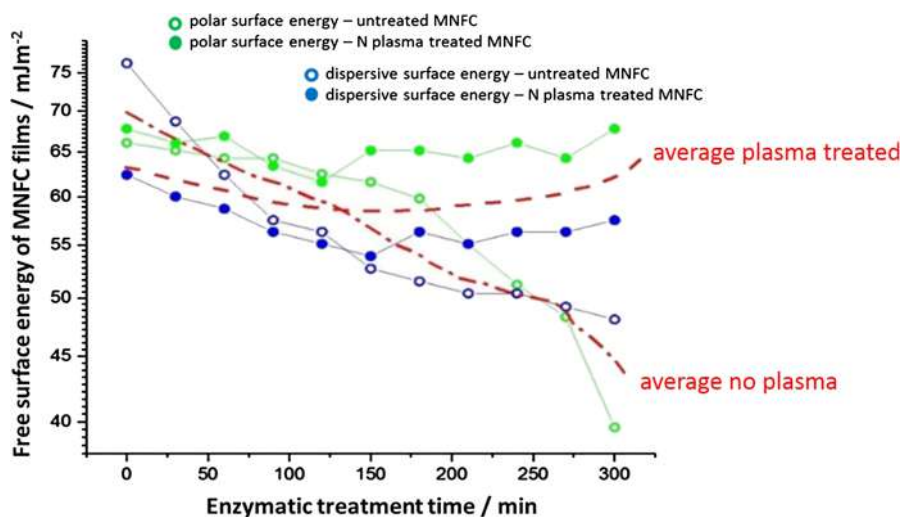
discharge on the amorphous part is initially to reduce the dispersive energy component, and so likely act, at least partially, to breakdown first the amorphous content resulting in debonding and hence roughening (Hansson et al. 2011). This effective etching of amorphous parts of fibrils is then replaced by the action of nitrogen attachment, such that the higher average FSE values regained in the more crystalline samples after plasma treatment are significantly higher than the theoretical FSE 59.4 mJ m^{-2} of cellulose, and this is achieved via the major contribution of the plasma-induced increase in polar component.

The increased contribution of the polar component in the FSE donated by the cationic N adsorption under plasma exposure is, therefore, expected to enhance the compatibility with the application of highly polar inks, especially if their components are anionic (Vanneste et al. 2017; Ma et al. 2010; Hoth et al. 2008). The images in Fig. 8 confirm this expectation, where the improved wetting of the surface by water as a function of plasma exposure time is paralleled by the greater pick-up (trapping) of ink colorant (Hoeng et al. 2016).

Summary and conclusions

Micro nanofibrillated cellulose films formed from aqueous suspension can be made stronger by pretreatment of the raw fibre using enzymatic hydrolysis. However, the wettability by ionic liquids, including functional inkjet printing inks, such as are suitably used for printed electronics, solar cells etc., decreases as a result, limiting the use of such films in practice. Nitrogen plasma treatment, however, enables wettability by such formulations to be improved. The mechanism by which this occurs has been studied in this work presented in this paper and the following conclusions can be drawn:

- Total free surface energy increases with nitrogen plasma treatment of highly enzymatically hydrolysed fibrillar films (contact angle decreases), with a major increase in the polar component.
  - Nitrogen is also included into the surface.
- Upon exposure to nitrogen plasma, dispersive surface energy initially decreases on those films made of pulp that was not treated or undergone short enzymatic treatment time, whereas the polar surface energy component remains relatively unchanged on such films.
  - This effect is related to the interaction of the nitrogen plasma with the amorphous cellulose component in the non-hydrolysed fibrils.
  - The dispersive energy component can once again be increased by exposure to nitrogen plasma in the case of the more crystalline fibrillar material derived from increased hydrolysis via enzymatic pretreatment.
- The surface area per unit mass was increased by the plasma treatment, apparently due to increased roughness on a nanometre scale.
- Highly ionic liquids, water and solvents typically used to disperse surfactant-containing organic-

Fig. 8 IP ink printed on MNFC/300/film showing the dependence on wettability of the surface after nitrogen plasma treatment (see also Fig. 4); lower water droplet CA on the film corresponds with a significant increase in print colour density: **a** untreated film, **b** plasma treated for 30 s and **c** plasma treated for 60 s
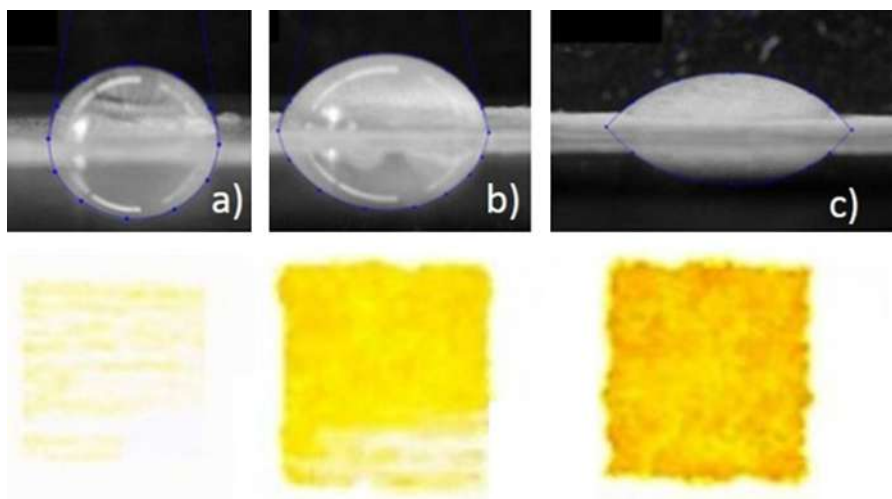
based inks, wet MNFC film better as hydrolysing pretreatment of fibres is increased and subsequent nitrogen plasma is applied.

Perspectives and future work arising from these findings include the need to study the origins of the surface roughening effect. Is this a random generation of surface disruption or is there a material transfer mechanism at play, involving perhaps vaporisation and redisposition? The impact on the amorphous component by plasma treatment could offer a means to induce a phase change at the material surface. Similarly, other gas plasma treatments should be investigated in the longer term to understand whether the role of atomic substitution versus the application of energy discharge has the greater treatment potential.

References

Afsahi G, Dimic-Misic K, Gane P, Budtova T, Maloney T, Vuorinen T (2018) The investigation of rheological and strength properties of NFC hydrogels and aerogels from hardwood pulp by short catalytic bleaching (H cat). Cellulose 25:1637–1655

Catia R, Castro G, Rana S, Fangueiro R (2015) Characterization of physical, mechanical and chemical properties of quiscal fibres: the influence of atmospheric DBD plasma treatment. Plasma Chem Plasma Process 35:863–878

Cernakova L, Stahel P, Kovacik C, Johansson K, Cernak M (2006) Low-cost high-speed plasma treatment of paper surfaces. In: 9th TAPPI advanced coating fundamentals symposium, Turku, Finland, pp 8–10

Chu PK, Chen JY, Wang LP, Huang N (2002) Plasma-surface modification of biomaterials. Mater Sci Eng R Rep 36:143–206

Dimic-Misic K, Puisto A, Gane P, Nieminen K, Alava M, Paltakari J, Maloney T (2013) The role of MFC/NFC swelling in the rheological behaviour and dewatering of high consistency furnishes. Cellulose 20:2847–2861

Dimic-Misic K, Karakoc A, Özkan M, Ghufran HS, Maloney T, Paltakari J (2015) Flow characteristics of ink-jet inks used for functional printing. J Appl Eng Sci 13:207–212

Dimic-Misic K, Maloney T, Gane P (2018) Effect of fibril length, aspect ratio and surface charge on ultralow shear-induced structuring in micro and nanofibrillated cellulose aqueous suspensions. Cellulose 25:117–136

Galagan Y, Rubingh JEJ, Andriessen R, Fan CC, Blom PW, Veenstra SC, Kroon JM (2011) ITO-free flexible organic solar cells with printed current collecting grids. Solar Energy Mater Solar Cells 95:1339–1343

Girifalco LA, Good RJ (1957) A theory for the estimation of surface and interfacial energies. I. Derivation and application to interfacial tension. J Phys Chem 61:904–909

Hansson PM, Skedung L, Claesson PM, Swerin A, Schoelkopf J, Gane PAC, Rutland MW, Thormann E (2011) Robust hydrophobic surfaces displaying different surface roughness scales while maintaining the same wettability. Langmuir 27:8153–8159

Hashmi SG, Özkan M, Halme J, Dimic-Misic K, ZakeeruddinSM PJ, Grätzel M, Lund PD (2015) High performance dye-sensitized solar cells with inkjet printed ionic liquid electrolyte. Nano Energy 17:206–215

Hoeng F, Denneulin A, Bras J (2016) Use of nanocellulose in printed electronics: a review. Nanoscale 8:13131–13154

Hoth CN, Schilinsky P, Choulis SA, Brabec CJ (2008) Printing highly efficient organic solar cells. Nano Lett 8(2008):2806–2813

Hubbe MA, Ferrer A, Tyagi P, Yin Y, Salas C, Pal L, Rojas OJ (2017a) Nanocellulose in thin films, coatings, and plies for packaging applications: a review. BioResources 12:2143–2233

Hubbe MA, Tayeb P, Joyce M, Tyagi P, Kehoe M, Dimic-Misic K, Pal L (2017b) Rheology of nanocellulose-rich aqueous suspensions: a review. BioResources 12:9556–9661

Jens V, Ennaert T, Vanhulsel A, Sels B (2017) Unconventional pretreatment of lignocellulose with low-temperature plasma. Chemsuschem 10:14–31

Johansson LS, Campbell JM (2004) Reproducible XPS on biopolymers: cellulose studies. Surf Interface Anal 36:1018–1022

Jun W, Fengcai Z, Bingqiang C (2008) The solubility of natural cellulose after DBD plasma treatment. Plasma Sci Technol 10:743

Jutila E, Koivunen R, Kiiski I, Bollström R, Sikanen T, Gane PAC (2018) Microfluidic lateral flow cytochrome P450 assay on a novel printed functionalized calcium carbonate-based platform for rapid screening of human xenobiotic metabolism. Adv Funct Mater 28(31):1802793–1802803

Kostić M, Radić N, Obradović BM, Dimitrijević S, Kuraica MM, Škundrić P (2009) Silver-loaded cotton/polyester fabric modified by dielectric barrier discharge treatment. Plasma Process Polym 6(1):58–67

Kramer F, Klemm D, Schumann D, Heßler N, Wesarg F, Fried W, Stadermann D (2006) Nanocellulose polymer composites as innovative pool for (bio) material development.

In: Macromolecular symposia, WILEY-VCH Verlag, vol 244, pp 136–148

Kumar P, Chand S (2012) Recent progress and future aspects of organic solar cells. Prog Photovolt Res Appl 20:377–415

Ma H, Yip HL, Huang F, Jen AKY (2010) Interface engineering for organic electronics. Adv Funct Mater 20:1371–1388

Maloney TC (2015) Network swelling of TEMPO-oxidized nanocellulose. Holzforschung 69:207–213

Mihailović D, Šaponjić Z, Radoičić M, Lazović S, Baily CJ, Jovančić P, Nedeljković J, Radetić M (2011) Functional-ization of cotton fabrics with corona/air RF plasma and colloidal TiO2 nanoparticles. Cellulose 18:811–825

Mohtaschemi M, Dimic-Misic K, Puisto A, Korhonen M, Maloney T, Paltakari J, Alava MJ (2014) Rheological characterization of fibrillated cellulose suspensions via bucket vane viscometer. Cellulose 21:1305–1312

Möller M, Leyland N, Copeland G, Cassidy M (2010) Self-powered electrochromic display as an example for inte-grated modules in printed electronics applications. Eur Phys J Appl Phys 5:33205

Özkan M, Dimic-Misic K, Karakoc A, Hashm SG, Lund P, Maloney T, Paltakari J (2016) Rheological characterization of liquid electrolytes for drop-on-demand inkjet printing. Organ Electron 38:307–315

Pääkkönen T, Dimic-Misic K, Orelma H, Pönni R, Vuorinen T, Maloney T (2016) Effect of xylan in hardwood pulp on the reaction rate of TEMPO-mediated oxidation and the rhe-ology of the final nanofibrillated cellulose gel. Cellulose 23(1):277–293

Pertile RA, Andrade FK, Alves JC, Gama M (2010) Surface modification of bacterial cellulose by nitrogen-containing plasma for improved interaction with cells. Carbohydr Polym 82:692–698

Prysiazhnyi V, Kramar A, Dojcinovic B, Zekic A, Obradovic BM, Kuraica MM, Kostic M (2013) Silver incorporation on viscose and cotton fibers after air, nitrogen and oxygen DBD plasma pretreatment. Cellulose 20:315–325

Rantanen J, Dimic-Misic K, Kuusisto J, Maloney TC (2015) The effect of micro and nanofibrillated cellulose water uptake on high filler content composite paper properties and fur-nish dewatering. Cellulose 22:4003–4015

Schultz J, Tsutsumi K, Donnet JB (1977) Surface properties of high-energy solids: II. Determination of the nondispersive component of the surface free energy of mica and its energy of adhesion to polar liquids. J Colloid Interface Sci 59:277–282

Singh M, Haverinen HM, Dhagat P, Jabbour GE (2010) Inkjet printing-process and its applications. Adv Mater 22:673–685

van de Vyver S, Geboers J, Jacobs PA, Sels BF (2011) Recent advances in the catalytic conversion of cellulose. Chem-CatChem 3:82–94

Vanneste J, Ennaert T, Vanhulsel A, Sels B (2017) Unconven-tional pretreatment of lignocellulose with low-temperature plasma. ChemSusChem 10(1):14–31

Willberg-Keyriläainen P, Ropponen J, Lahtinen M, Pere J (2019) Improved reactivity and derivitization of cellulose after pre-hydrolysis with commercial enzymes. BioRe-sources 14(1):561–574

Yinhua Z, Fuentes-Hernandez C, Khan TM, Liu JC, Hsu J, Shim JW, Dindar A, Youngblood JP, Moon RJ, Kippelen B (2013) Recyclable organic solar cells on cellulose nanocrystal substrates. Sci Rep 3:1536

Zhu H, Narakathu BB, Fang Z, Aijazi AT, Joyce M, Atashbar M, Hu L (2014) A gravure printed antenna on shape-stable transparent nanopaper. Nanoscale 6(15):9110–9115

SISSA

PAPER

Growth signals determine the topology of evolving networks

View the article online for updates and enhancements.

IOP ebooks™

Bringing together innovative digital publishing with leading authors from the global scientific community.

Start exploring the collection–download the first chapter of every title for free.

Growth signals determine the topology of evolving networks

Ana Vranić and Marija Mitrović Dankulov*

Institute of Physics Belgrade, University of Belgrade, Pregrevia 118, 11080 Belgrade, Serbia

E-mail: anav@ipb.ac.rs and mitrovic@ipb.ac.rs

Abstract. Network science provides an indispensable theoretical framework for studying the structure and function of real complex systems. Different network models are often used for finding the rules that govern their evolution, whereby the correct choice of model details is crucial for obtaining relevant insights. Here, we study how the structure of networks generated with the aging nodes model depends on the properties of the growth signal. We use different fluctuating signals and compare structural dissimilarities of the networks with those obtained with a constant growth signal. We show that networks with power-law degree distributions, which are obtained with time-varying growth signals, are correlated and clustered, while networks obtained with a constant growth signal are not. Indeed, the properties of the growth signal significantly determine the topology of the obtained networks and thus ought to be considered prominently in models of complex systems.

Keywords: random graphs, networks, network dynamics, stochastic processes

S Supplementary material for this article is available online

Contents

*Author to whom any correspondence should be addressed.

1. Introduction

Emergent collective behavior is an indispensable property of complex systems [1]. It occurs as a consequence of interactions between a large number of units that compose a complex system, and it cannot be easily predicted from the knowledge about the behavior of these units. The previous research offers definite proof that the interaction network structure is inextricably associated with the dynamics and function of the complex system [2–9]. The structure of complex networks is essential for understanding the evolution and function of various complex systems [10–13].

The structure and dynamics of real complex systems are studied using complex network theory [1, 10, 11]. It was shown that real networks have similar topological properties regardless of their origins [14]. They have broad degree distribution, degree–degree correlations, and power-law scaling of clustering coefficient [11, 14]. Understanding how these properties emerge in complex networks leads to the factors that drive their evolution and shape their structure [2].

The complex network models substantially contribute to our understanding of the connection between the network topology and system dynamics and uncover underlying mechanisms that lead to the emergence of distinctive properties in real complex networks [15–17]. For instance, the famous Barabási–Albert model [15] finds the emergence of broad degree distribution to be a consequence of preferential attachment and network growth. Degree–degree anti-correlations of the internet can be explained, at least to a certain extent, by this constraint [18, 19]. Detailed analysis of the emergence of clustered networks shows that clustering is either the result of finite memory of the nodes [20] or occurs due to triadic closure [21].

Network growth, in combination with linking rules, shapes the network topology [22]. While various rules have been proposed to explain the topology of real networks [10], most models assume a constant rate of network growth, i.e., the addition of a fixed number of nodes at each time step [15, 20, 21]. However, empirical analysis of numerous technological and social systems shows that their growth is time-dependent [23–26]. The time-dependent growth of the number of nodes and links in the networks has been considered as a parameter in uncovering network growth mechanisms [27]. The accelerated growth of nodes in complex networks is the cause of the high heterogeneity in the distribution of web pages among websites [23] and the emergence of highly cited authors in citation networks [26]. The accelerated growth of the number of new links added in each time step changes the shape and scaling exponent of degree distribution

in the Barabási–Albert model [28] and model with preferential attachment with aging nodes [29].

The growth of real systems is not always accelerated. The number of new nodes joining the system varies in time, has trends, and exhibits circadian cycles typical for human behavior [24, 25, 30]. These signals are multifractal and have long-range correlations [31]. Some preliminary evidence shows that the time-varying growth influences the structure and dynamics of the social system and, consequently, the structure of interaction networks in social systems [25, 30, 32–34]. Still, which properties of the real growth signal have the most considerable influence, how different properties influence the topology of the generated networks, and to what extent is an open question.

In this work, we explore the influence of real and computer-generated time-varying growth signals on complex networks' structural properties. We adapt the aging nodes model [35] to enable time-varying growth. We compare the networks' structure using the growing signals from empirical data and randomized signals with ones grown with the constant signal using D-measure [36]. We demonstrate that the growth signal determines the structure of generated networks. The networks grown with time-varying signals have significantly different topology compared to networks generated through constant growth. The most significant difference between topological properties is observed for the values of model parameters for which we obtain networks with broad degree distribution, a common characteristic of real networks [10]. Our results show that real signals, with trends, cycles, and long-range correlations, alter networks' structure more than signals with short-range correlations.

This paper is divided as follows. In section 2, we provide a detailed description of growth signals. In section 3, we briefly describe the original model with aging nodes and structural properties of networks obtained for different values of model parameters [35]. We also describe the changes in the model that we introduce to enable time-varying growth. We describe our results in section 4 and show that the values of D-measure indicate large structural differences between networks grown with fluctuating and ones grown with constant signals. This difference is particularly evident for networks with power-law degree distribution and real growth signals. The networks generated with real signals are correlated and have hierarchical clustering, properties of real networks that do not emerge if we use constant growth. We discuss our results and give a conclusion in section 5.

2. Growth signals

The *growth signal* is the number of new nodes added in each time step. Real complex networks evolve at a different pace, and the dynamics of link creation define the time unit of network evolution. For instance, the co-authorship network grows through establishing a link between two scientists when they publish a paper [37]. In contrast, the links in an online social network are created at a steady pace, often interrupted by sudden bursts [38]. A paper's publication is thus a unit of time for the evolution of co-authorship networks, while the most appropriate time unit for social networks is 1 min or 1 h. While systems may evolve at a different pace, their evolution is often driven by the related mechanisms reflected by the similarity of their structure [10].
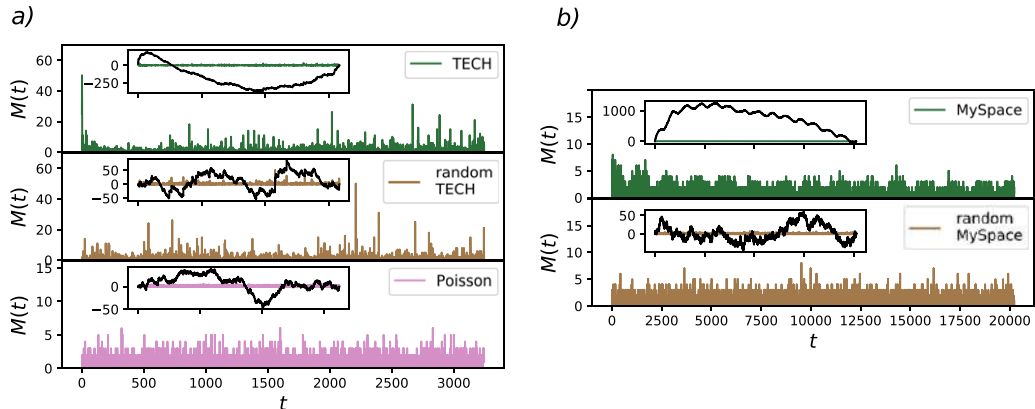
Figure 1. Growth signals for TECH (a) and MySpace (b) social groups, their randomized counterparts, and random signal drawn from Poissonian distribution with mean 1. The cumulative sums of signals' deviations from average mean value are shown in insets.

In this work, we use two different growth signals from real systems figure 1: (a) the data set from TECH community from Meetup social website [39] and (b) two months dataset of MySpace social network [40]. TECH is an event-based community where members organize offline events through the Meetup site [39]. The time unit for TECH is event since links are created only during offline group meetings. The growth signal is the number of people that attend the group's meetings for the first time. MySpace signal shows the number of new members occurring for the first time in the dataset [40] with a time resolution of 1 min. The number of newly added nodes for the TECH signal is $N = 3217$, and the length of the signal is $T_s = 3162$ steps. We have shortened the MySpace signal to $T_s = 20\,221$ time steps to obtain the network with $N = 10\,000$ nodes. The signals in the inset of figures 1(a) and (b) show the cumulative sum of deviations of signals from their average mean value, which is 1.017 for TECH and random TECH signal, 0.47 for MySpace and random MySpace, and 1 for Poissonian signal.

Real growth signals have long-range correlations, trends and cycles [25, 30, 40]. We also generate networks using randomized signals and one computer-generated white-noise signal to explore the influence of signals' features on evolving networks' structure. We randomize real signals using a reshuffling procedure. The reshuffling procedure consists of E steps. We randomly select two signal values at two distinct time steps and exchange their position in each step. The number of reshuffling steps is proportional to the length of the signal T_s, and in our case, it equals $100T_s$. Using this procedure, we keep the signal length and mean value, the number of added nodes, and the probability density function of fluctuations intact, but destroy cycles, trends, and long-range correlations. Besides, we generate a white-noise signal from a Poissonian probability distribution with a mean equal to 1. The length of the signal is $T = 3246$, and the number of added nodes in the final network is the same as for the TECH signal.

We characterize the long-range correlations of the growth signals calculating Hurst exponent [41, 42]. Hurst exponent describes the scaling behavior of time series $M(xt) = x^H M(t)$. It takes values between 0.5 and 1 for long-range correlated signals

and $H = 0.5$ for short-range correlated signals. The most commonly used method for estimating Hurst exponent of real, often non-stationary, temporal signals is detrended fluctuation analysis (DFA) [41]. The DFA removes trends and cycles of real signals and estimates Hurst exponent based on residual fluctuations. The DFA quantifies the scaling behavior of the second-moment fluctuations. However, signals can have deviations in fractal structure with large and small fluctuations that are characterized by different values of Hurst exponents [31].

We use multifractal detrended fluctuation analysis (MFDFA) [31, 43] to estimate multifractal Hurst exponent $H(q)$. For a given time series $\{x_i\}$ with length N, we first define global profile in the form of cumulative sum equation (1), where $\langle x \rangle$ represents an average of the time series:

$$Y(j) = \sum_{i=0}^{j} (x_i - \langle x \rangle), \quad j = 1, \ldots, N. \tag{1}$$

Subtracting the mean of the time series is supposed to eliminate global trends. Insets of figure 1 show global profiles of TECH, MySpace, their randomized signals and Poissonian distribution. The profile of the signal Y is divided into $N_s = \text{int}(N/s)$ non overlapping segments of length s. If N is not divisible with s the last segment will be shorter. This is handled by doing the same division from the opposite side of time series which gives us $2N_s$ segments. From each segment ν, local trend $p_{\nu,s}^m$—polynomial of order m—should be eliminated, and the variance $F^2(\nu, s)$ of detrended signal is calculated as in equation (2):

$$F^2(\nu, s) = \frac{1}{s} \sum_{j=1}^{s} \left[Y(j) - p_{\nu,s}^m(j) \right]^2. \tag{2}$$

Then the qth order fluctuating function is:

$$
\begin{aligned}
F_q(s) &= \left\{ \frac{1}{2N_s} \sum_{\nu}^{2N_s} \left[F^2(\nu, s) \right]^{\frac{q}{2}} \right\}^{\frac{1}{q}}, \qquad q \neq 0 \\
F_0(s) &= \exp \left\{ \frac{1}{4N_s} \sum_{\nu}^{2N_s} \ln \left[F^2(\nu, s) \right] \right\}, \quad q = 0.
\end{aligned}
\tag{3}
$$

The fluctuating function scales as power-law $F_q(s) \sim s^{H(q)}$ and the analysis of log–log plots $F_q(s)$ gives us an estimate of multifractal Hurst exponent $H(q)$. Multifractal signal has different scaling properties over scales while monofractal is independent of the scale, i.e., $H(q)$ is constant.

Figures 1(a) and 2 show that the TECH signal has long trends and a broad probability density function of fluctuations. The trends are erased from the randomized TECH signal, but the broad distribution of the signal and average value remain intact. MFDFA analysis shows that real signals have long-range correlations with Hurst exponent approximately 0.6 for $q = 2$, figure 2. The TECH signal is multifractal, resulting from both broad probability distribution for the values of time series and different long-range correlations of the intervals with small and large fluctuations. Reshuffling of the
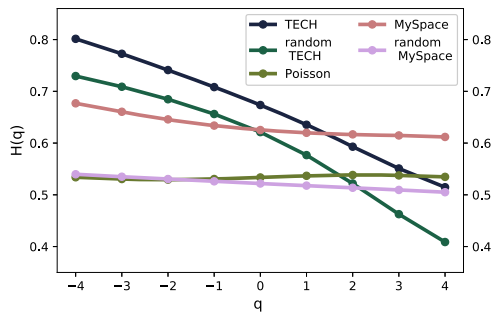
Figure 2. Dependence of Hurst exponent on parameter q for all five signals shown in figure 1 obtained with MFDFA.

time series does not destroy the broad distribution of values, which is the cause for the persistent multifractality of the TECH randomized signal is figure 2.

MySpace signal has a long trend with additional cycles that are a consequence of human circadian rhythm, figure 1(b). Circadian rhythm is an internal process that regulates the sleep-wake cycle and activity, and its period for humans is 24 h [44]. Circadian rhythm leads to periodic changes in online activity during the day and the emergence of a well-defined daily rhythm of activity that we see in figure 1(b). MySpace signal is multifractal for $q < 0$, and has constant value of $H(q)$ for $q > 0$, figure 2. In MFDFA, with negative values of q, we emphasize segments with smaller fluctuations, while for positive q, the emphasis is more on segments with larger fluctuations [43]. Segments with smaller fluctuations have more persistent long-range correlations in both real signals, see figure 2. Randomized MySpace signal and Poissonian signal are monofractal and have short-range with $H = 0.5$ correlations typical for white noise.

Detailed MDFA analysis of real, shuffled, and computer-generated signals are shown in figure S1 and table S1 of the supplementary material (https://stacks.iop.org/JSTAT/2021/013405/mmedia). In figure S1 we show in details how the $F_q(s)$ depends on s for different values of parameter q. The curve $F_q(s)$ exhibits different slopes for different values of q for multifractal signals, i.e., TECH, random TECH, and MySpace. $F_q(s)$ curves for monofractal signals are parallel. We provide the estimated values of $H(q)$ with estimated errors for q in a range from -4 to 4 for all five signals in table S1 of the supplementary material.

3. Model of aging nodes with time-varying growth

To study the influence of temporal fluctuations of growth signal on network topology, we need a model with linking rules where linking probability between network nodes depends on time. We use a network model with aging nodes [35]. In this model, the probability of linking the newly added node and the old one is proportional to their age difference and an old node's degree. In the original version of the model, one node is added to the network and linked to one old node in each time step. The old node is

chosen according to probability

$$\Pi_i(t) \sim k_i(t)^\beta \tau_i^\alpha \tag{4}$$

where $k_i(t)$ is a degree of a node i at time t, and τ_i is age difference between node i and newly added node. As was shown in [35], the values of model parameters β and α determine the topological properties of the resulting networks grown with the constant signal. According to this work, the networks generated using constant growth signals are uncorrelated trees for all values of model parameters. The phase diagram in $\alpha-\beta$ plain, obtained for $\beta > 0$ and $\alpha < 0$, shows that the degree distribution $P(k) \sim k^{-\gamma}$ with $\gamma = 3$ is obtained only along the line $\beta(\alpha^*)$, see [35] and figure S2 in the supplementary material. For $\alpha > \alpha^*$ networks have gel-like small world behavior, while for $\alpha < \alpha^*$ but close to line $\beta(\alpha^*)$ networks have stretched exponential shape of degree distribution [35].

Here we slightly change the original aging model [35] to enable the addition of more than one node and more than one link per newly added node in each time step. In each time step, we add $M \geqslant 1$ new nodes to the network and link them to $L \geqslant 1$ old nodes according to probability Π_i given in equation (4). Again, the networks with broad degree distribution are only generated for the combination of the model parameters along the critical line $\beta(\alpha^*)$. This line's position in the $\alpha-\beta$ plane changes with link density, while the addition of more than one node in each time step does not influence its position. Our analysis shows that the critical line's position is independent of the growth signal's properties, see figure S2 in the supplementary material showing phase diagram. For instance, for $L = 1$ networks and $\alpha = -1.25$ and $\beta = 1.5$ we obtain networks with power-law degree, while for $L = 2$ and $\beta = 1.5$ we need to increase the value of parameter α to -1.0 in order to obtain networks with broad degree distribution. Networks obtained for the values of model parameters $\beta(\alpha^*)$, $L \geqslant 2$, and constant growth have power-law degree distribution, are uncorrelated and have a finite non-zero value of clustering coefficient which does not depend on node degree, figure 4(b). If we fix the value of parameter β and lower down the value of parameter α to -1.5, the resulting networks are uncorrelated with a small value of clustering coefficient, see figure 4(a). For $\alpha < \alpha^*$ we obtain networks with stretched exponential degree distribution, without degree–degree correlations and small value of clustering exponent that does not depend on node degree (see figure S2 in the supplementary material). For $\alpha \ll \alpha^*$ the resulting networks are regular graphs. If we keep the value of α to 1.0 but increase the value β to 2.0 we enter the region of small world gels, see figure 4(c). The networks created for the values of $\alpha > \alpha^*$ are correlated networks with power-law dependence of the clustering coefficient on the degree (see figure S2 in the supplementary material). However, these networks do not have a power-law degree distribution.

The master equation approach is useful for studying the model with aging nodes when $M(t) = 1$ [45]. However, this approach is not sufficient for time-varying growth signals. In this work, we use numerical simulations to explore the case when $M(t)$ is a correlated time-varying function and study how these properties influence the structure of generated networks for different values of parameter $-\infty < \alpha \leqslant 0$ and $\beta \geqslant 1$ and constant L.

4. Structural differences between networks generated with different growth signals

We generate networks for different values of L, and different growth signal profiles $M(t)$. To examine how these properties influence the network structure, we compare the network structure obtained with different growth signals with networks of the same size grown with constant signal $M = 1$. The $M = 1$ is the closest constant value to average values of the signals, which are 1.017 for TECH, 0.47 for MySpace, and 1 for Poissonian signals. We explore the parameter space of the model by generating networks for pairs of values (α, β) in the range $-3 \leqslant \alpha \leqslant -0.5$ and $1 \leqslant \beta \leqslant 3$ with steps 0.5. For each pair of (α, β) we generated networks of different link density by varying parameter $L \in 1, 2, 3$, and for each combination of (α, β, L), we generate a sample of 100 networks and compare the structure of the networks grown with $M = 1$ with the ones grown with $M(t)$ shown in figure 1.

We quantify topological differences between two networks using D-measure defined in [36]

$$D(G, G') = \omega \left| \sqrt{\frac{J(P_1, \ldots, P_N)}{\log(d+1)}} - \sqrt{\frac{J(P'_1, \ldots, P'_N)}{\log(d'+1)}} \right| + (1 - \omega) \sqrt{\frac{J(\mu_G, \mu_{G'})}{log\,2}}. \quad (5)$$

D-measure captures the topological differences between two networks, G and G', on a local and global level. The first term in equation (5) evaluates dissimilarity between two networks on a local level. For each node in the network G one can define the distance distribution $P_i = \{p_i(j)\}$, where $p_i(j)$ is a fraction of nodes in network G that are connected to node i at distance j. The set of N node-distance distributions $\{P_1, \ldots, P_N\}$ contains a detailed information about network's topology. The heterogeneity of a graph G in terms of connectivity distances is measured through node network dispersion (NND). In [36] authors estimate NND as Jensen–Shannon divergence between N distance distributions $J(P_1, \ldots, P_N)$ normalized by $\log(d+1)$, where d is diameter of network G, and show that NND captures relevant features of heterogeneous networks. The difference between NNDs for graph G and G' captures the dissimilarity between the graph's connectivity distance profile.

However, certain graphs, such as k-regular graphs, have NND $= 0$ and can not be compared using NND. For these reasons, authors also introduce average node distance distribution of a graph $\mu(G) = \{\mu(1), \ldots, \mu(d)\}$, where $\mu(k)$ is the fraction of all pair of nodes in the network G that are at a distance k. The Jensen–Shannon divergence between $\mu(G)$ and $\mu(G')$ measures the difference between nodes' average connectivity in a graph G and G'. This term captures the differences between nodes on a global scale.

The original definition of D-measure also includes the third term, which quantifies dissimilarity in node α-centrality. The term can be omitted without precision loss [36]. The parameter ω in equation (5) determines the weight of each term. The extensive analysis shows that the choice $\omega = 0.5$ is the most appropriate for quantifying structural differences between two networks [36].

The D-measure takes the value between 0 and 1. The lower the value of D-measure is the more similar two networks are, with $D = 0$ for isomorphic graphs. The D-measure
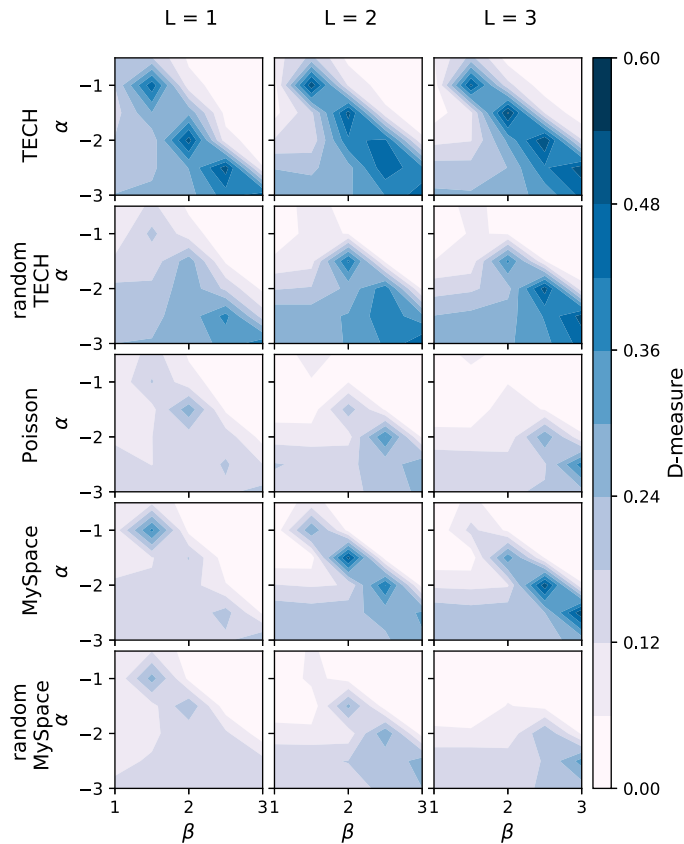
Figure 3. The comparison of networks grown with growth signals shown in figure 1 versus ones grown with constant signal $M = 1$, for value of parameter $\alpha \in [-3, -1]$ and $\beta \in [1, 3]$. $M(t)$ is the number of new nodes, and L is the number of links added to the network in each time step. The compared networks are of the same size.

outperforms previously used network dissimilarity measures such as Hamming distance and graph editing distance and clearly distinguishes between networks generated with the same model but with different values of model parameters [36].

For each pair of networks, one grown with constant and one with the fluctuating signal, we calculate the D-measure. The structural difference between networks grown with constant and fluctuating growth signal for fixed L and values of parameters α and β is obtained by averaging the D-measure calculated between all possible pairs of networks, see figure 3. We observe the non-zero value of D-measure for all time-varying signals. The D-measure has the largest value in the region around the line $\beta(\alpha^*)$. The values of D-measure in this region are similar to ones observed when comparing Erdös–Rényi graphs grown with linking probability below and above critical value [36]. For values $\beta < \beta(\alpha^*)$, the structural differences between networks grown with constant signal and $M(t)$ still exist, but they become smaller as we are moving away from the critical line. Networks obtained with constant signal and fluctuating signals have statistically similar structural properties in the region of small-world network gels, i.e., $\alpha > \alpha^*$.

We focus on the region around the critical line and observe the significant structural discrepancies between networks created for constant versus time-dependent growth signals for all signals regardless of their features. However, the value of D-measure depends on the signal's properties, figure 3. Networks grown with multifractal signals, TECH, random TECH, and MySpace signals, are the most different from those created by a constant signal. The D-measure has the maximum value for the original TECH signal, with $D_{\max} = 0.552$, the signal with the most pronounced multifractal properties among all signals shown in figure 2. Networks generated with randomized MySpace signal and Poisson signal are the least, but still notably dissimilar from those created with $M = 1$.

Randomized MySpace signal and Poissonian signal are monofractal signals with Hurst exponent $H = 0.5$. To investigate the influence of monofractal correlated signals on the network structure, we generate six signals with a different value of $H \in \{0.5, 0.6, 0.7, 0.8, 0.9, 1.0\}$, see figure S3 in the supplementary material. We use each of these signals to generate networks following the same procedure as for signals shown in figure 1. The results shown in figure S4 of the supplementary material confirm that short-range correlated signals create networks with different structures from ones grown with the constant signal. The increase of the Hurst exponent leads to increases in the D-measure. However, D-measure's maximal value is smaller than one observed for multifractal signals shown in figure 3.

The value of D-measure rises with a decline of α^*. This observation can be explained by examining linking rules and how model parameters determine linking dynamics between nodes. The ability of a node to acquire a link declines with its age and grows with its degree. A node's potential to become a hub, node with a degree significantly larger than average network degree, depends on the number of nodes added to the network in the T time steps after its birth. The length of the interval T decreases with parameter α. For constant signal, the number of nodes added during this time interval is constant and equal to MT. For fluctuating growth signals, the number of added nodes during the time T varies with time. In signals that have a broad distribution of fluctuations, like TECH signals, the peaks of the number of newly added nodes lead to the emergence of one or several hubs and super hubs. The emergence of super hubs, nodes connected to more than 30% of the nodes in the network, significantly alters the network's topology. For instance, super hubs' existence lowers the value of average path length and network diameter [10]. The emergence of hubs occurs for values of parameter α relative close to -1.0 for signals with long-range correlations. As we decrease the parameter α, the fluctuations present in the time-varying signals become more important, and we observe the emergence of hubs even for the white-noise signals. The trends present in real growth signals further promote the emergence of hubs. The impact of fluctuations and their temporal features on the structure of complex networks increases with link density.

The large number of structural properties observed in real networks are often consequences of particular degree distributions, degree correlations, and clustering coefficient [47]. Figure 4 shows the degree distribution $P(k)$, dependence of average neighboring degree on node degree $\langle k \rangle_{nn}(k)$, and dependence of clustering coefficient on node degree $c(k)$ for networks with average number of links per node $L = 2$. The significant structural differences between networks grown with real time-varying and constant signals
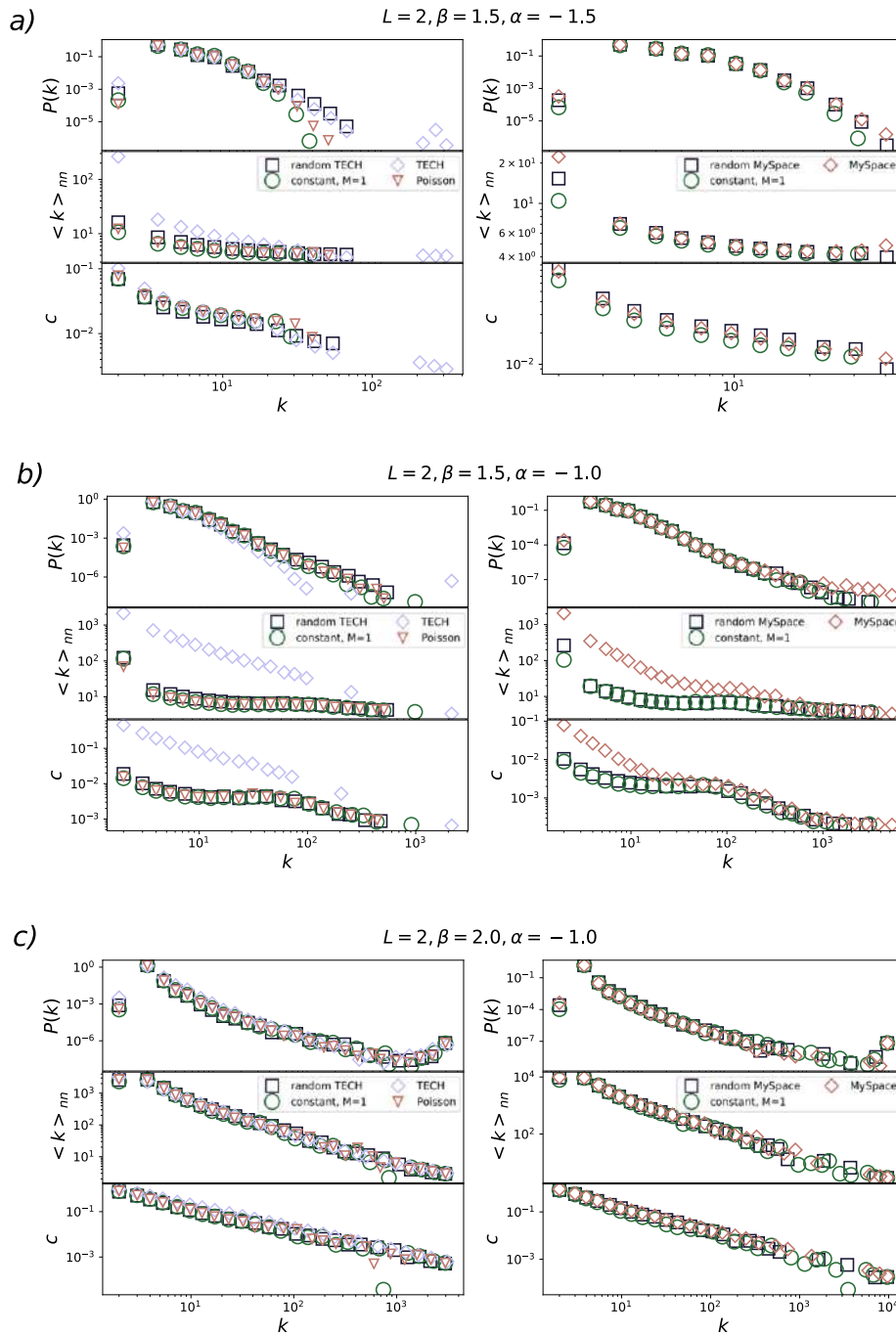
Figure 4. Degree distribution, the dependence of average first neighbor degree on node degree, dependence of node clustering on node degree for networks grown with different time-varying and constant signals. Model parameters have the values $\alpha = -1.5$, $\beta = 1.5$ (a), $\alpha = -1.0$, $\beta = 1.5$ (b), $\alpha = -1.0$, $\beta = 2.0$ (c), and $L = 2$ for all networks.

are observed for the values of model parameters $\alpha = -1.0$ and $\beta = 1.5$, figures 3 and 4(b). The degree distribution of networks generated for real signals shows the occurrence of super hubs in these networks. In contrast, degree distributions of networks generated with white-noise like signals do not differ from one created with constant signal, figure 4(b). Networks obtained for the real signals are disassortative and have a hierarchical structure, i.e., their clustering coefficient decreases with the degree. On the other hand, networks generated with constant and randomized signals are uncorrelated, and their clustering weakly depends on the degree.

We observe a much smaller, but still noticeable, difference between the topological properties of networks evolved with constant and time-varying signal for $\alpha < \alpha^*$, figure 4(a). The difference is particularly observable for degree distribution and dependence of average neighboring degree on node degree of networks grown with real TECH signal. The fluctuations of time-varying growth signals do not influence the topological properties of small-world gel networks, figure 4(c). For $\alpha > \alpha^*$, the super hubs emerge even with the constant growth. Since this is the mechanism through which the fluctuations alter the structure of evolving networks for $\alpha \lesssim \alpha^*$, the features of the growth signals cease to be relevant.

5. Discussion and conclusions

We demonstrate that the resulting networks' structure depends on the time-varying signal features that drive their growth. The previous research [25, 30] indicated the possible influence of temporal fluctuations on network properties. Our results show that growth signals' temporal properties generate networks with power-law degree distribution, non-trivial degree–degree correlations, and clustering coefficient even though the local linking rules, combined with constant growth, produce uncorrelated networks for the same values of model parameters [35].

We observe the most substantial dissimilarity in network structure along the critical line, the values of model parameters for which we generate broad degree distribution networks. Figure 3 shows that dissimilarity between networks grown with time-varying signals and ones grown with constant signals always exists along this line regardless of the features of the growth signal. However, the magnitude of this dissimilarity strongly depends on these features. We observe the largest structural difference between networks grown with multifractal TECH signal and networks that evolve by adding one node in each time step. The identified value of D-measure is similar to one calculated in the comparison between sub-critical and super-critical Erdös–Rényi graphs [36] indicating the considerable structural difference between these networks. Our findings are further confirmed in figure 4(b). The networks generated with signals with trends and long-range temporal correlations differ the most from those grown with the constant signal. Our results show that even white-noise type signals can generate networks significantly different from ones created with constant signal for low values of α^*.

Randomized and computer-generated signals do not have trends or cycles. Nevertheless, networks grown with these signals have a significantly different structure from ones grown with constant M. Our results demonstrate that growth signals' temporal

fluctuations are the leading cause for the structural differences between networks evolved with the constant and time-varying signal. We observe the smallest, but significant, difference between networks generated with constant M and monofractal signal with short-range correlations. As we increase the Hurst exponent, the value of the D-measure increases. The most considerable differences are observed for multifractal signals TECH, random TECH, and MySpace.

The value of D-measure declines as we move away from the critical line, figure 3. The primary mechanism through which the fluctuations influence the structure of evolved networks is the emergence of hubs and super hubs. For values of $\alpha \ll \alpha^*$, the nodes attache to their immediate predecessors creating regular networks without hubs. For $\alpha \lesssim \alpha^*$ graphs have stretched exponential degree distribution with low potential for the emergence of hubs. Still, multifractal signal TECH enables the emergence of hub even for the values of parameters for which we observe networks with stretched-exponential degree distribution in the case of constant growth figure 4(a). By definition, small-world gels generated for $\alpha > \alpha^*$ have super-hubs [35] regardless of the growth signal. Therefore the effects that fluctuations produce in the growth of networks do not come to the fore for values of model parameters in this region of $\alpha-\beta$ plane.

In this work, we focus on the role of the node growth signal in evolving networks' structure. However, real networks do not evolve only due to the addition of new nodes, but also through addition of new links [27–29, 38]. Furthermore, the deactivation of nodes [48] and the links [48] influence the evolving networks' structure. Each of these processes alone can result in a different network despite having the same linking rules. The next step would be to examine how different combinations of these processes influence the evolving networks' structure. For instance, in [28], authors have examined the influence of the time-dependent number of added links $L(t)$ on the Barabási–Albert networks' structure. They show that as long as the average value of time-dependent signal $\langle L(t) \rangle$ is independent of time, the generated networks have a similar structure as Barabási–Albert networks, and that the degree distribution depends strongly on the behavior of $\langle L(t) \rangle$. It would be interesting to examine how correlated $L(t)$ signals influence networks' structure with aging nodes, where the age of a node plays a vital role in linking between new and old nodes. Moreover, we expect that the combination of time-varying growth of the number of nodes and the number of links will significantly influence these networks' structure.

Evolving network models are an essential tool for understanding the evolution of social, biological, and technological networks and mechanisms that drive it [10]. The most common assumption is that these networks evolve by adding a fixed number of nodes in each time step [10]. So far, the focus on developing growing network models was on linking rules and how different rules lead to networks of various structural properties [10]. Growth signals of real systems are not constant [25, 30]. They are multifractal, characterised with long-range correlations [25], trends and cycles [40]. Research on temporal networks has shown that temporal properties of edge activation in networks and their properties can affect the dynamics of the complex system [12]. Our results imply that modeling of social and technological networks should also include non-constant growth. Its combination with local linking rules can significantly alter the structure of generated networks.

Acknowledgments

References

[1] Ladyman J, Lambert J and Wiesner K 2013 What is a complex system? *Euro J. Phil. Sci.* **3** 33

[2] Barrat A, Barthelemy M and Vespignani A 2008 *Dynamical Processes on Complex Networks* (Cambridge: Cambridge University Press)

[3] Pascual M *et al* 2006 *Ecological Networks: Linking Structure to Dynamics in Food Webs* (Oxford: Oxford University Press)

[4] Castellano C, Fortunato S and Loreto V 2009 Statistical physics of social dynamics *Rev. Mod. Phys.* **81** 591

[5] Gosak G, Markovič R, Dolenšek J, Rupnik M S, Marhl M, Stožer A and Perc M 2018 Network science of biological systems at different scales: a review *Phys. Life Rev.* **24** 118–35

[6] Arenas A, Díaz-Guilera A, Kurths J, Moreno Y and Zhou C 2008 Synchronization in complex networks *Phys. Rep.* **469** 93

[7] Boccaletti S, Almendral J A, Guan S, Leyva I, Liu Z, Sendiña-Nadal I, Wang Z and Zou Y 2016 Explosive transitions in complex networks' structure and dynamics: percolation and synchronization *Phys. Rep.* **660** 1

[8] Chen H, Zhang H and Shen C 2018 Double phase transition of the Ising model in core-periphery networks *J. Stat. Mech.* 063402

[9] Kuga K and Tanimoto J 2018 Impact of imperfect vaccination and defense against contagion on vaccination behavior in complex networks *J. Stat. Mech.* 113402

[10] Boccaletti S, Latora V, Moreno Y, Chavez M and Hwang D 2006 Complex networks: structure and dynamics *Phys. Rep.* **424** 175

[11] Newman M E J 2010 *Networks: An Introduction* (Oxford: Oxford University Press)

[12] Holme P and Saramäki J 2012 Temporal networks *Phys. Rep.* **519** 97

[13] Boccaletti S, Bianconi G, Criado R, Del Genio C I, Gómez-Gardeñes J, Romance M, Sendiña-Nadal I, Wang Z and Zanin M 2014 The structure and dynamics of multilayer networks *Phys. Rep.* **544** 1

[14] Barabási A-L 2009 Scale-free networks: a decade and beyond *Science* **325** 412

[15] Barabási A-L and Albert R 1999 Emergence of scaling in random networks *Science* **286** 509

[16] Tadić B 2001 Dynamics of directed graphs: the world-wide web *Physica* A **293** 273

[17] Mitrović M and Tadić B 2009 Spectral and dynamical properties in classes of sparse networks with mesoscopic inhomogeneities *Phys. Rev.* E **80** 026123

[18] Maslov S, Sneppen K and Zaliznyak A 2004 Detection of topological patterns in complex networks: correlation profile of the internet *Physica* A **333** 529

[19] Park J and Newman M E J 2003 Origin of degree correlations in the internet and other networks *Phys. Rev.* E **68** 026112

[20] Klemm K and Eguiluz V M 2002 Highly clustered scale-free networks *Phys. Rev.* E **65** 036123

[21] Serrano M A and Boguná M 2005 Tuning clustering in random networks with arbitrary degree distributions *Phys. Rev.* E **72** 036133

[22] Vázquez A 2003 Growing network with local rules: preferential attachment, clustering hierarchy, and degree correlations *Phys. Rev.* E **67** 056104

[23] Huberman B A and Adamic L A 1999 Growth dynamics of the world-wide web *Nature* **401** 131

[24] Mitrović M and Tadić B 2010 Bloggers behavior and emergent communities in blog space *Eur. Phys. J.* B **73** 293

[25] Dankulov M M, Melnik R and Tadić B 2015 The dynamics of meaningful social interactions and the emergence of collective knowledge *Sci. Rep.* **5** 1

[26] Liu J, Li J, Chen Y, Chen X, Zhou Z, Yang Z and Zhang C-J 2019 Modeling complex networks with accelerating growth and aging effect *Phys. Lett.* A **383** 1396

[27] Pham T, Sheridan P and Shimodaira H 2016 Joint estimation of preferential attachment and node fitness in growing complex networks *Sci. Rep.* **6** 32558

[28] Sen P 2004 Accelerated growth in outgoing links in evolving networks: deterministic versus stochastic picture *Phys. Rev.* E **69** 046107

[29] Dorogovtsev S N and Mendes J F F 2001 Effect of the accelerating growth of communications networks on their structure *Phys. Rev.* E **63** 025101

[30] Mitrović M and Tadić B 2012 *Emergence and Structure of Cybercommunities* (*Springer Optimization and Its Applications*) vol 57 (Berlin: Springer) p 209

[31] Kantelhardt J W, Zschiegner S A, Koscielny-Bunde E, Havlin S, Bunde A and Stanley H E 2002 Multifractal detrended fluctuation analysis of nonstationary time series *Physica* A **316** 87

[32] Mitrović M, Paltoglou G and Tadić B 2011 Quantitative analysis of bloggers' collective behavior powered by emotions *J. Stat. Mech.* P02005

[33] Tadić B, Dankulov M M and Melnik R 2017 Mechanisms of self-organized criticality in social processes of knowledge creation *Phys. Rev.* E **96** 032307

[34] Tadić B and Šuvakov M 2013 Can human-like bots control collective mood: agent-based simulations of online chats *J. Stat. Mech.* P10014

[35] Hajra K B and Sen P 2004 Phase transitions in an aging network *Phys. Rev.* E **70** 056103

[36] Schieber T A, Carpi L, Díaz-Guilera A, Pardalos P M, Masoller C and Ravetti M 2017 Quantification of network structural dissimilarities *Nat. Commun.* **8** 1

[37] Sarigöl E, Pfitzner R, Scholtes I, Garas A and Schweitzer F 2014 Predicting scientific success based on coauthorship networks *EPJ Data Sci.* **3** 9

[38] Myers S A and Leskovec J 2014 The bursty dynamics of the twitter information network *Proc. 23rd Int. Conf. on World Wide Web* 913

[39] Smiljanić J and Dankulov M M 2017 Associative nature of event participation dynamics: a network theory approach *PloS One* **12** e0171565

[40] Šuvakov M, Mitrovi M, Gligorijevi V and Tadi B 2013 How the online social networks are used: dialogues-based structure of MySpace *J. R. Soc. Interface* **10** 20120819

[41] Peng C-K, Buldyrev S V, Havlin S, Simons M, Stanley H E and Goldberger A L 1994 Mosaic organization of DNA nucleotides *Phys. Rev.* E **49** 1685

[42] Kantelhardt J W, Koscielny-Bunde E, Rego H H A, Havlin S and Bunde A 2001 Detecting long-range correlations with detrended fluctuation analysis *Physica* A **295** 441

[43] Fürst EAFI Ihlen E A 2012 Introduction to multifractal detrended fluctuation analysis in Matlab *Front. Physiol.* **3** 141

[44] Wever R A 2013 *The Circadian System of Man: Results of Experiments under Temporal Isolation* (Berlin: Springer)

[45] Dorogovtsev S N and Mendes J F F 2001 Scaling properties of scale-free evolving networks: continuous approach *Phys. Rev.* E **63** 056125

[46] Orsini C *et al* 2015 Quantifying randomness in real networks *Nat. Commun.* **6** 8627

[47] Tian L, Zhu C-P, Shi D-N, Gu Z-M and Zhou T 2006 Universal scaling behavior of clustering coefficient induced by deactivation mechanism *Phys. Rev.* E **74** 046103

[48] Gagen M J and Mattick J S 2005 Accelerating, hyperaccelerating, and decelerating networks *Phys. Rev.* E **72** 016123

Article

Evolution of Cohesion between USA Financial Sector Companies before, during, and Post-Economic Crisis: Complex Networks Approach

Vojin Stević [1] , **Marija Rašajski** [1] and **Marija Mitrović Dankulov** [2,*]

[1] University of Belgrade-School of Electrical Engineering, Bulevar Kralja Aleksandra 73, 11120 Belgrade, Serbia;
vojin.stevic@gmail.com (V.S.); rasajski@etf.bg.ac.rs (M.R.)
[2] Institute of Physics Belgrade, University of Belgrade, Pregrevica 118, 11080 Belgrade, Serbia
[*] Correspondence: mitrovic@ipb.ac.rs

Abstract: Various mathematical frameworks play an essential role in understanding the economic systems and the emergence of crises in them. Understanding the relation between the structure of connections between the system's constituents and the emergence of a crisis is of great importance. In this paper, we propose a novel method for the inference of economic systems' structures based on complex networks theory utilizing the time series of prices. Our network is obtained from the correlation matrix between the time series of companies' prices by imposing a threshold on the values of the correlation coefficients. The optimal value of the threshold is determined by comparing the spectral properties of the threshold network and the correlation matrix. We analyze the community structure of the obtained networks and the relation between communities' inter and intra-connectivity as indicators of systemic risk. Our results show how an economic system's behavior is related to its structure and how the crisis is reflected in changes in the structure. We show how regulation and deregulation affect the structure of the system. We demonstrate that our method can identify high systemic risks and measure the impact of the actions taken to increase the system's stability.

Keywords: complex networks; time series; economic systems; evolution of community structure

1. Introduction

Economic crises negatively impact people's lives. They influence every aspect of individual and social development. Therefore, it is essential to prevent a crisis or alleviate its impact by promptly taking appropriate action. Thus, it is necessary to understand the economic system's functioning and behavior before, after, and during the crisis. Different approaches have been applied towards that end, including economic [1,2] and quantitative approaches [3–9].

The economic system is a complex system consisting of many interacting units whose collective behavior cannot be inferred from individual units' behavior. The behavior of the complex system is determined by its structure [10,11]. To understand the behavior and function of a complex system, one needs to describe its structure and understand how this structure evolves. Complex networks theory provides tools for the inference of the structure of a wide range of systems, including biological [12], social [11], technological [13], and economic systems [14]. The construction of economic networks is mostly achieved by mapping the flow of funds between companies [15] or transforming time series into a correlation matrix [14]. These two networks are complementary, although they overlap to a certain extent. The former network requires more time-consuming data collection, while the advantage of obtaining a network from time series is in its simplicity and the availability of data. The appropriate method for efficiently extracting information from time series is essential since it provides insights into the system's structure at a relatively low data collection cost.

Existing methods for obtaining networks use as input data either the time series of logarithmic returns [16] or methods based on detrended logarithmic returns [14]. Both are derived from the time series of prices. The direct use of the prices is often avoided because they are non-stationary and contain trends.

Current works obtain a network from a correlation matrix by applying filtering methods such as the minimum spanning tree (MST) [17], planar maximally filtered graph (PMFG) [14], and threshold method [18]. Complex networks, including economic networks, are characterized by the rich mesoscopic structure, known as communities [19]. MST and PMFG techniques are not appropriate for analyzing communities in a network and their interconnections as they focus on including all nodes in the network, disregarding stronger intra-community connectivity [20]. Existing methods that use the threshold method do not differentiate between relevant and less relevant edges, filtering out essential information about the system.

The economic crisis is a common research topic [3–9]. These works contribute to a better understanding of crisis by examining the system's functioning using different quantitative methodologies. This diverse approach is especially beneficial for a better understanding of the crisis. Input data are the time series of market indices coming from different countries [21] or time series of companies' stock prices constituting, for example, the S&P 500 index [6]. The former data do not include many constituents and are used in works where the subject understands global interaction. The latter focuses on the USA market, which can have up to 500 constituents, the largest USA companies, and all sectors.

In this work, we propose a new approach for obtaining the network from price time series, which provides insight into the system's structure. Our motivation is to obtain an optimal network containing sufficient information and as few edges as possible, allowing efficient analysis. Moreover, we want to gain insight into the change in the system's structure due to the economic crisis. We use a time series of prices and apply detrending to those series. We demonstrate that the system's structure can be inferred from these data, thus broadening the dataset options for the empirical study of complex systems. We applied this approach for obtaining a network from a correlation matrix that differentiates between edges based on their relevance to network topology. Using this approach, we studied the evolution of the USA financial sector's network structure. The financial sector is the heart of the economy since companies in this sector enable a flow of funds through the economy. The 2008 economic crisis was catalyzed by subprime mortgage-backed securities in the USA and spread to mutual funds, pensions, and other parts of the financial sector, with national and global impacts. The USA financial sector occupied a central place in the emergence and development of the 2008 crisis. For these reasons, we focus on studying the evolution of company relations in this sector. Our input data present all companies from the domestic financial sector, the source of the 2008 crisis. We identify the relationship between the economic system's structure and behavior before, during, and after the crisis. Our analysis shows how an economic crisis affects a system's structure on a mesoscopic level. We examine the relationship between inter- and intra-community connectivity. We show that, using this approach, we can detect crises and different interventions by governments and policy-makers by examining the community structure and their connections. By adding the different perspectives of observing the system's behavior in crisis, we contribute to a better understanding of connectivity and relations within the economic system.

The rest of the paper is organized as follows. In Section 2, we give an overview of previous work on the following topics: time series processing, obtaining networks from correlation matrix, and methods for studying economic crisis. In Section 3.2, we provide a detailed description of our approach. We present our results in Section 3.1, and discuss these results and conclude in Section 5.

2. Related Work

2.1. Time Series Processing

The time series of prices are not suitable for calculating the correlation matrix since they have strong trends and are non-stationary. Different methods were applied in order to overcome these problems. One of the methods is based on the simple transformation of prices into logarithmic returns [6,21–25], derived as $r_t = \ln \frac{P_t}{P_{t-1}}$, where P_t presents the price at time t. These series fluctuate around the mean, which is constant and close to zero.

The other methods apply detrending techniques on the time series of returns [14,26–28]. These detrending techniques differ according to trend calculation. Zhao et al. [14] make a cumulative time series of returns and calculate the trend for each series separately based on the detrending fluctuation analysis technique [29]. Random matrix theory is used to calculate market component, representing the trend in [26,28]. Musmeci et al. [27] calculate the market component based on average returns for all companies considered in the analysis.

Some works used the auto-correlation of time series of returns to derive residuals, which are then used to calculate the correlation matrix [17,30]. Dynamic conditional correlation multivariate generalized autoregressive conditionally heteroscedastic (GARCH) model, DCC-MVGARCH, is used in these works.

2.2. Obtaining Network from Correlation Matrix

Obtaining a network from a correlation matrix suitable for gaining insights into the system's structure is a complex problem. It involves the usage of an appropriate filtering method. The method should ensure that relevant information is present in the network and that redundant edges are removed. Not satisfying any of the two requirements can lead to false conclusions. Existing filtering methods include the minimum-spanning tree (MST) [17,30–32], planar maximally filtered graph (PMFG) [6,14,33,34] and threshold method [18,20–23,28].

The threshold method filters out information based on correlation strength, while MST and PMFG combine correlation strength to include all graph nodes and planarity. From the perspective of inter- and intra-connectivity between communities, inclusion and planarity criteria result in a connected graph at the price of not including all relevant edges. Onnela et al. [20] compared the threshold method with MST and showed that a threshold network with the same number of edges as MST results in a disconnected graph. These results imply that intra-community edges are more robust than edges between the communities. Moreover, in [6], PMFG leads to the conclusion that in times of crisis, communities are less connected than they out of crisis, which is in contrast to results obtained using random matrix theory [3,7].

The threshold method is more suitable for analyzing community structure in the network. However, the problem is finding the optimal threshold value. A lower threshold is desirable to include as much information as possible. On the other hand, a higher threshold is preferable since it provides a sparse network, which is easier for analysis. The optimal threshold is the one that filters out noise from the network structure and leaves the edges that carry relevant information about mutual relations between entities. Onnela et al. [20] proposed clustering coefficient as the criteria for determining the threshold value. However, there is no substantial evidence that the clustering coefficient is more relevant than other network measures.

X. Cao et al. [28] calculated the optimal threshold by comparing clustering coefficients, the average shortest path length, and the size of the giant component between random graph and empirical network for different threshold values. They determine the optimal threshold at which the structural difference between empirical and random networks is at the highest level. While these network properties are one of the most investigated ones, they are not inclusive of other topological properties [35]. The work from C. Orsini et al. [35] indicates that the degree sequence, joint degree matrix, average clustering coefficient, and its dependence on the node degree are sufficient to describe the topology of most of the

networks. In contrast, the giant component's average shortest path length and size depend on these properties.

Xue Guo et al. [18] determine the threshold based on the community's correlation strength. This approach underestimates the inter-connectivity between communities as higher importance is given to intra-community edges. Inter-community edges impact the diffusion process in a network and should be recognized appropriately. Moreover, according to the max-flow min-cut theorem, edges between communities are essential since the information flow is maximal through them.

S. Kumar et al. [21] set different thresholds to show how network characteristics, such as the component number and maximum clique size, change with the threshold. Xia et al. [23] determine the threshold by using the probability distribution of correlation coefficients and setting the threshold at the expected value plus multiples of standard deviations.

The mentioned threshold methods do not provide quantitative insights into how much information is filtered from the network. The complete correlation matrix carries all information about the structure of the network. Once threshold filtering is applied, a certain amount of information is lost. Therefore, it is essential to have quantitative insight into how much information we included in the network. It is vital to see which edges carry the relevant information about the systems' topology and which are redundant. Here, we propose a quantitative measure based on the network's spectral properties to determine the optimal value of the threshold.

2.3. Crisis Examined Using Quantitative Methodologies

Different quantitative methods have been applied to better understand the impact of the crisis on the economic system. V. Filimonov et al. [5] used the Poisson Hawkes model and developed a measure to determine whether price fluctuations are due to an endogenous feedback process as opposed to exogenous news. A. M. Petersen et al. [4] studied cascading dynamics and related the Omori, productivity, and Bath laws with financial shocks. G. Oh et al. [36] used entropy density function in return time series, while K. Yim et al. [37] used the Hurst exponent.

Complex networks theory is also used for the analysis of crisis impact. X. Cao et al. [28] have shown that the crisis impacts the average degree, size of the giant component, and clustering coefficient. S. Kumar et al. [21] presented how the crisis affects the formation of clusters and the structure of minimum spanning trees. A. Nobi [24] showed the impact of the crisis on degree distribution and cluster formation. M. Wilinski [38] showed that MST changes structure from a hierarchical scale-free MST to a superstar-like MST decorated by a scale-free hierarchy of trees. L. Zhao et al. [6] examined how the crisis affects the number of communities and inter-sector edges.

Existing methods that use complex networks to analyze the impact of a crisis primarily consider either mapping country indices [21] or the constituents of leading indices S&P 500 [6]. The former network comprises nodes representing different countries, while the latter network nodes represent companies from different sectors. These companies are, for example, for index S&P 500, the largest 500 companies in the USA. This work demonstrates our approach to studying the evolution of relations between companies in the USA financial sector. We show that laws and policies strongly influence the system's structure. The network's community structure reflects the pre-crisis, crisis, and post-crisis periods.

3. Materials and Methods

3.1. Data

Innovative solutions such as derivatives and securitization in the financial sector that were not followed by developing the system's regulatory framework created a bubble in the housing and credit supply markets. The bubble burst in 2008 due to the subprime mortgage crisis, which led to a worldwide economic crisis. This work studies the long-term

relations between companies in the USA's financial sector and its evolution from 2002 until 2017. This period includes the time before the 2008 crisis, the period during the crisis, and the economic recovery period. The financial sector includes companies whose main economic activity is asset management, real estate investment trusts (REITs), banks, insurance, and municipal funds.

We obtained data from the publicly available Finance Yahoo database https://finance.yahoo.com/ accessed on 27 September 2018 which contains various information data about the company's values and how they have changed with time. The database comprises different data types, for instance, opening, closing, intraday, adjusted closing prices, and trading volume. The information is given for different aggregation intervals: day, week, and month. For this study, we used adjusted daily closing prices. The closing price means that the price is taken at the end of the business day after trading is closed. The price fluctuates between the opening and closing of a business day. Adjusted means that the price is corrected to exclude the effect of dividend pay-out and stock split. The impact of dividend pay-out and splits of stock would provide misleading information. A split or dividend pay-out can significantly change the price, although the company's real value did not change.

For each year $T \in \{2002, \ldots, 2017\}$, we collected the time series $x_i^T(t)$ of the adjusted closing price at the end of each trading day t for each company i. Each time series' length equals the one-year or 252 trading days. The number of companies in each year $N_c(T)$ varies since some companies were founded after 2002, and some of them were closed before 2017. Table 1 shows the number of companies active in year T in the USA financial sector according to the Yahoo Finance database.

Table 1. The number of USA financial sector companies in each year T according to the Yahoo database.

T	$N_c(T)$	T	$N_c(T)$
2002	518	2010	762
2003	558	2011	786
2004	609	2012	804
2005	653	2013	825
2006	695	2014	855
2007	711	2015	884
2008	740	2016	892
2009	748	2017	888

Table 1 shows that the number of companies in the USA financial sector grew by 7.5% per year on average before 2007. The crisis and economic recovery period from 2007 until 2015 had much slower growth, with an average relative increase in the number of companies of approximately 2.7%. There was a certain stagnation of this growth in 2016 and 2017.

3.2. Methodology

This work proposes a method for determining the network of relations between companies based on their stock price time series. We use this method to study the evolution of cohesion of financial sector companies whose stocks are publicly traded on the USA's stock exchange. With this method, we explore the evolution of mutual influences and how this evolution is shaped by different critical events, such as the world economic crisis in 2008.

Our method consists of three steps. In the first step, we perform a detrending time series of stock prices for each considered company. In the second step, we calculate the matrix of Pearson correlation coefficients using this detrended time series. In the final step, we apply the threshold filtering of correlation coefficients to extract the companies'

network of relations. We then analyze and compare the topology of the networks obtained for different years.

3.2.1. Time Series of Prices for Obtaining Network

In our approach, the companies are represented by nodes, and edges represent companies' relationships. As input data, we use the time series of stock prices. We consider the time series of each company's adjusted daily closing prices. The considered time series are non-stationary and have strong trends, as can be seen in Figure 1 (green line), which are often the consequence of different external influences. The non-stationarity of the time series and the trends can lead to false, highly positive, or negative correlations between companies. To avoid this, we remove the trends by detrending the time series using the method proposed in [29]. The detrended time series is the time series of the fluctuations.

Original time series $x_i^T(t)$ consists of 252 values of adjusted daily stock prices of the company i during year T. In [29], the authors considered the differential time series of fluctuations and then performed detrending on the cumulative time series. Our original time series are already cumulative, thus omitting this step in our calculations. We divide the time series on k non-overlapping segments of equal size l, so that $k = \frac{n}{l}$. We determine the linear trend of time series x_i by fitting the equation $y_i^j(t) = a_i^j \times t + b_i^j$ and determining the coefficients a_i^j and b_i^j for each segment j, as can be seen in Figure 1 (red line). The detrended time series equals the original time series minus the trend on each segment, i.e., $\overline{x_i^T}(t) = x_i^T(t) - y_i^j(t)$. The resulting time series is stationary, and its average value is approximately zero. By removing the trend typical for period l, we only consider fluctuations that result from mutual influence between companies.
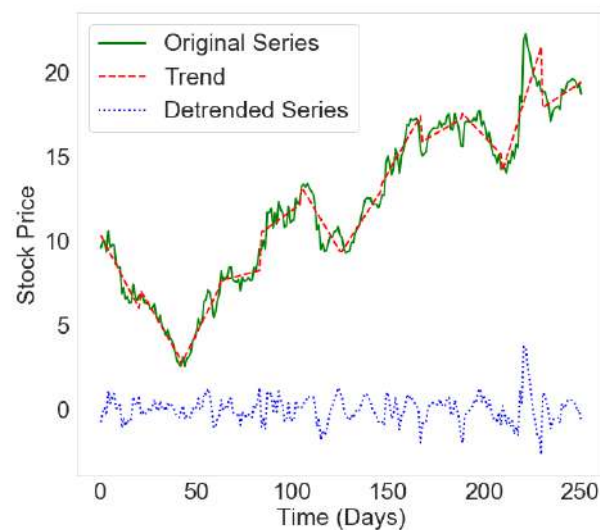
Figure 1. Example of a time series of prices for one company belonging to the USA financial sector. The green line is the original time series, the red line shows the trend, and the blue line is a detrended time series of prices.

We apply detrending to each company's time series. The detrended time series are used for the calculation of the Pearson correlation coefficient matrix for year T, where each element of the matrix is calculated using the following formula:

$$\hat{\rho}_{i,j}^T = \frac{\sum_{t=1}^n (\overline{x_i^T}(t) - \hat{\mu}_{\overline{x_i^T}})(\overline{x_j^T}(t) - \hat{\mu}_{\overline{x_j^T}})}{\sqrt{\sum_{t=1}^n (\overline{x_i^T}(t) - \hat{\mu}_{\overline{x_i^T}})^2 \sum_{t=1}^n (\overline{x_j^T}(t) - \hat{\mu}_{\overline{x_j^T}})^2}}, \tag{1}$$

where $\overline{x_i^T}(t)$ and $\overline{x_j^T}(t)$ are the detrended time series of companies i and j in year T, $\hat{\mu}_{\overline{x_i^T}}$ and $\hat{\mu}_{\overline{x_j^T}}$ are estimated average values over period $n = 252$ for the detrended time series of companies i and j. The matrix with $\hat{\rho}_{i,j}^T$ elements is symmetrical and takes values from -1 to 1.

In order to obtain the network of mutual influences between considered companies for the year T, we only take into account the correlation coefficient with a value above a certain threshold θ, i.e.,

$$w_{i,j}^T = \begin{cases} \hat{\rho}_{i,j}^T & \text{if } \hat{\rho}_{i,j}^T > \theta \\ 0 & \text{if } \hat{\rho}_{i,j}^T \leq \theta \end{cases} . \tag{2}$$

Determining the threshold value θ is not a simple task. In their approach, Živković et al. [39] assumed that the most optimal value of the threshold can be determined from the relation between the threshold value and the size of the largest component in the network obtained for that value. The giant component is the largest set of connected nodes in the network [10]. The dependence of the size of giant component S on the threshold value θ has a characteristic steep decline in the giant component's size for a particular value of the threshold θ_c. The abrupt deterioration implies the detachment of a group of nodes forming separate components. θ_c is the threshold value for which one can observe essential changes in the network structure. The threshold value is determined as the one which is slightly smaller than θ_c.

Figure 2 shows the dependence of the size of the giant component S on the value of threshold θ for the financial sector in the year 2015. There are two steep drops in the value of the giant component's size, one for the values of the threshold starting from 0.54 and ending at 0.56, and one starting from 0.78 and ending at 0.82. This indicates that observed networks elapse through a series of significant structural changes; thus, it is hard to determine the optimal threshold value.
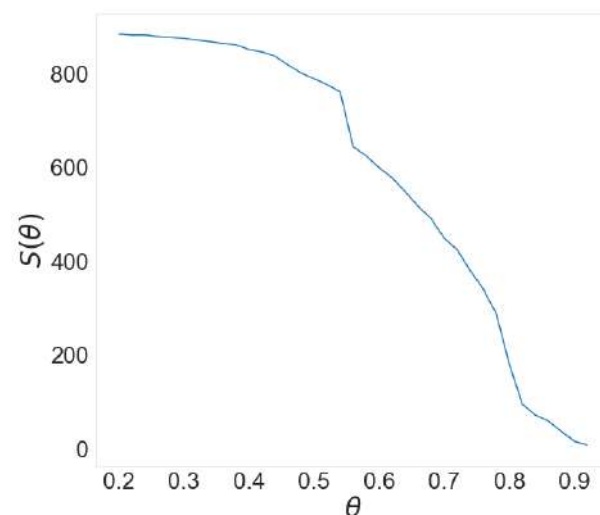
Figure 2. Dependence of size of giant component (S) on value of threshold (θ).

For these reasons, we adopted a different approach. We determine the optimal threshold for filtering the correlation matrix based on the networks' spectral properties. The probability distribution of the eigenvalues of the adjacency matrix fundamentally describes a system and contains the complete information about its topology [40–42]. Different networks, such as Erdos–Reniy and Barabasi–Albert graphs, have different probability distributions of eigenvalues. The difference between the two networks is proportional to their structural differences.

We compare the empirical economic network's spectra with the spectra of different random networks to demonstrate our claims. C. Orsini et al. [35] proposed a method to

create a line of random networks, each topologically more similar to an empirical network. We obtained a network from the correlation matrix by applying the threshold method and using it as the empirical network. We generated three random networks (RNs) based on the empirical network properties. RN1 has the same average degree as the empirical network, while other topological properties are random. The RN2 has the same degree sequence and consequently, the average degree as the original network. The RN3 has the same joint degree matrix, degree sequence, average degree, and the most similar topology to the empirical network.

Figure 3 shows the spectra of the empirical network obtained for 2005 and three random networks. The RN1 has the most different spectral properties than the empirical network, while the RN3 has the most similar spectra. Each random network only contains a fraction of information about the relations between nodes in the empirical network. The difference between spectra decreases as we increase the number of properties similar to the empirical network. Our analysis demonstrates that we can use the comparison between spectra to evaluate the optimal threshold.
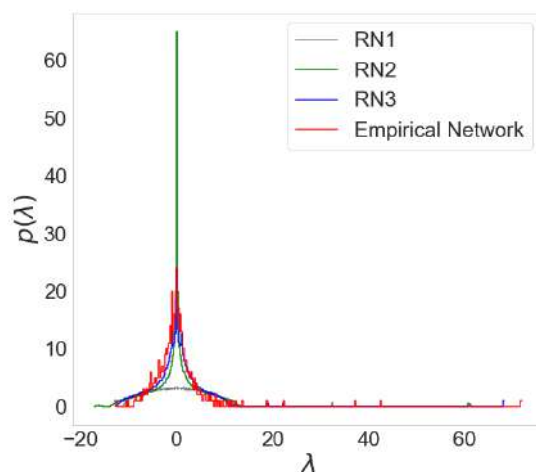
Figure 3. The probability distribution of eigenvalues for empirical network for year 2005 and three random networks.

We used the same approach to compare the full correlation matrix, represented as a weighted and filtered network.

The correlation matrix contains complete information about the system and can be represented as a weighted graph. Once the threshold is applied to the correlation matrix, edges with weights less than the threshold are removed. A filtered network thus only has a fraction of information about companies' relations. By comparing the probability distributions of eigenvalues for original and filtered matrices, we understand how much information is lost due to filtering.

To quantify this difference, we use the Kolmogorov–Smirnov (KS) distance. We calculate the KS distance between the probability distributions of eigenvalues for the original and filtered correlation matrices for different threshold values. The lower value of KS distance implies a better agreement between spectra and higher similarity between networks' topologies. Therefore, we want the KS distance to be as low as possible.

Figure 4 shows the KS dependence on the threshold for 2008, 2009, 2014, and 2015. As we expect, the KS distance increases with the threshold value. At the threshold -0.5, the KS distance is equal to zero as complete information is included in the network, while the KS distance reaches its maximum for a threshold close to 1. The dependence of KS on threshold has a local minimum at the value $\theta_m > 0$ and is similar to the KS distance at the threshold $\theta = 0$. We keep the same information about the network structure by fixing the threshold's value at 0 or θ_m. However, a network at 0 is denser and thus more complicated for the analysis. By setting the threshold value to θ_m for which we observe

the local minimum for the KS distance, and we obtain the optimal network with enough information about the relations between companies, which is not excessively dense.
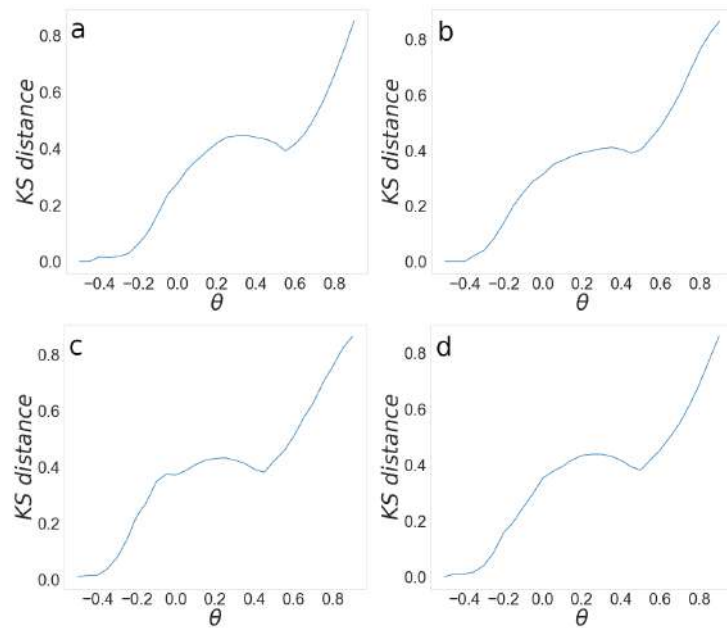
Figure 4. Kolmogorov–Smirnov distance between the probability distributions of the eigenvalues of correlation matrices obtained from original and filtered matrix for the years 2008 (**a**), 2009 (**b**), 2014 (**c**), and 2015 (**d**).

The probability distributions of correlation coefficients differ for each year, as can be seen in Figure 5; thus, it is not surprising that the local minimum is different for each year. We calculate the local minimum for each year separately and obtain the network based on corresponding thresholds. Table 2 shows the local minima θ for different years.
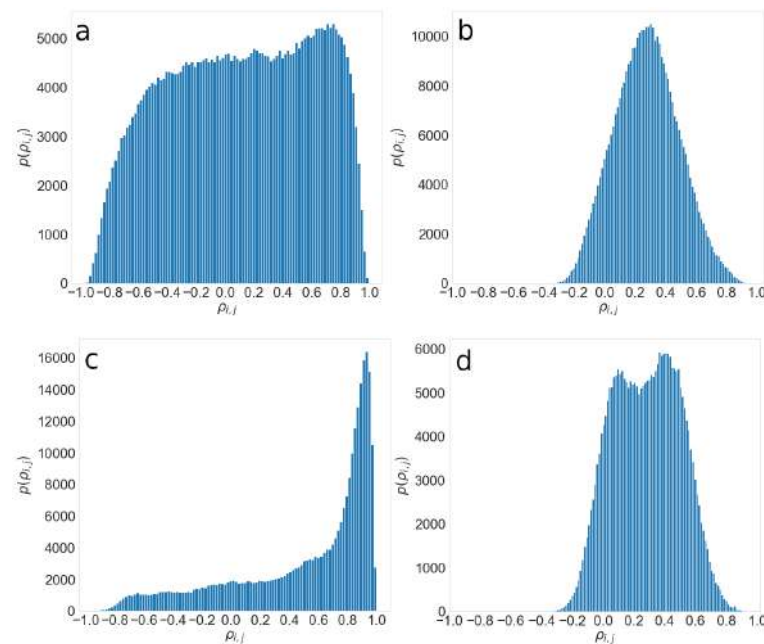
Figure 5. Probability distribution of the correlation coefficients for the years 2015, (**a**,**b**), and 2009, (**c**,**d**). (**a**,**c**) present distributions obtained from original time series, while (**b**,**d**) are the distributions obtained from detrended series.

Table 2. The threshold values obtained for different years.

Year	θ	Year	θ
2002	0.35	2010	0.525
2003	0.325	2011	0.55
2004	0.425	2012	0.475
2005	0.35	2013	0.475
2006	0.35	2014	0.45
2007	0.475	2015	0.5
2008	0.55	2016	0.55
2009	0.475	2017	0.375

3.2.2. Measuring the Intra- and Inter-Community Connectivity

We are interested in the mesoscopic structure of the networks and how it changes with time. A community is a group of nodes more densely connected than the rest of the network [19]. Communities are an indicator of the system's collective behavior, and the network's community structure provides essential information about its dynamics and function [19]. In this work, we apply the Louvain algorithm [43] to find communities in weighted networks. The results of the Louvain algorithm for a single run may differ due to different initial conditions. We run a Louvain algorithm each year 100 times for these reasons. For each community $CM_i^{T,r}$, where T denotes a year and r denotes the run of the Louvain algorithm, we calculate the ratio between edges inside the community and all edges formed by nodes belonging to that community. We calculate the ratio using the following equation

$$P_{in}^{CM_i^{T,r}} = \frac{L_{in}^{CM_i^{T,r}}}{L_{Total}^{CM_i^{T,r}}}, \quad i = 1, 2, ..., R^{T,r} \tag{3}$$

where $L_{in}^{CM_i^{T,r}}$ is the sum of weighted edges inside the community $CM_i^{T,r}$, $L_{Total}^{CM_i^{T,r}}$ is the total sum of weighted edges of nodes in the community $CM_i^{T,r}$, and $R^{T,r}$ is the number of communities for the network obtained for time period T and run r. First, the average $P_{in}^{CM_i^{T,r}}$ over all communities obtained in the single run

$$\langle P_{in}^{T,r} \rangle = \frac{\sum_i P_{in}^{CM_i^{T,r}}}{R^{T,r}}, \tag{4}$$

and then we obtain the average over all runs

$$\langle P_{in}^{T} \rangle = \frac{\sum_r \langle P_{in}^{T,r} \rangle}{100}, \tag{5}$$

and standard deviation

$$\sigma_{P_{in}^{T}} = \sqrt{\frac{\sum_r \left(\langle P_{in}^{T} \rangle - \langle P_{in}^{T,r} \rangle \right)^2}{99}} \tag{6}$$

4. Results

This work focuses on how the network structure changed when the system went through the 2008 economic crisis. We selected the period between 2002 and 2017, which covers the time before, during, and after the crisis. The number of companies varies between 518 in 2002 and 888 in 2017, as can be seen in Table 2.

We detrended each segment separately and calculated the correlation matrix $\{\hat{\rho}_{i,j}\}$ between the companies for each year $T \in \{2002, ..., 2017\}$. We detrended the time series for the interval $l = 21$ trading days, which equals one average trading month. We then mapped the correlation matrix to the adjacency matrix using the threshold method and

obtained an undirected weighted network for the year T. We use the approach described in Section 3.2 to determine the threshold. We performed community structure analysis and calculated the $\overline{P_{in}^T}$ and $\sigma_{P_{in}^T}$ for each year. The analysis of community structure and the evolution of their cohesion shows how the network structure evolves.

4.1. Characteristics of the Correlation Matrix and Obtained Network

Detrending helps extract information about the economic system's internal behavior and relationships between companies. Figure 5 shows a probability distribution of correlation coefficients $p(\hat{\rho}_{i,j})$ for the original time series and detrended time series for years 2009 and 2015. The $p_{ij}(\hat{\rho}_{i,j})$ shown in Figure 5a was calculated for the original time series for the year 2015 and resembles a uniform distribution. Figure 5b shows the probability distribution obtained from the detrended series and is more similar to Gaussian distribution. The center of Gaussian varies between years. The distribution of correlation coefficients changes during the economic crisis period, as can be seen in Figure 5c,d. If we obtain the correlation matrix from the original time series, most companies are highly correlated, with correlation coefficients between 0.9 and 1, as can be seen in Figure 5c. The distribution of correlation coefficients obtained from detrended time series during an economic crisis is a convolution of two Gaussians, Figure 5d.

After detrending the time series and calculating the correlation matrix, we used a method described in Section 3.2 to obtain an undirected weighted network. We ran Louvain on the networks and found the community structure. Figure 6 shows the networks for the years 2004, 2006, 2008, and 2015. Based on examining communities by comparing their constituents' characteristics, we concluded that their edges imply exposure to similar factors. Namely, the nodes belonging to a community, i.e., companies in the same sector, have different owners, operate in different states, and have different clients. Common to these companies is their economic activity, i.e., their functioning is similar. Therefore, we obtain a network where edges reflect exposure to similar factors.
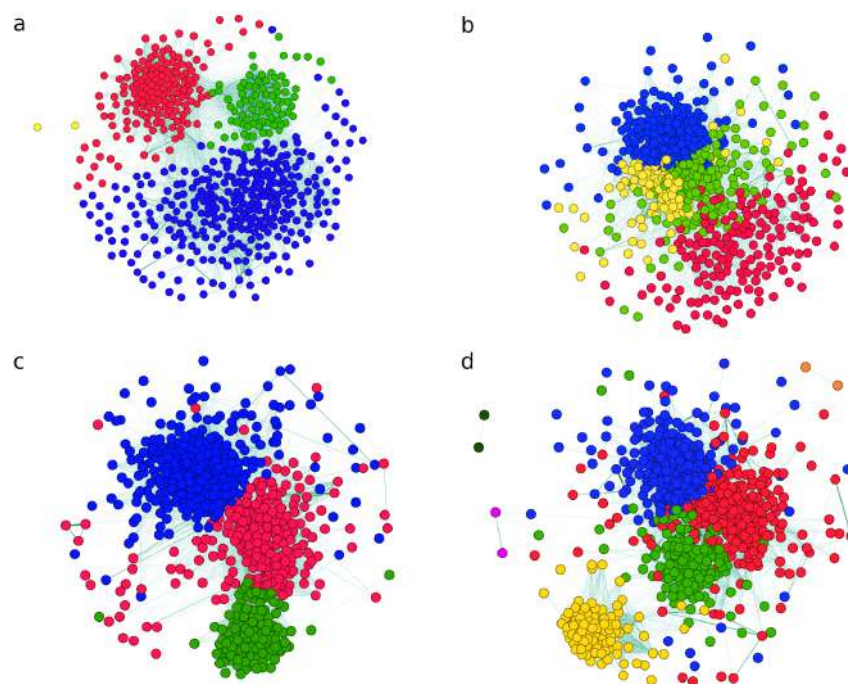
Figure 6. Networks obtained from detrended time series in the years 2004 (**a**), 2006 (**b**), 2008 (**c**), and 2015 (**d**). Networks are obtained by applying the threshold given in Table 1. The number of nodes and edges, respectively, is equal to 554 and 23,340 (**a**), 652 and 43,167 (**b**), 677 and 47,590 (**c**), and 793 and 58,291 (**d**). Nodes of the same color belong to the same community.

Inter-community edges indicate that even companies belonging to different subsectors may operate under similar conditions. For example, a bank and REIT company are exposed to similar external factors if they are linked to the residential project, where a bank lends money to home buyers while REIT invests in project development.

Robust intra-community connectivity indicates that companies in the same community operate under similar conditions and are susceptible to the same factors. The low value of correlation coefficients between companies belonging to different sub-sectors suggests that specific factors typically affect them. Strong connectivity between network nodes is an indicator of its high vulnerability. A system with a distinct community structure and stronger connectivity within the communities than between the communities is more robust than one with similar strengths of connections between and within the communities.

We are interested in the evolution of the ratio between intra- and inter-community connectivity and how this ratio changes when the system is in different states, such as during crisis and out-of-crisis periods.

4.2. Relation between Inter and Intra-Connectivity of Communities and Its Evolution

We analyze the community structure of networks for each year from 2002 to 2017 using the Louvain method. The results of applying the Louvain method, which includes the number and structure of communities, depend on the initial conditions. As a result, different runs of the Louvain algorithm on the same network may result in a different number of communities depending on how network nodes are assigned to these communities. For these reasons, we ran the Louvain algorithm 100 times on each of the 16 networks and calculated the average number of communities and the average connectivity of these communities. Figure 7a shows the evolution of the number of communities between 2002 and 2017. The number of communities fluctuates with time and grows after the peak of the crisis in 2008, with two distinctive local minima in 2010 and 2013. Furthermore, the number of detected communities in 2004, 2008, and 2015 is equal for each of the 100 runs of the Louvain algorithm, suggesting a stable community structure in these networks. We observe the lowest number of communities for 2008, which indicates the lowest differentiation between sectors within the financial industry during the financial crash.
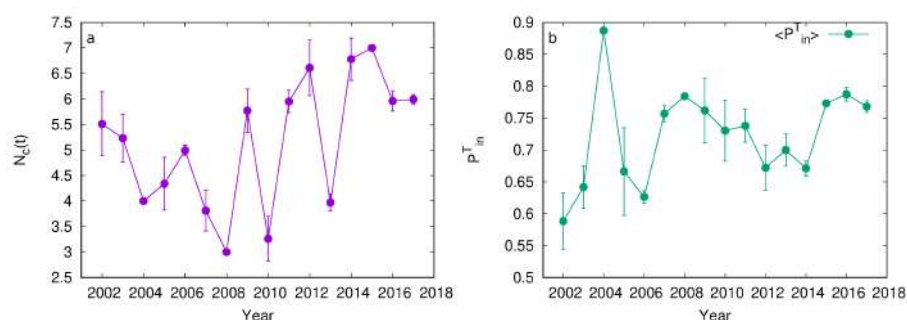
Figure 7. The evolution of the average number of communities (**a**) and average intra-connectivity (**b**) for networks from 2002 to 2017.

We analyze the intra- and inter-community connectivity for the networks obtained for each year from 2002 to 2017. Figure 7b shows $<P_{in}^T>$ for the years from 2002 to 2017. Higher values of $<P_{in}^T>$ imply higher community intra-connectivity, while lower values indicate higher community inter-connectivity. The error bars shown in Figure 7b are standard deviations calculated on the sample of 100 runs. Low standard deviation implies similar intra-community connectivity among communities. The peak of intra-community connectivity is observed in the year 2004. The interconnectivity then drops to its minimal value in 2006, where the connectivity within the communities grows and has local maxima in 2008, which slowly decreases until 2014. In 2015, we observed another smaller rise in connectivity.

The networks of the years 2004, 2008, and 2015 have two essential features. They have a very stable number of communities, independent from the initial conditions of the Louvain method. Furthermore, the intra-community connectivity for these networks has a local maximum in these years with a very low standard deviation.

Figure 7a shows that the intra-community connectivity $\langle P_{in}^T \rangle$ has the local minima in the year 2006, indicating high connectivity between communities. In 2006, the system had the highest potential for diffusion between communities, meaning that one community's disturbance could easily be transmitted to any other community. If this disturbance is a failure, the system is at high risk of efficiently spreading failure and breaking down. Our result matches what happened to the USA financial sector since 2006 was the year before the crisis started in 2008. Other researchers have predicted the beginning of the crisis [1]. High and consistent inter-community connectivity in 2006 indicates that companies in different sectors were susceptible to the influence of the same factor. This factor was real estate lending, which pulled most of the financial industry. Many financial sectors were directly or indirectly involved in real estate lending, leading to the relationship network's almost homogeneous structure. The local minima in 2006 preceded a peak in 2004, where communities were well defined.

A crisis is followed by a period of recession, which is recognized by lower values of economic indicators such as employment, gross domestic product, household net worth, and federal surplus or deficit. Figure 8 shows the relative change of these four indicators for the USA economy between 2002 and 2017. We see that the recession period lasted from 2009 and ended in 2014. Our results indicate that the standard deviation for intra-community connectivity has higher values for the same period, while its values decrease between 2014 and 2017. We see from Figure 7b that standard deviation $\langle P_{in}^T \rangle$ was higher during the economic recovery compared to the post-crisis period.
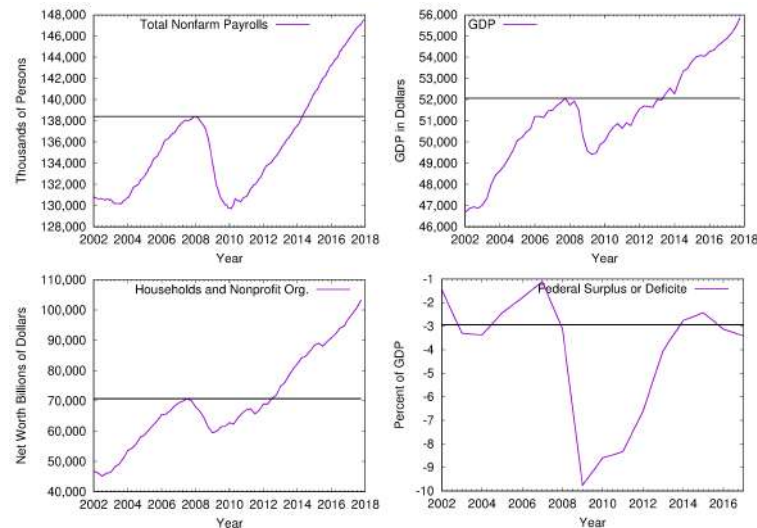
Figure 8. Economic variables which indicate whether the system is in the state of crisis or out of crisis such as (left top) employment, (right top) gross domestic product, (left bottom) household net worth, and (right bottom) federal surplus or deficit.

We observe the increase in $\langle P_{in}^T \rangle$, see Figure 7a in the years 2007 and 2008, after reaching its minimum in 2006. In 2007, companies in the financial system understood that the economy was in bad condition and that interconnection was high. Communities tried to depart from each other, leading to a high $\langle P_{in}^T \rangle$ and low $\sigma_{P_{in}^T}$ in the year 2008. However, the number of communities $N_C(T)$ decreased in 2008 because two communities merged into one, regional banks and REITs. We observed the homogenizing of the system in a different form where the number of different sectors decreased.

Effect of Regulation on Structure and Behavior

The USA financial system has to be controlled to prevent the system break down and decrease systemic risk [1]. Control is realized through appropriate regulations. High restrictive regulation may prevent a crisis. However, it can jeopardize economic growth since it limits companies' profits [1]. Less stringent regulations enable higher yields but increase systemic risk. Therefore, an optimal level of regulation has to be implemented, allowing a thriving economy while decreasing systemic risk. The regulations impose restrictions on companies' behavior, while deregulation provides companies with a higher degree of freedom. They both define the behavior of comprising elements of the system. The effect of regulation and deregulation on the system results from the collective behavior of incorporating elements. One needs tools to measure the impact of regulations on the system to create optimal regulations. Our methodology provides insight into the influence of regulation and deregulation on a system's structure and behavior.

Deregulation took place in 2004 [44] and proposed a system of voluntary regulations where investment banks can hold less capital in reserve. Having less money in reserve means that companies become more dependent on other companies and more vulnerable. Higher connectivity between companies leads to an increase in systemic risk. Deregulation is considered one of the leading causes of crisis [45]. Our results show that $\langle P_{in}^T \rangle$ sharply decreased in 2005, indicating higher inter-dependence between communities and higher systemic risk. $\langle P_{in}^T \rangle$ and the standard deviation $\sigma_{P_{in}^T}$ further decreased in 2006, implying higher homogeneity within the system.

Regulations were implemented between 2011 and 2014 to respond to the crisis. The Dodd–Frank Wall Street Reform and Consumer Protection Act of 2010 was designed to increase financial stability and prevent future crises [46]. As this was the most comprehensive overhaul of the financial system [47], it took time to be implemented. Implementation started in 2011 and reached 50% of planned regulations in 2014 [48]. Our results show a sharp increase in $\langle P_{in}^T \rangle$ in 2014 when the economy recovered. Standard deviation $\sigma_{P_{in}^T}$ is higher during the crisis period compared to the period of the recovering economy, 2014–2017.

5. Discussion and Conclusions

In this work, we used a novel method to infer network structure from time series to study the cohesion between USA companies in the financial sector. Compared to existing methods, we used detrended prices instead of detrended returns. We introduced a technique for obtaining an optimal network from a correlation matrix and used a measure based on community structure that allows us to examine the evolution of cohesion. Our results show that the USA financial system's network structure between 2002 and 2017 underwent several phases: deregulation, crisis, and post-crisis. Each of these periods is characterized by different intra-community connectivity and standard deviation. The strength of connections between communities is directly related to the system's level of risk and stability.

Understanding the connections between the system's components is crucial for preventing crises. Our approach can identify the points of high systemic risk. This knowledge enables timely actions to increase the system's stability. Moreover, measuring the effect of these actions, such as regulation and deregulation, can be performed using our method. This is of great importance as inadequate efforts can further deteriorate financial stability. In 2008, the government's actions to increase financial stability and save the economy in the form of capital injection into the financial system were inadequate, which further pushed the economy into recession [1]. The price of wrong measures for recovering the economy is high in times of crisis because resources are even more limited. Our results show that the system's structure did not change due to these measures.

The economic system has to be regulated to prevent crises while securing the unrestrained behavior of individual companies to allow economic growth and prosperity. The economic system is dynamic and should be constantly monitored by policymakers

to secure an optimal trade-off between of the economic growth and limiting behavior of composing elements. Policymakers must act on time since a delay of adequate actions can have a negative impact. Our method allows policymakers to see whether their actions are adequate and act promptly. As per our analysis, deregulation, which took place in 2004 to enable economic growth, had a strong impact on increasing systemic risk. This signal can be seen in 2005, where $\langle P_i n \rangle$ sharply decreased, while standard deviation implied that connectivity between a certain number of communities increased. In addition, in 2006, all communities were strongly interconnected, which presents a high systemic risk and can be seen in low $\langle P_{in} \rangle$, standard deviation, and the number of communities. This led to the 2008 crisis, when some of the communities merged.

Existing techniques for constructing networks from the correlation matrix, MST and PMFG, put strict constraints on the network structure. MST forbids cycles between nodes and conditions the number of links to $N - 1$, where N is the number of nodes. PMFG only allows short cycles and the maximal number of links $3(N - 2)$. There is no economic reason behind these topological constraints for economic systems. Furthermore, the limit on the number of connections is too strict and may filter out some critical information about the network's connectivity. The lack of this forbids the study of the cohesion of the network and its dynamics.

Our method can be used by researchers interested in studying collective behavior in real systems such as economic, social, biological, and technological systems. The prerequisite is the availability of data in the form of time series. Our method enables discovering hidden relationships between the constituents of the system, leading to a better understanding of the system, predicting its behavior and controlling it.

Author Contributions: V.S., M.R. and M.M.D. designed the research; V.S. collected the data; V.S. and M.M.D. contributed program tools, performed the analysis, and analyzed data; V.S. and M.M.D. produced the figures; V.S., M.R. and M.M.D. wrote the and reviewed the manuscript. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data used in this research are publicly available and can be found at https://figshare.com/articles/dataset/Financial_data/20088311 accessed on 17 June 2022.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Stiglitz, J.E. *Freefall: America, Free Markets, and the Sinking of the World Economy*; WW Norton & Company: New York, NY, USA, 2010.
2. Krugman, P. The profession and the crisis. *East. Econ. J.* **2011**, *37*, 307–312. [CrossRef]
3. Junior, L.S.; Franca, I.D.P. Correlation of financial markets in times of crisis. *Phys. A* **2012**, *391*, 187–208.
4. Petersen, A.M.; Wang, F.; Havlin, S.; Stanley, H.E. Market dynamics immediately before and after financial shocks: Quantifying the Omori, productivity, and Bath laws. *Phys. Rev. E* **2010**, *82*, 036114. [CrossRef] [PubMed]
5. Filimonov, V.; Sornette, D. Quantifying reflexivity in financial markets: Toward a prediction of flash crashes. *Phys. Rev. E* **2012**, *85*, 056108. [CrossRef] [PubMed]
6. Zhao, L.; Li, W.; Cai, X. Structure and dynamics of stock market in times of crisis. *Phys. Lett. A* **2016**, *380*, 654–666. [CrossRef]
7. Fenn, D.J.; Porter, M.A.; Williams, S.; McDonald, M.; Johnson, N.F.; Jones, N.S. Temporal evolution of financial-market correlations. *Phys. Rev. E* **2011**, *84*, 026109. [CrossRef]
8. Yang, L.; Rea, W.; Rea, A. Impending doom: The loss of diversification before a crisis. *J. Financ. Stud.* **2017**, *5*, 29. [CrossRef]

9. Meng, H.; Xie, W.J.; Jiang, Z.Q.; Podobnik, B.; Zhou, W.X.; Stanley, H.E. Systemic risk and spatiotemporal dynamics of the US housing market. *Sci. Rep.* **2014**, *4*, 3655. [CrossRef]

10. Boccaletti, S.; Latora, V.; Moreno, Y.; Chavez, M.; Hwang, D. Complex networks: Structure and dynamics. *Phys. Rep.* **2006**, *424*, 175–308. [CrossRef]

11. Barabási, A. *Network Science*; Cambridge University Press: Cambridge, UK, 2016.

12. Barzel, B.; Barabási, A. Universality in network dynamics. *Nat. Phys.* **2013**, *9*, 673–681. [CrossRef]

13. Caldarelli, G. *Scale-Free Networks: Complex Webs in Nature and Technology*; Oxford University Press: Oxford, UK, 2007.

14. Zhao, L.; Li, W.; Fenu, A.; Podobnik, B.; Wang, Y.; Stanley, H.E. The q-dependent detrended cross-correlation analysis of stock market. *J. Stat. Mech. Theory Exp.* **2018**, *2018*, 023402. [CrossRef]

15. Squartini, T.; Van Lelyveld, I.; Garlaschelli, D. Early-warning signals of topological collapse in interbank networks. *Sci. Rep.* **2013**, *3*, 3357. [CrossRef] [PubMed]

16. Ross, G.J. Dynamic multifactor clustering of financial networks. *Phys. Rev. E* **2014**, *89*, 022809. [CrossRef] [PubMed]

17. Huang, W.Q.; Zhuang, X.T.; Yao, S.; Uryasev, S. A financial network perspective of financial institutions' systemic risk contributions. *Phys. A* **2016**, *456*, 183–196. [CrossRef]

18. Guo, X.; Zhang, H.; Jiang, F.; Tian, T. Development of stock correlation network models using maximum likelihood method and stock big data. In Proceedings of the 2018 IEEE International Conference on Big Data and Smart Computing (BigComp), Shanghai, China, 15–17 January 2018; pp. 455–461.

19. Fortunato, S. Community detection in graphs. *Phys. Rep.* **2010**, *486*, 75–174. [CrossRef]

20. Onnela, J.; Kaski, K.; Kertész, J. Clustering and information in correlation based financial networks. *Eur. Phys. J. B* **2004**, *38*, 353–362. [CrossRef]

21. Kumar, S.; Deo, N. Correlation and network analysis of global financial indices. *Phys. Rev. E* **2012**, *86*, 026101. [CrossRef]

22. Nobi, A.; Lee, S.; Kim, D.H.; Lee, J.W. Correlation and network topologies in global and local stock indices. *Phys. Lett. A* **2014**, *378*, 2482–2489. [CrossRef]

23. Xia, L.; You, D.; Jiang, X.; Guo, Q. Comparison between global financial crisis and local stock disaster on top of Chinese stock network. *Phys. A* **2018**, *490*, 222–230. [CrossRef]

24. Nobi, A.; Maeng, S.E.; Ha, G.G.; Lee, J.W. Effects of global financial crisis on network structure in a local stock market. *Phys. A* **2014**, *407*, 135–143. [CrossRef]

25. Zhao, L.; Wang, G.J.; Wang, M.; Bao, W.; Li, W.; Stanley, H.E. Stock market as temporal network. *Phys. A* **2018**, *506*, 1104–1112. [CrossRef]

26. Namaki, A.; Shirazi, A.; Raei, R.; Jafari, G. Network analysis of a financial market based on genuine correlation and threshold method. *Phys. A* **2011**, *390*, 3835–3841. [CrossRef]

27. Musmeci, N.; Aste, T.; Di Matteo, T. Risk diversification: A study of persistence with a filtered correlation-network approach. *Ournal Netw. Theory Financ.* **2014**, *1*, 77–98. [CrossRef]

28. Cao, X.; Shi, Y.; Wang, P.; Chen, L.; Wang, Y. The evolution of network topology structure of Chinese stock market. In Proceedings of the 2018 IEEE 3rd International Conference on Big Data Analysis (ICBDA), Shanghai, China, 9–12 March 2018; pp. 329–333.

29. Peng, C.; Buldyrev, S.V.; Havlin, S.; Simons, M.; Stanley, H.E.; Goldberger, A.L. Mosaic organization of DNA nucleotides. *Phys. Rev. E* **1994**, *49*, 1685. [CrossRef] [PubMed]

30. Lyócsa, Š.; Vÿrost, T.; Baumöhl, E. Stock market networks: The dynamic conditional correlation approach. *Phys. A* **2012**, *391*, 4147–4158. [CrossRef]

31. Coelho, R.; Hutzler, S.; Repetowicz, P.; Richmond, P. Sector analysis for a FTSE portfolio of stocks. *Phys. A* **2007**, *373*, 615–626. [CrossRef]

32. Eom, C.; Oh, G.; Jung, W.S.; Jeong, H.; Kim, S. Topological properties of stock networks based on minimal spanning tree and random matrix theory in financial time series. *Phys. A* **2009**, *388*, 900–906. [CrossRef]

33. Song, D.M.; Tumminello, M.; Zhou, W.X.; Mantegna, R.N. Evolution of worldwide stock markets, correlation structure, and correlation-based graphs. *Phys. Rev. E* **2011**, *84*, 026108. [CrossRef]

34. Yang, L.; Zhao, L.; Wang, C. Portfolio optimization based on empirical mode decomposition. *Phys. A* **2019**, *531*, 121813. [CrossRef]

35. Orsini, C.; Dankulov Mitrović, M.; Colomer-de Simón, P.; Jamakovic, A.; Mahadevan, P.; Vahdat, A.; Bassler, K.E.; Toroczkai, Z.; Boguná, M.; Caldarelli, G.; et al. Quantifying randomness in real networks. *Nat. Commun.* **2015**, *6*, 1–10. [CrossRef]

36. Oh, G.; Kim, H.Y.; Ahn, S.W.; Kwak, W. Analyzing the financial crisis using the entropy density function. *Phys. A* **2015**, *419*, 464–469. [CrossRef]

37. Yim, K.; Oh, G.; Kim, S. An analysis of the financial crisis in the KOSPI market using Hurst exponents. *Phys. A Stat. Mech. Its Appl.* **2014**, *410*, 327–334. [CrossRef]

38. Wiliński, M.; Sienkiewicz, A.; Gubiec, T.; Kutner, R.; Struzik, Z.R. Structural and topological phase transitions on the German Stock Exchange. *Phys. A* **2013**, *392*, 5963–5973. [CrossRef]

39. Živković, J.; Tadić, B.; Wick, N.; Thurner, S. Statistical indicators of collective behavior and functional clusters in gene networks of yeast. *Eur. Phys. J. B* **2006**, *50*, 255–258. [CrossRef]

40. Van Mieghem, P. *Graph Spectra for Complex Networks*; Cambridge University Press: Cambridge, UK, 2010.

41. Mitrović, M.; Tadić, B. Spectral and dynamical properties in classes of sparse networks with mesoscopic inhomogeneities. *Phys. Rev. E* **2009**, *80*, 026123. [CrossRef]

42. Farkas, I.J.; Derényi, I.; Barabási, A.; Vicsek, T. Spectra of "real-world" graphs: Beyond the semicircle law. *Phys. Rev. E* **2001**, *64*, 026704. [CrossRef]

43. Blondel, V.D.; Guillaume, J.L.; Lambiotte, R.; Lefebvre, E. Fast unfolding of communities in large networks. *J. Stat. Mech. Theory Exp.* **2008**, *2008*, P10008. [CrossRef]

44. Sherman, M. A short history of financial deregulation in the United States. *Cent. Econ. Policy Res.* **2009**, *7*, 1–17.

45. Coffee, J.C., Jr. What went wrong? An initial inquiry into the causes of the 2008 financial crisis. *J. Corp. Law Stud.* **2009**, *9*, 1–22. [CrossRef]

46. Baily, M.N.; Klein, A.; Schardin, J. The impact of the Dodd-Frank Act on financial stability and economic growth. *RSF Russell Sage Found. J. Soc. Sci.* **2017**, *3*, 20–47.

47. Akhigbe, A.; Martin, A.D.; Whyte, A.M. Dodd–Frank and risk in the financial services industry. *Rev. Quant. Financ. Account.* **2016**, *47*, 395–415. [CrossRef]

48. Barth, J.R.; Prabha, A.P.; Wihlborg, C. The Dodd-Frank act: Key features, implementation progress, and, financial system impact. In *The First Great Financial Crisis of the 21st Century: A Retrospective*; World Scientific: Singapore, 2016; pp. 337–376.

Check for updates

Analysis of Worldwide Time-Series Data Reveals Some Universal Patterns of Evolution of the SARS-CoV-2 Pandemic

Marija Mitrović Dankulov[1]*, Bosiljka Tadić[2,3] and Roderick Melnik[4,5]

[1]*Institute of Physics Belgrade, University of Belgrade, Belgrade, Serbia, [2]Department for Theoretical Physics, Jožef Stefan Institute, Ljubljana, Slovenia, [3]Complexity Science Hub, Vienna, Austria, [4]MS2Discovery Interdisciplinary Research Institute, M2NeT Laboratory and Department of Mathematics, Wilfrid Laurier University, Waterloo, ON, Canada, [5]BCAM–Basque Center for Applied Mathematics, Bilbao, Spain*

Predicting the evolution of the current epidemic depends significantly on understanding the nature of the underlying stochastic processes. To unravel the global features of these processes, we analyse the world data of SARS-CoV-2 infection events, scrutinising two 8-month periods associated with the epidemic's outbreak and initial immunisation phase. Based on the correlation-network mapping, K-means clustering, and multifractal time series analysis, our results reveal several universal patterns of infection dynamics, suggesting potential predominant drivers of the pandemic. More precisely, the Laplacian eigenvectors localisation has revealed robust communities of different countries and regions that break into clusters according to similar profiles of infection fluctuations. Apart from quantitative measures, the immunisation phase differs significantly from the epidemic outbreak by the countries and regions constituting each cluster. While the similarity grouping possesses some regional components, the appearance of large clusters spanning different geographic locations is persevering. Furthermore, characteristic cyclic trends are related to these clusters; they dominate large temporal fluctuations of infection evolution, which are prominent in the immunisation phase. Meanwhile, persistent fluctuations around the local trend occur in intervals smaller than 14 days. These results provide a basis for further research into the interplay between biological and social factors as the primary cause of infection cycles and a better understanding of the impact of socio-economical and environmental factors at different phases of the pandemic.

Keywords: complex networks, k-means, time-series analysis, spectral analysis, community structure, SARS-CoV-2

1 INTRODUCTION

In cooperative social dynamics [1, 2], the genesis of a collective phenomenon arising from contagious social interactions involves mechanisms of self-organised criticality [3, 4]. It depends on each individual involved, based on its actual contacts, psychology and behaviour. In the presence of viruses, these mechanisms are additionally shaped by firm biological factors. Recent developments of SARS-CoV-2 pandemic [5, 6] revealed a specific global phenomenon emerging from the stochastic multi-scale processes. The infection incidence occurs with a high temporal resolution at the

interactions between the virus and human hosts, whose biological features, social behaviours and mobility [7] significantly contribute to the epidemic's spreading [8]. At the molecular scale, the virus-host interactions [9–11] crucially depend on the virus biology and genetic factors determining the host's immunity towards the virus in question [12, 13]. Thus, the occurrence of an infection event and the infection manifestation may lead to a range of different scenarios from asymptomatically infected to severe health issues and fatalities [14–17]. Multiple other factors may play a role [18], depending on the population genetic features and social life [19]. They include cultural, political and economic aspects, official and spontaneous reaction to the crisis, and the organisation of the health care system, all of which may significantly differ between different geographical locations [20]. Moreover, the actual impact of these factors changes over time as the epidemic develops, in particular, since the appropriate vaccines targeting SARS-CoV-2 viruses [3, 21] are available, thus enabling potentially substantial changes due to massive immunisation of the population given the theoretical analysis in [22–24]. Attempts were made to identify different parameters that may influence the epidemic and estimate their mutual interdependence and impact. For example, the human-development index, built-up-area-per-capita, and the immunisation coverage appear among the statistically high-ranking drivers of SARS-CoV-2 epidemic [18].

In addition, temporal variations occur at all scales, from the virus mutations [11] to changed behaviours of each individual and population groups, e.g., due to the government imposed measures [6, 25], or adaptation caused by the awareness of the current epidemiological situation [26, 27]. These variations increase the stochasticity of the infection and contact processes, making the prediction of their output even more difficult. For real-time epidemic management and the predictions of further developments, it is crucial to understand the nature of the underlying stochastic processes and the factors that can significantly influence them. For this purpose, the empirical data analysis and theoretical modelling [28] provide complementary views of these complex processes. For example, agent-based models capture the interplay of the bio-social factors at the elementary scale of the virus-host interactions at high temporal resolution [8, 29–37]. On the other hand, more traditional compartmental models [38] consider a coarse-grained picture of the population groups having different roles in the process. Another research line aims at the mathematical description of the exact empirical data, in particular, for the outbreak phase [39, 40]. For instance, different studies provided tangible arguments for the cause of the changing shape of the infection curve comprising the appearance of linear and power-law segments [41, 42], prolonged stagnation periods, and multiple waves [43]. Since the beginning of the epidemic, empirical data were collected over different countries or provinces [44]. Despite the coarse-grained spatial and temporal structure (daily resolution), these data may contain relevant information about the temporal aspects of the epidemic at different geographical locations. Previous studies, based on the empirical data regarding the dynamics of interacting units in many complex systems, provided valuable information about the related stochastic processes. Some striking examples across different spatial and temporal scales include the influence of the world financial index dynamics on different countries [45, 46], traffic jamming [47, 48], brain-to-brain coordination dynamics [49, 50], and the cooperative gene expressions along different phases of the cell cycle [51, 52]. Similarly, the collected data of SARS-CoV-2 spreading enable a possibility to investigate the infection dynamics in various details and more appropriate modelling of the emergent behaviours. In this respect, a larger-scale picture may emerge by studying temporal fluctuations of the world infection dynamics. More subtle questions regard the indicators for hidden mechanisms arising from the interplay of the above-mentioned biological factors and different social behaviours [8, 27, 29, 53, 54] behind the observed epidemic development.

In this work, we address some of these critical issues aiming to unveil the inherent features of infection dynamics by studying time-series data that are publicly available at GitHub [44] collected over different countries or regions (provinces). Using the datasets of the *daily recorded* number of confirmed infection cases, we consider two separate segments of time series. Defining two distinct 8-month periods in the epidemic's evolution is motivated by the appearance of SARS-CoV-2 vaccines in the latter period, enabling pharmaceutical intervention measures not available in the outbreak phase, cf. **Figure 1**. Namely, the records for the first 8 months of the epidemic, starting from the first registered case in each country, represent the epidemic's *outbreak phase*. Meanwhile, the last 8 months (preceding the data collection on 30 September 2021), during which the pharmaceutical intervention was available in most of the countries, characterises the initial *immunisation phase* of this pandemic. Our quantitative analysis comprises three levels of information: the network mapping and spectral analysis, K-means clustering of pairs of time series, and detrended fractal analysis of individual time series. Each of these methods provides just partial information about the studied dynamics. We combine them to create a comprehensive picture of the course of the epidemic in different countries and how they relate to each other. In addition to quantifying the differences between the outbreak and immunisation phase, our results reveal two global features of the SARS-CoV-2 pandemic. Firstly, the worldwide groups of countries (and provinces) robustly appear in clusters having a similar temporal evolution of the infection dynamics. This clustering suggests that the environmental and socio-economical factors and government-imposed measures can certainly influence small-scale fluctuation characteristics of the clusters but do not significantly change the course of the process on larger scales. Secondly, the epidemic evolution exhibits ubiquitous waves driven by the cyclic infection dynamics, where several typical cycles appear associated with the identified clusters. Again, the shape of these specific cycles coincides with the mentioned clustering mechanisms. Hence, their origin and potential control will remain challenging within purely social measures. A more detailed analysis of the complex feedback between biological and social factors at all scales is needed.
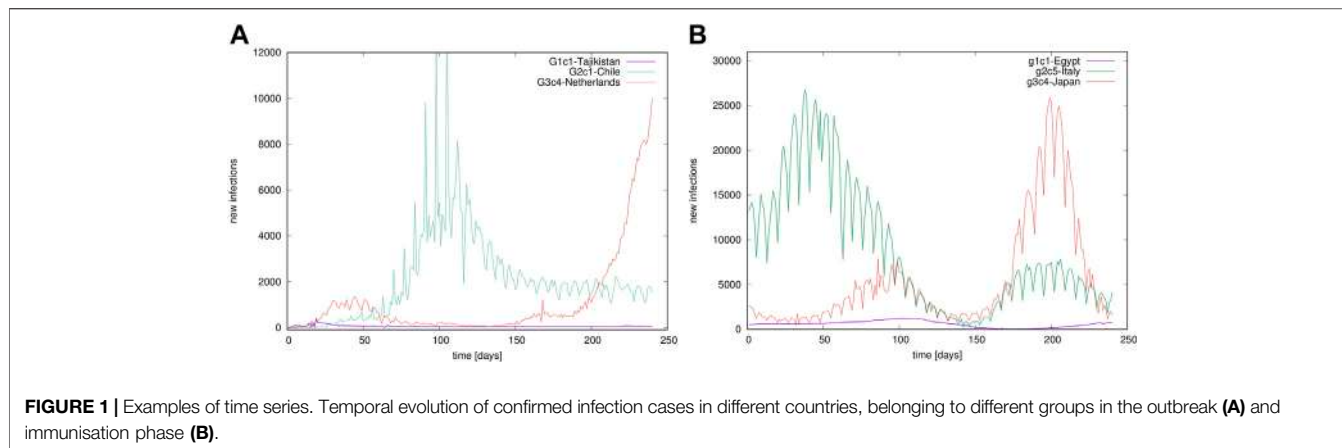
FIGURE 1 | Examples of time series. Temporal evolution of confirmed infection cases in different countries, belonging to different groups in the outbreak **(A)** and immunisation phase **(B)**.

2 MATERIALS AND METHODS

2.1 Data Acquisition, Preparation, and Mapping

We consider the worldwide data of the number of new infection cases downloaded from GitHub [44]. The dataset contains the number of daily detected new cases for 279 countries including separated data for some provinces. For this work, we select time series in two eight-mount periods comprising the epidemic's outbreak phase (starting from the first registered case in a given country or province) and the immunisation phase (22 January 2020 until 30 September 2021). The corresponding number of countries and provinces with the active epidemic's data traced in both periods is 255. For instance, the first case in France was detected on 24 January 2020, and thus the outbreak time series covers the period from that date until 19 September 2020. However, Slovenia had the first registered case on 5 March 2020; hence its outbreak time series cover 5 March until 30 October 2020. Meanwhile, the immunisation period is from 3 February 2021 to 30 September 2021, equal for all considered countries and provinces.

By mapping these datasets, we obtain two correlation networks for the outbreak and immunisation phase, respectively, where the network's links stand for significant positive correlations. We first compute the Pearson's correlation coefficient for the corresponding pairs (i, j) of the time series

$$C_{ij}^\tau = \frac{1}{N_T - 1} \sum_{t=1}^{N_T} \frac{X_i^\tau(t) - \mu_i^\tau}{\sigma_i^\tau} \frac{X_j^\tau(t) - \mu_j^\tau}{\sigma_j^\tau}, \quad (1)$$

where $\tau \in \{O, V\}$, μ_i^τ is average value of the time series of country i during period τ, σ_i^τ is standard deviation of time series $X_i^\tau(t)$, and $N_T = 240$ is the length of time series. To remove spurious correlations, we apply the filtering procedure standardly used in these type of network mapping [47, 49, 51]. More precisely, the matrix elements C_{ij}^τ are first transformed to the interval [0, one] by $CP_{ij}^\tau = \frac{1}{2}(C_{ij}^\tau + 1)$, and then multiplied by a factor M_{ij}^τ which is obtained in the following way. From the rows i and j, the diagonal elements are removed and the considered elements CP_{ij}^τ and CP_{ji}^τ
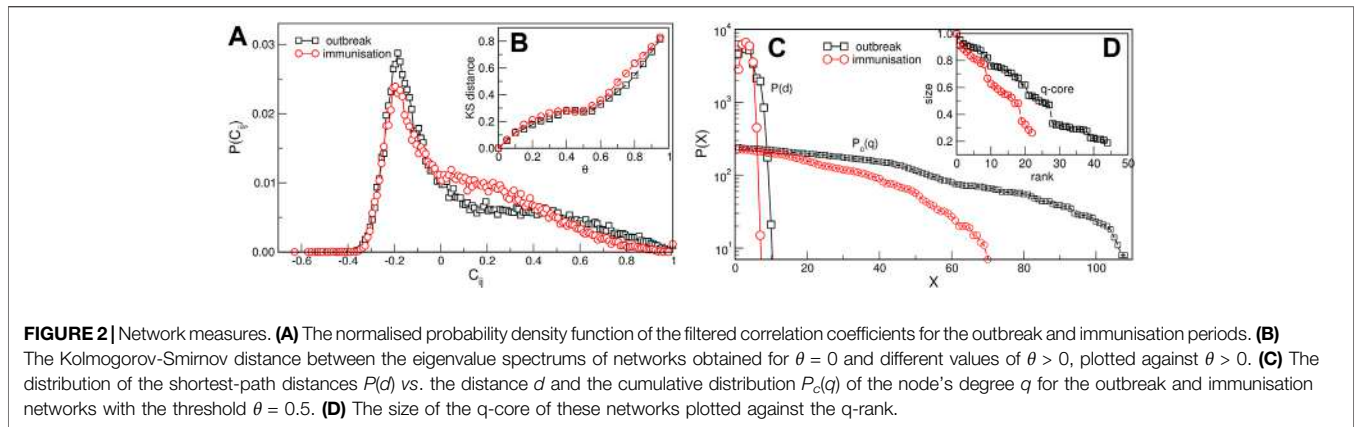
are placed at the beginning of the row i and j, respectively, thus obtaining two $n = N - 1$ dimensional vectors $\widetilde{CP_i^\tau}$ and $\widetilde{CP_j^\tau}$. Then M_{ij}^τ is computed as Pearson's coefficient between these two vectors. The matrix element of the filtered correlation matrix $\overline{C_{ij}^\tau} = M_{ij}^\tau CP_{ij}^\tau$ is then mapped back to the interval $[-1, 1]$. Finally, the elements of the network's adjacency matrix are defined as $A_{ij}^\tau = 1$ when the matrix elements $\overline{C_{ij}^\tau} > \theta$ exceed a specified threshold value θ, and zero otherwise. The threshold value θ is determined concerning the network's spectral properties, as described below.

2.2 Network's Spectral Analysis and Community Detection

The above-described data mapping should lead to undirected unweighted networks; the nodes represent countries (or provinces), and links indicate the positive correlations between infection incidences exceeding a threshold θ. We use the spectral properties of networks to obtain the adequate threshold value, where the guiding criteriums are the network's sparseness and the relative stability of the community structure. Starting from $\theta = 0$, we increase it by the value 0.05 and solve the eigenvalue problem of the corresponding adjacency matrix, $\mathbf{A}\nu_i = \lambda_i \nu_i|_\theta$, and calculate the spectrum $\{\lambda_1, \ldots, \lambda_N\}_\theta$ for each threshold θ. We compare the adjacency matrix spectrum for network $\theta = 0$ with spectra of each network obtained for considered $\theta > 0$ using Kolmogorov-Smirnov (KS) distance. For each $\theta > 0$ we obtain one KS distance and plot its dependence of θ, see **Figure 2B**. The KS-distance has a minimum of around $\theta = 0.5$ for the outbreak and immunisation phase. We use this value of θ to obtain the networks used in our analysis.

We study the community structure of the networks for the outbreak and immunisation period using spectral analysis and the eigenvalue problem of the normalised Laplacian related to the network's adjacency matrix. In mathematics theory [55, 56], the number of smallest non-zero eigenvalues of the Laplacian matrix is a good indicator of the number of communities. The matrix elements of the normalised Laplacian for undirected binary network represented by the adjacency matrix **A** are defined as

FIGURE 2 | Network measures. **(A)** The normalised probability density function of the filtered correlation coefficients for the outbreak and immunisation periods. **(B)** The Kolmogorov-Smirnov distance between the eigenvalue spectrums of networks obtained for $\theta = 0$ and different values of $\theta > 0$, plotted against $\theta > 0$. **(C)** The distribution of the shortest-path distances $P(d)$ vs. the distance d and the cumulative distribution $P_c(q)$ of the node's degree q for the outbreak and immunisation networks with the threshold $\theta = 0.5$. **(D)** The size of the q-core of these networks plotted against the q-rank.

$$L_{ij} = \delta_{ij} - \frac{A_{ij}}{\sqrt{q_i q_j}}, \qquad (2)$$

where q_i and q_j are degrees of nodes i and j. For the normalised Laplacian [2], we solve the eigenvalue equation $\mathbf{L} v^L = \lambda_i^N v_i^L$ and determine all eigenvalues and eigenvectors. In the case of a connected network, these eigenvalues are non-negative. One zero-eigenvalue appears with strictly positive eigenvector's components [55]. Consequently, the orthogonal eigenvectors corresponding to the three smallest non-zero eigenvalues localise on the communities of the network. Hence, the scatter plot of the components of these eigenvectors shows a branching structure. Each branch contains indexes of the non-zero eigenvector components, that is, the nodes belonging to a network's community [56].The size of the q-core of the networks is determined by removing the nodes with the increasing degree q. Several other graph properties are determined, and the networks are visualised using *Gephi* software [57].

2.3 K-Means Clustering of Time Series

The implementation of the K-means algorithm for clustering of time series in Python known as *tslearn* [58] is used. K-means is an unsupervised machine learning algorithm that aggregates data points according to similarities, starting with K randomly positioned centroids. Based on these centroids, data points are assigned to the centroid closest to that data point according to some distance metric. The algorithm consists of a certain number of iterative (repetitive) calculations used to optimise the positions of the centroids. Considering each time series of length N_T as a data point in N_T dimensional space, the appropriate measures enable calculating the distances between these data points. We use the Dynamic Time Wrapping (DTW) algorithm to align time series with centroids and measure their similarities. The DTW is a widely used algorithm measuring similarities between time series and their classification. It does not transform the time series; it only finds the minimal distance between time series beyond simple correlation. Specifically, it performs an optimal alignment between two time series by matching the indices from the first time series to the second time series, subject to several constraints. The mapping of indices from the first series to the second series must be monotonically increasing. For the indices $i > j$ from the first

time series, there must be two indices from the second series $l > k$ such that i is matched with l and j is matched with k. Meanwhile, the first index from the first series must match the first index of the second time series, and similarly, the last index from the first series must be matched to the last index of the second time series, but these points may have more other matches. The optimal alignment is the one that satisfies all of these restrictions with the minimal cost, where cost is the sum of absolute differences of values for each matched pair of indices. The DTW distance in the K-means algorithm is the value of cost. We use the K-means algorithm with DTW distance to cluster time series and find centroids. Each centroid is again a time series that describes the average behaviour of the time series belonging to one cluster.

2.4 Trends and Fractal Analysis of Time Series

Temporal fluctuations are studied by the fractal detrended analysis of each time series. For each time series $x(k)$, $k = 1, 2, \cdots T$, the profile $Y(i) = \sum_{k=1}^{i} (x(k) - <x>)$ of the time series is divided in N_s segments of the length n. The fluctuation function $F_q(n)$ with the varied segment length n is defined as

$$F_q(n) = \left(\frac{1}{N_s} \sum_{\mu=1}^{N_s} \left[F^2(\mu, n) \right]^{q/2} \right)^{1/q} \sim n^{h_q}, \qquad (3)$$

Here, $F^2(\mu, n) = \frac{1}{n} \sum_{i=1}^{n} [Y((\mu - 1)n + i) - y_\mu(i)]^2$ is the standard deviation from a local trend $y_\mu(i)$ on the segment μ. For $q = 2$, we determine the Hurst exponent h_2 from the straight-line segments of the log-log plot of the fluctuation function $F_2(n)$. For the multifractal analysis, the values of $q \in [-4, 4]$ are varied.

To determine cyclic trends, we use the local adaptive detrending algorithm, see [59, 60], where time series is divided into segments of the length $2m + 1$ overlapping over $m + 1$ points. The polynomial interpolation is applied in each segment, and its contribution in the overlapped region is weighted such that it decreases linearly with the distance from the segment's centre. As stated in the Introduction, we consider worldwide recorded time series of the infection cases. For illustration, a few examples of time series recorded in different countries are shown in **Figure 1**.
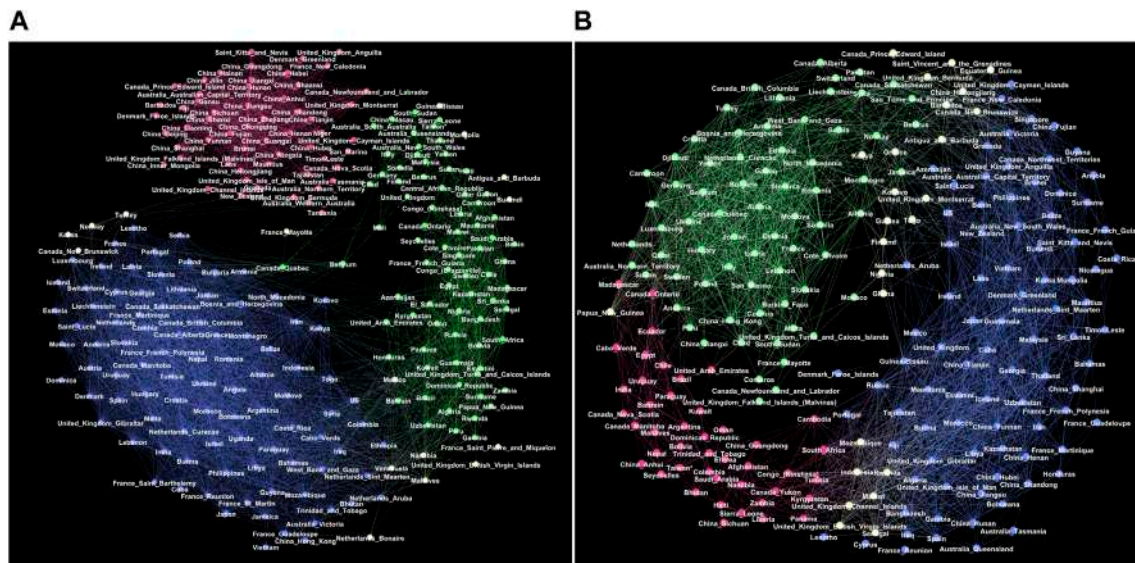
FIGURE 3 | Giant connected components of the correlation networks at the threshold θ =0.5 for the outbreak period **(A)** and the immunisation period **(B)**. Red, green and blue colours indicate groups of nodes in three respective communities G1, G2, G3 in the outbreak network, and g1, g2, g3 in the immunisation period network, determined by the eigenvector-localisation, see text and **Figure 4**. Unclassified borderline nodes are shown in white colour. Labels on nodes identify the corresponding country or province. The complete lists of nodes in each community are given in **Supplementary Tables S1–S6** in Supplementary Information.

2.5 The Correlation Networks Mapping in the Outbreak and Immunisation Phase

The network mapping is based on the cross-correlation coefficient C_{ij} of the pairs of time series $\{i, j\}$ and a suitably selected threshold. Hence, the correlations exceeding the threshold θ are accepted, making the adjacency-matrix elements $A_{ij}(\theta) = \Theta(C_{ij} - \theta) - \delta_{ij}$ of an undirected unweighted network. Before selecting the threshold, a filtering procedure was applied to the complete correlation matrix to enhance the positive correlations of interest in this work (see Methods). The applied methodology was proved useful in quantifying correlations of time series in diverse type of data [45, 47–52]. **Figure 2A** shows probability distributions of filtered correlations coefficients for the outbreak and immunisation period. While both probability distributions have a peak at a value $C_{ij} < 0$, they have slightly different shapes. They both have a pronounced tail for positive values of correlation coefficients, where the distribution $P(C_{ij})$ for the outbreak period has a slower decay at correlations $C_{ij} > 0.2$. The appropriate threshold is selected considering changes in spectral properties of the adjacency matrix with the increasing threshold, as explained in the following. **Figure 2B** shows the Kolmogorov-Smirnov (KS) distance between the eigenvalues of the $A_{ij}(\theta)$ compared to the one at $\theta = 0$ depending on the threshold θ for the outbreak and immunisation networks. We see that the KS distance grows slowly with θ up to the value ~ 0.4; meanwhile, the growth becomes rapid for the values of $\theta > 0.5$ for both networks, suggesting a profound change in the networks' structure when the threshold exceeds $\theta = 0.5$. Thus, we select this turning point as the optimal threshold value. Moreover, the networks obtained by applying the threshold weight $\theta = 0.5$ are sufficiently sparse; meanwhile, their spectral properties do not

differ drastically from the corresponding outbreak and immunisation period networks at $\theta = 0$ containing all positive correlations. The resulting networks for $\theta = 0.5$ are visualised in **Figure 3**. See also Supplementary Information (SI) for more details.

The giant connected component of each network exhibits a community structure, i.e., the occurrence of groups of nodes that are better connected among themselves than with the nodes outside that group, cf. **Figure 3**. The identity of nodes comprising each community is determined using the localisation of the eigenvectors associated with the three lowest nonzero eigenvalues of the normalised Laplacian operator [55], as explained in Methods [56]. The eigenvalues of the normalised Laplacians for two networks are shown in ranking order in **Figure 4**, middle panel. Several lowest nonzero eigenvalues appear to be separated from the bulk in both networks. This network feature is compatible with the existence of mesoscale communities, on which the corresponding eigenvectors tend to localise [55, 56]. The scatterplots of the eigenvectors associated with three lowest nonzero eigenvalues, see **Figure 4**, show three differentiable branches, here indicated as $G1$, $G2$, $G3$ for the outbreak, and $g1$, $g2$, $g3$ for the immunisation phase network. The indexes with a nonzero component of the eigenvectors in each branch mark the IDs of the nodes belonging to the corresponding community. The complete lists of nodes in each community (group) are given in **Supplementary Tables S1–S6** in SI.

Even though both networks exhibit three major communities, the structural differences between the two networks in **Figure 3** are apparent. They indicate the corresponding differences in the fluctuations of the infection rates in the world regions during the immunisation phase, compared to the epidemic's outbreak, when
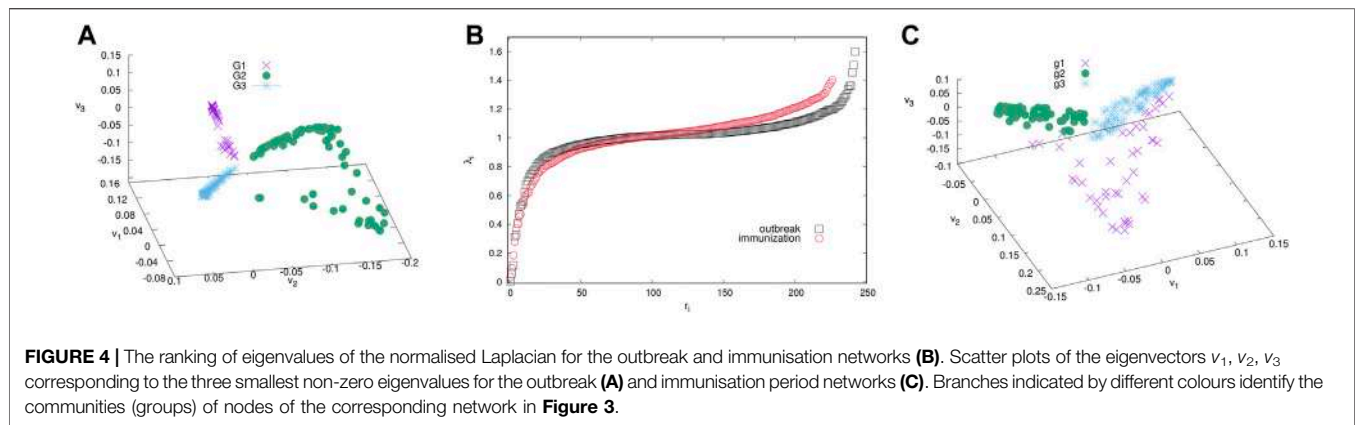
FIGURE 4 | The ranking of eigenvalues of the normalised Laplacian for the outbreak and immunisation networks **(B)**. Scatter plots of the eigenvectors v_1, v_2, v_3 corresponding to the three smallest non-zero eigenvalues for the outbreak **(A)** and immunisation period networks **(C)**. Branches indicated by different colours identify the communities (groups) of nodes of the corresponding network in **Figure 3**.

TABLE 1 | For the outbreak and immunisation period networks: the number of nodes N, edges E, and triangles #$\triangle$; the graph diameter D and density ρ; the average path length $<\ell>$, degree $<q>$, and clustering coefficient $<C_c>$.

	N	E	D	$<\ell>$	$<q>$	$<C_c>$	#$\triangle$	ρ
Outbreak	243	9,284	11	3.55	38.2	0.735	58,508	0.158
immunis	227	5,660	7	3.10	24.9	0.656	18,857	0.112

the whole population was practically susceptible to the infection. These differences are quantified by several graph measures, see **Figures 2A–D** and **Table 1**. Compatible with these graph-theory measures are the span of the exponentially-decaying degree distributions $P_c(q)$ and different distributions of the shortest-path distances $P(d)$, shown in **Figure 2C**. We also show the prominent differences in the q-core structure of these networks, cf. **Figure 2D**.

More importantly, the majority of nodes that belong to the same community in the outbreak phase network appear to be a part of entirely different communities in the immunisation phase network, cf. **Figure 3** and the corresponding lists in Supplementary Information. More precisely, we find that only 625 edges established in the outbreak phase persist in the immunisation phase network. They are shown in **Supplementary Figure S2** left, in Supplementary Information. A more systematic comparison is made by computing the overlap (Jaccard index defined in Methods) for the correlation networks determined from the successive 2-month intervals, see **Supplementary Figure S2**, right. The overlap systematically remains below 15%, suggesting that the fluctuation patterns at these intervals can vary between the countries or even provinces within the same country.

2.6 K-Means Clustering and Multi-Fractality of Time Series Within Identified Communities
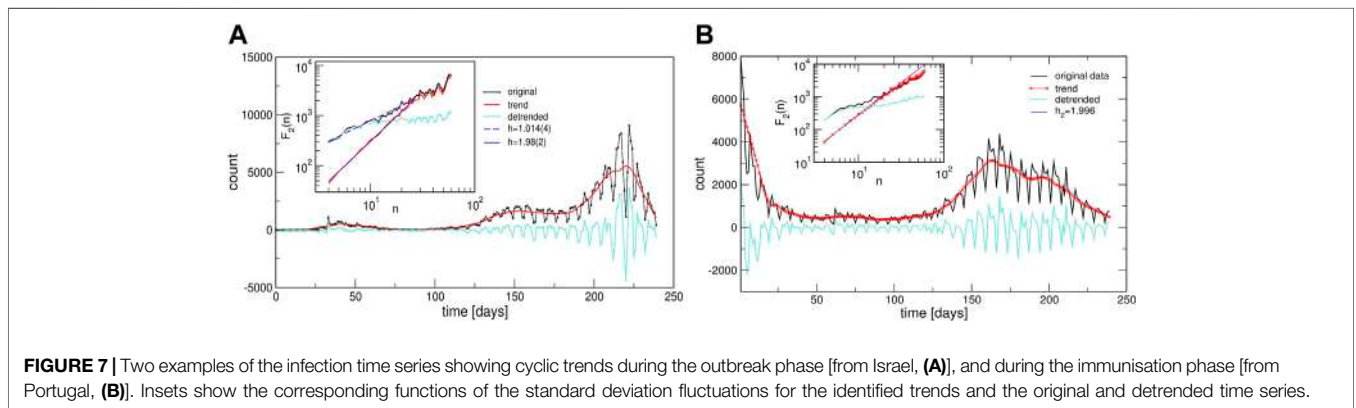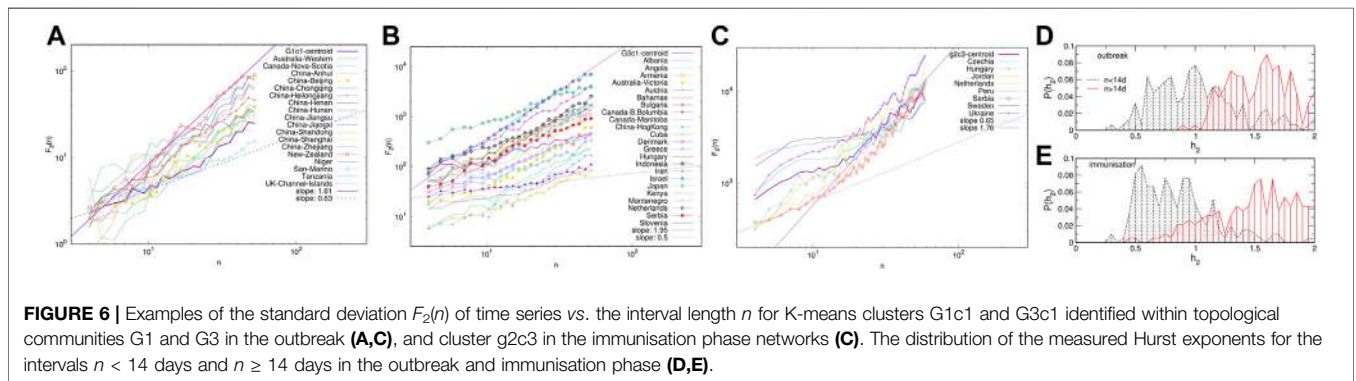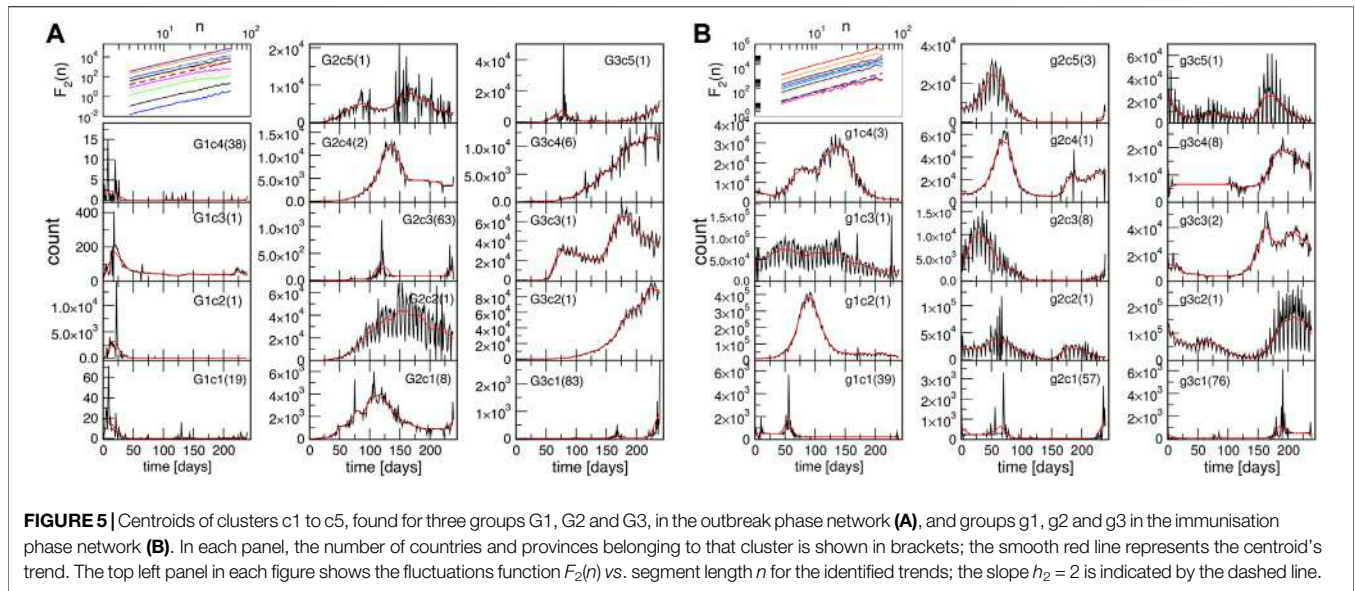
To further explore the nature of temporal fluctuations of the infection time series of the countries and provinces within each community found using spectral analysis, we apply the K-means

algorithm adapted for time series analysis [58], see Methods. It appears that each topological community is further partitioned into several clusters, for example, $G1c1\cdots G1c4$, for the group $G1$, and so on. Inside each cluster, the corresponding time series have a similar evolution pattern. Hence, the cluster's typical time series (centroid) is determined for each identified cluster. The results are shown in **Figure 5** both for the outbreak and immunisation phase; in the figure legends, the number of countries or provinces belonging to a given cluster is indicated in the brackets in each panel. The names of countries and provinces belonging to each cluster in each group are given in **Supplementary Tables S1–S6** in Supplementary Information. Notably, in each network's group, there is one large and one medium-size cluster. Meanwhile, there are several single-country centroids; as a rule, they indicate a large-population country.

Next we consider the fluctuation function $F_2(n)$ *vs.* the interval length n for each time series separately, see some examples in **Figure 6**, and **Supplementary Figure S4** in SI. We realised that the similarity of the time series belonging to each cluster manifests itself in the apparent similarity of the slopes of their fluctuation function, which defines the corresponding Hurst exponent. As **Figure 6** shows, two different slopes of the fluctuation function can be identified for a majority of time series. At the intervals $n < 14$, a Hurst exponent $0.5 \lesssim h_2 \lesssim 1$ can be determined, indicating persistent fluctuations occurring at these time intervals. Meanwhile, an exponent $h_2 > 1$, characteristic to the fractional Brownian motion, is found for $n \geq 14$. In some cases, the determined Hurst exponent reaches values close to two. The histograms of the observed Hurst exponents are shown in **Figures 6D,E**. Compatible with the grouping and different shapes of centroids in the immunisation phase, the distributions of lower and higher values of the Hurts exponents are also different with the increased incidence of the value $h_2 \simeq 0.5$ (white noise), and $h_2 \simeq 2$ (periodic signals) in the immunisation phase. In the following, we show that these large values of the Hurst exponent in many of the studied time series can be related to the occurrence of *cyclic trends*.

Two prominent examples are shown in **Figure 7**. The methodology of determining local trends in these time series is described in Methods. The original time series shows a cyclic

FIGURE 5 | Centroids of clusters c1 to c5, found for three groups G1, G2 and G3, in the outbreak phase network **(A)**, and groups g1, g2 and g3 in the immunisation phase network **(B)**. In each panel, the number of countries and provinces belonging to that cluster is shown in brackets; the smooth red line represents the centroid's trend. The top left panel in each figure shows the fluctuations function $F_2(n)$ *vs.* segment length n for the identified trends; the slope $h_2 = 2$ is indicated by the dashed line.

FIGURE 6 | Examples of the standard deviation $F_2(n)$ of time series *vs.* the interval length n for K-means clusters G1c1 and G3c1 identified within topological communities G1 and G3 in the outbreak **(A,C)**, and cluster g2c3 in the immunisation phase networks **(C)**. The distribution of the measured Hurst exponents for the intervals $n < 14$ days and $n \geq 14$ days in the outbreak and immunisation phase **(D,E)**.

FIGURE 7 | Two examples of the infection time series showing cyclic trends during the outbreak phase [from Israel, **(A)**], and during the immunisation phase [from Portugal, **(B)**]. Insets show the corresponding functions of the standard deviation fluctuations for the identified trends and the original and detrended time series.

trend, where the cycle length can vary from region to region. A separate analysis of the fluctuation functions for the trend and the fluctuations around the local trend (detrended signal) reveals that the trend drives the fluctuations beyond the intervals of approximately 14 days; see the insets to **Figure 7**. The trend has true cyclic fluctuations (the Hurst exponent equals two, within error bars) in the range up to $n \lesssim 30$ days. Meanwhile,

beyond this range, both the original signal and trend have a lower Hurst exponent in the range $h_2 \gtrsim 1$, characterising a fractional Brownian motion. By extending a similar analysis to the above-mentioned typical time series (centroids), we find that they also exhibit cyclic trends but with different cycles characterising different clusters of countries. The corresponding trends are also shown in each panel of **Figure 5** as a red line on the top

of the related centroid. The trend fluctuation functions $F_2(n)$ vs. n shows the cycle characteristics in a large range of the intervals n, cf. top left panels of **Figure 5**. They differ from cluster to cluster and, even for the same country, the cycles also differ in the outbreak and immunisation phase. Generally, larger cycles (in the length and amplitude) are observed in the immunisation phase as compared to the outbreak period, cf. **Supplementary Figure S5** in SI. Remarkably, these findings imply that the cycles (or the infection waves) represent an inherent feature of current pandemic which may have some long-lasting consequences.

3 DISCUSSION AND CONCLUSION

In search of universal characteristics of infection dynamics, we have analysed the worldwide empirical data of the SARS-CoV-2 epidemic [44], focusing on the new-infection time series with a daily resolution. The data are purposefully divided into two periods, corresponding to the epidemic's outbreak and the initial immunisation phase, respectively. Three complementary methods of quantitative analysis have been performed. Specifically, we have analysed the mesoscopic structure of the networks, which embody the significant pairwise correlations among the infection time series of different countries or provinces. The further similarity in the pairs of time series has been analysed by K-means clustering. Finally, the fluctuation function of each time series has been determined using the detrended time series analysis. Our analysis has revealed global clustering and several universal features of the infection dynamics. Our main conclusions are:

- *the worldwide clustering* represented by fourteen temporal patterns of evolution of infection reveals significant similarities transcending geographical regions;
- *the cyclic trends* dominate the infection fluctuations, implying the prevalent infection waves and multi-scale fluctuations around these cycles; typically determined cycles appear in conjunction with the identified clusters;
- *the immunisation phase differs* from the epidemic outbreak phase in all measures considered here, thus quantifying the impact of the (partial) immunisation coverage on the underlying stochastic process and the course of the pandemic.

The mesoscopic (community) structure, as shown in **Figure 3** is one of the striking characteristics of the infection-correlation networks; remarkably, it occurs already at zero thresholds, see **Supplementary Figure S1** in SI. What comes as a surprise is that these communities constitute almost entirely different nodes (countries or provinces) in the immunisation phase compared to the outbreak phase. Only a few edges established during the outbreak phase persist throughout the entire evolution of the epidemic, as shown in **Supplementary Figure S2** in SI. Consequently, the same applies to the contents of the clusters found in these two phases, cf. **Supplementary Tables S1–S6** in SI. Notably, a given geographic location and potentially similar cultural and economic development levels, similar healthcare systems and other related factors play some role. However,

even such regional groups appear to be a part of a worldwide cluster in both representative phases of the pandemic. Such a picture probably emerges under another dominant driver, common to countries at different locations, and with different cultural and economic developments. In this context, the *biology factors*, the virus mutations in the interplay with the *social behaviour* of individuals and groups in the crisis seems to be of the primary importance for the genesis of sustained infection waves, quantified by cyclic trends in different clusters, cf. **Figure 7**. Our analysis suggests that the waves are ubiquitous in all countries and regions in both representative phases of the pandemic. Meanwhile, the timing, duration and amplitude of these waves vary between different clusters of countries and provinces, likely depending on the applied measures and the corresponding variations in the population behaviours. Moreover, the small-scale fluctuations around these cyclic trends seem to be more region-specific, and depending on the immunisation measures; two comparative examples are shown in **Supplementary Figure S5** in SI. A more systematic analysis of these fluctuations and the impact of the immunisation level on the infection dynamics merits future study.

Our analysis of the world infection dynamics of the SARS-CoV-2 pandemic revealed several universal features of the underlying multiscale stochastic processes that go beyond the geographical impact, locally-imposed governmental measures, and partial immunisation phases. Indeed, while these measures are truly valuable for short-term effects, saving lives, and maintaining the functional healthcare system in each country [25], they are much less effective in changing the fundamental nature of the infection process, rooted in the interplay of biology and social behaviours. This work has provided an in-depth analysis of the pandemic's fundamental phases with an overview that can guide further research into the nature of biosocial interdependencies. The latter factor plays a critical role in the SARS-CoV-2 evolution, where individual biological features of the participants and their role in the collective behaviours need to be better understood. Our effective long-term management of the pandemic and prediction of its future developments rely upon our ability to continue unfolding critical attributes of the underlying biosocial stochastic dynamics.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: https://github.com/CSSEGISandData/COVID-19/.

AUTHOR CONTRIBUTIONS

BT, RM, and MM designed research, MM collected data, MM and BT contributed program tools and performed analysis, BT, MM, and RM analysed data, BT produced figures, BT and RM wrote the manuscript, all authors reviewed the manuscript.

FUNDING

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fphy.2022.936618/full#supplementary-material

REFERENCES

1. Jusup M, Holme P, Kanazawa K, Takayasu M, Romić I, Wang Z, et al. Social Physics. *Phys Rep* (2022) 948:1–148. doi:10.1016/j.physrep.2021.10.005

2. Perc M, Jordan JJ, Rand DG, Wang Z, Boccaletti S, Szolnoki A. Statistical Physics of Human Cooperation. *Phys Rep* (2017) 687:1–51. doi:10.1016/j.physrep.2017.05.004

3. Tadić B, Melnik R. Self-organised Critical Dynamics as a Key to Fundamental Features of Complexity in Physical, Biological, and Social Networks. *Dynamics* (2021) 1:181–97.

4. Tadić B, Dankulov MM, Melnik R. Mechanisms of Self-Organized Criticality in Social Processes of Knowledge Creation. *Phys Rev E* (2017) 96:032307. doi:10.1103/PhysRevE.96.032307

5. Chilamakuri R, Agarwal S. Covid-19: Characteristics and Therapeutics. *Cells* (2021) 10:206. doi:10.3390/cells10020206

6. Gerotziafas GT, Catalano M, Theodorou Y, Dreden PV, Marechal V, Spyropoulos AC, et al. The Covid-19 Pandemic and the Need for an Integrated and Equitable Approach: an International Expert Consensus Paper. *Thromb Haemost* (2021) 121:992–1007. doi:10.1055/a-1535-8807

7. Hâncean MG, Slavinec M, Perc M. The Impact of Human Mobility Networks on the Global Spread of Covid-19. *J Complex Networks* (2020) 8:cnaa041.

8. Tadić B, Melnik R. Modeling Latent Infection Transmissions through Biosocial Stochastic Dynamics. *PloS one* (2020) 15:e0241163. doi:10.1371/journal.pone.0241163

9. Doms RW. Basic Concepts. In: *Viral Pathogenesis*. Amsterdam: Elsevier (2016). p. 29–40. doi:10.1016/b978-0-12-800964-2.00003-3

10. Schneider M, Johnson JR, Krogan NJ, Chanda SK. The Virus-Host Interactome. In: *Viral Pathogenesis*. Amsterdam: Elsevier (2016). p. 157–67. doi:10.1016/b978-0-12-800964-2.00012-4

11. Callaway E. The Coronavirus Is Mutating - Does it Matter? *Nature* (2020) 585:174–7. doi:10.1038/d41586-020-02544-6

12. Lu C, Gam R, Pandurangan AP, Gough J. Genetic Risk Factors for Death with Sars-Cov-2 from the uk Biobank. *MedRxiv* (2020). doi:10.1101/2020.07.01.20144592

13. Zhang Y, Geng X, Tan Y, Li Q, Xu C, Xu J, et al. New Understanding of the Damage of Sars-Cov-2 Infection outside the Respiratory System. *Biomed Pharmacother* (2020) 127:110195. doi:10.1016/j.biopha.2020.110195

14. Cevik M, Kuppalli K, Kindrachuk J, Peiris M. Virology, Transmission, and Pathogenesis of Sars-Cov-2. *BMJ* (2020) 371:m3862. doi:10.1136/bmj.m3862

15. Meyers KJ, Jones ME, Goetz IA, Botros FT, Knorr J, Manner DH, et al. A Cross-sectional Community-based Observational Study of Asymptomatic SARS-CoV-2 Prevalence in the Greater Indianapolis Area. *J Med Virol* (2020) 92:2874–9. doi:10.1002/jmv.26182

16. Chen J. Pathogenicity and Transmissibility of 2019-nCoV-A Quick Overview and Comparison with Other Emerging Viruses. *Microbes Infect* (2020) 22:69–71. doi:10.1016/j.micinf.2020.01.004

17. Wang Y, Wang Y, Chen Y, Qin Q. Unique Epidemiological and Clinical Features of the Emerging 2019 Novel Coronavirus Pneumonia (COVID-19) Implicate Special Control Measures. *J Med Virol* (2020) 92:568–76. doi:10.1002/jmv.25748

18. Djordjevic M, Salom I, Markovic S, Rodic A, Milicevic O, Djordjevic M. Inferring the Main Drivers of Sars-Cov-2 Global Transmissibility by Feature Selection Methods. *GeoHealth* (2021) 5:e2021GH000432. doi:10.1029/2021GH000432

19. Bavel JJV, Baicker K, Boggio PS, Capraro V, Cichocka A, Cikara M, et al. Using Social and Behavioural Science to Support Covid-19 Pandemic Response. *Nat Hum Behav* (2020) 4:460–71. doi:10.1038/s41562-020-0884-z

20. Chaudhry R, Dranitsaris G, Mubashir T, Bartoszko J, Riazi S. A Country Level Analysis Measuring the Impact of Government Actions, Country Preparedness and Socioeconomic Factors on Covid-19 Mortality and Related Health Outcomes. *EClinicalMedicine* (2020) 25:100464. doi:10.1016/j.eclinm.2020.100464

21. Funk CD, Laferrière C, Ardakani A. A Snapshot of the Global Race for Vaccines Targeting Sars-Cov-2 and the Covid-19 Pandemic. *Front Pharmacol* (2020) 11:937. doi:10.3389/fphar.2020.00937

22. Wang Z, Bauch CT, Bhattacharyya S, d'Onofrio A, Manfredi P, Perc M, et al. Statistical Physics of Vaccination. *Phys Rep* (2016) 664:1–113. doi:10.1016/j.physrep.2016.10.006

23. Khajanchi S, Das DK, Kar TK. Dynamics of Tuberculosis Transmission with Exogenous Reinfections and Endogenous Reactivation. *Physica A: Stat Mech its Appl* (2018) 497:52–71. doi:10.1016/j.physa.2018.01.014

24. Khajanchi S, Perc M, Ghosh D. The Influence of Time Delay in a Chaotic Cancer Model. *Chaos* (2018) 28:103101. doi:10.1063/1.5052496

25. Haug N, Geyrhofer L, Londei A, Dervic E, Desvars-Larrive A, Loreto V, et al. Ranking the Effectiveness of Worldwide Covid-19 Government Interventions. *Nat Hum Behav* (2020) 4:1303–12. doi:10.1038/s41562-020-01009-0

26. Weitz JS, Park SW, Eksin C, Dushoff J. Awareness-driven Behavior Changes Can Shift the Shape of Epidemics Away from Peaks and toward Plateaus, Shoulders, and Oscillations. *Proc Natl Acad Sci U.S.A* (2020) 117:32764–71. doi:10.1073/pnas.2009911117

27. Tkachenko AV, Maslov S, Elbanna A, Wong GN, Weiner ZJ, Goldenfeld N. Time-dependent Heterogeneity Leads to Transient Suppression of the Covid-19 Epidemic, Not Herd Immunity. *Proc Natl Acad Sci* (2021) 118. doi:10.1073/pnas.2015972118

28. Brauer F. Mathematical Epidemiology: Past, Present, and Future. *Infect Dis Model* (2017) 2:113–27. doi:10.1016/j.idm.2017.02.001

29. Tadić B, Melnik R. Microscopic Dynamics Modeling Unravels the Role of Asymptomatic Virus Carriers in Sars-Cov-2 Epidemics at the Interplay between Biological and Social Factors. *Comput Biol Med* (2021) 133:104422. doi:10.1016/j.compbiomed.2021.104422

30. Nagel K, Rakow C, Müller SA. Realistic Agent-Based Simulation of Infection Dynamics and Percolation. *Physica A: Stat Mech its Appl* (2021) 584:126322. doi:10.1016/j.physa.2021.126322

31. Burda Z. Modelling Excess Mortality in Covid-19-like Epidemics. *Entropy* (2020) 22:1236. doi:10.3390/e22111236

32. Chang SL, Harding N, Zachreson C, Cliff OM, Prokopenko M. Modelling Transmission and Control of the Covid-19 Pandemic in australia. *Nat Commun* (2020) 11:5710–3. doi:10.1038/s41467-020-19393-6

33. Jackson ML. Low-impact Social Distancing Interventions to Mitigate Local Epidemics of Sars-Cov-2. *Microbes Infect* (2020) 22:611–6. doi:10.1016/j.micinf.2020.09.006

34. Lin Q, Zhao S, Gao D, Lou Y, Yang S, Musa SS, et al. A Conceptual Model for the Coronavirus Disease 2019 (Covid-19) Outbreak in Wuhan, china with Individual Reaction and Governmental Action. *Int J Infect Dis* (2020) 93:211–6. doi:10.1016/j.ijid.2020.02.058

35. Magal P, Webb G. Predicting the Number of Reported and Unreported Cases for the Covid-19 Epidemic in south korea, italy, france and germany (2020). Italy, France and Germany *(March 19, 2020)*.

36. Hoertel N, Blachier M, Blanco C, Olfson M, Massetti M, Rico MS, et al. A Stochastic Agent-Based Model of the Sars-Cov-2 Epidemic in france. *Nat Med* (2020) 26:1417–21. doi:10.1038/s41591-020-1001-6

37. Rice BL, Annapragada A, Baker RE, Bruijning M, Dotse-Gborgbortsi W, Mensah K, et al. Variation in Sars-Cov-2 Outbreaks across Sub-saharan Africa. *Nat Med* (2021) 27:447–53. doi:10.1038/s41591-021-01234-8

38. Giordano G, Blanchini F, Bruno R, Colaneri P, Di Filippo A, Di Matteo A, et al. Modelling the Covid-19 Epidemic and Implementation of Population-wide Interventions in italy. *Nat Med* (2020) 26:855–60. doi:10.1038/s41591-020-0883-7

39. Anastassopoulou C, Russo L, Tsakris A, Siettos C. Data-based Analysis, Modelling and Forecasting of the Covid-19 Outbreak. *PloS one* (2020) 15: e0230405. doi:10.1371/journal.pone.0230405

40. Christopoulos DT. A Novel Approach for Estimating the Final Outcome of Global Diseases like Covid-19. *medRxiv* (2020). doi:10.1101/2020.07.03. 20145672

41. Thurner S, Klimek P, Hanel R. A Network-Based Explanation of Why Most Covid-19 Infection Curves Are Linear. *Proc Natl Acad Sci U.S.A* (2020) 117: 22684–9. doi:10.1073/pnas.2010398117

42. Vasconcelos GL, Macêdo AMS, Duarte-Filho GC, Brum AA, Ospina R, Almeida FAG. Power Law Behaviour in the Saturation Regime of Fatality Curves of the Covid-19 Pandemic. *Sci Rep* (2021) 11:4619–2. doi:10.1038/s41598-021-84165-1

43. Tkachenko AV, Maslov S, Wang T, Elbana A, Wong GN, Goldenfeld N. Stochastic Social Behavior Coupled to Covid-19 Dynamics Leads to Waves, Plateaus, and an Endemic State. *Elife* (2021) 10:e68341. doi:10.7554/eLife. 68341

44. [Dataset] CSSE. Covid-19 Data Repository by the center for Systems Science and Engineering (Csse) at the Johns hopkins university (2020).

45. Plerou V, Gopikrishnan P, Rosenow B, Amaral LAN, Stanley HE. Econophysics: Financial Time Series from a Statistical Physics point of View. *Physica A: Stat Mech its Appl* (2000) 279:443–56. doi:10.1016/s0378-4371(00)00010-8

46. Maslov S. Measures of Globalization Based on Cross-Correlations of World Financial Indices. *Physica A: Stat Mech its Appl* (2001) 301:397–406. doi:10. 1016/s0378-4371(01)00370-3

47. Tadić B, Mitrović M. Jamming and Correlation Patterns in Traffic of Information on Sparse Modular Networks. *The Eur Phys J B* (2009) 71:631–40.

48. Isufaj R, Koca T, Piera MA. Spatiotemporal Graph Indicators for Air Traffic Complexity Analysis. *Aerospace* (2021) 8:364. doi:10.3390/aerospace8120364

49. Baruchi I, Ben-Jacob E. Functional Holography of Recorded Neuronal Networks Activity. *Ni* (2004) 2:333–52. doi:10.1385/ni:2:3:333

50. Tadić B, Andjelković M, Boshkoska BM, Levnajić Z. Algebraic Topology of Multi-Brain Connectivity Networks Reveals Dissimilarity in Functional Patterns during Spoken Communications. *PLoS One* (2016) 11:e0166787. doi:10.1371/journal.pone.0166787

51. Madi A, Friedman Y, Roth D, Regev T, Bransburg-Zabary S, Jacob EB. Genome Holography: Deciphering Function-form Motifs from Gene Expression Data. *PLoS One* (2008) 3:e2708. doi:10.1371/journal.pone. 0002708

52. Živković J, Tadić B, Wick N, Thurner S. Statistical Indicators of Collective Behavior and Functional Clusters in Gene Networks of Yeast. *Eur Phys J B-Condensed Matter Complex Syst* (2006) 50:255–8.

53. Lahiri D, Dubey S, Ardila A. Impact of Covid-19 Related Lockdown on Cognition and Emotion: A Pilot Study. *medRxiv* (2020). 2020.06.30.20138446. doi:10.1101/2020.06.30.20138446

54. Browning R, Sulem D, Mengersen K, Rivoirard V, Rousseau J. Simple Discrete-Time Self-Exciting Models Can Describe Complex Dynamic Processes: A Case Study of Covid-19. *PloS one* (2021) 16:e0250015. doi:10.1371/journal.pone. 0250015

55. Biyikoglu T, Leydold J, Stadler PF. *Laplacian Eigenvectors of Graphs: Perron-Frobenius and Faber-Krahn Type Theorems*. Heidelberg: Springer (2007).

56. Mitrović M, Tadić B. Spectral and Dynamical Properties in Classes of Sparse Networks with Mesoscopic Inhomogeneities. *Phys Rev E* (2009) 80: 026123.

57. Bastian M, Heymann S, Jacomy M. Gephi: an Open Source Software for Exploring and Manipulating Networks. *Proc Int AAAI Conf web Soc media* (2009) 3:361–2.

58. Tavenard R, Faouzi J, Vandewiele G, Divo F, Androz G, Holtz C, et al. Tslearn, a Machine Learning Toolkit for Time Series Data. *J Mach Learn Res* (2020) 21:1–6.

59. Hu J, Gao J, Wang X. Multifractal Analysis of sunspot Time Series: the Effects of the 11-year Cycle and Fourier Truncation. *J Stat Mech* (2009) 2009:P02066. doi:10.1088/1742-5468/2009/02/p02066

60. Šuvakov M, Mitrovic M, Gligorijevic V, Tadic B. How the Online Social Networks Are Used: Dialogues-Based Structure of Myspace. *J R Soc Interf* (2013) 10:20120819. doi:10.1098/rsif.2012.0819

(19) **REPUBLIKA SRBIJA** (12) **Spis malog patenta** (11) **1566 U1**

(73) Nosilac malog patenta:
     INSTITUT ZA FIZIKU ZEMUN,
     Pregrevica 118, 11080 Beograd-Zemun, RS

(72) Pronalazači:
     VUKOVIĆ, Đorđe;
     LAZOVIĆ, Saša;
     DIMITRIJEVIĆ, Dragoljub;
     MITROVIĆ-DANKULOV, Marija;
     JOVANOVIĆ, Stevan;
     ĐUKIĆ, VUKOVIĆ, Aleksandra

(74) Zastupnik:

(54) Naziv: **PODESIVI MEHANIZAM LABORATORIJSKE MEŠALICE**

(57) Apstrakt:

Podesivi mehanizam laboratorijske mešalice: Rotirajuća ruka 3 pogonjena vratilom elektromotora 11 prenosi orbitalno kretanje na laboratorijski uzorak preko osovinice 4 čija se radijalno odstojanje od a time i amplituda orbitalnog kretanja, podešava putem vretena 2. Na ovaj način je omogućeno kontinualno podešavanje amplitude orbitalnog kretanja laboratorijskog uzorka na mešalici.

RS 1566 U1
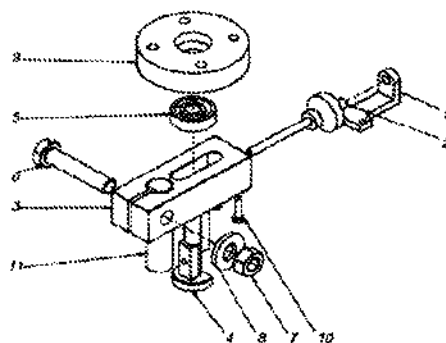
Област технике на коју се проналазак односи

Проналазак се односи на унапређење постојећих лабораторијских мешалица које се користе за припрему различитих фармацеутских, биолошких и хемијских агенаса и производа, лекова, боја итд.

Према Међународној класификацији патената (МКП) ознака је: **F 16 H 51/02.**

Технички проблем

Проналазак решава проблем конструкције механизма за подешавање амплитуде орбиталног кретања тиме што омогућава линеарно подешавање центрифугалне силе која делује на узорак постављен на лабораторијску мешалицу.

Стање технике

Лабораторијске мешалице се користе за временски и брзински подесиво мешање у хоризонталној равни различитих узорака. Комерцијалне мешалице, које се могу наћи на тржишту, производе се са механизмом са непроменљивом амплитудом, тј. радијусом ексцентричног кретања од ½ инча (12,7 mm) или 1 инча (25,4 mm). Код њих се регулација центрифугалне силе врши само променом броја обртаја погона, што за последицу има да се центрифугална сила мења квадратно (мање прецизно), изузев патента бр. GB2277043A где је то изводљиво променом позиције узорка.

Главна предност овог патента је у подесивом механизму који омогућава да се амплитуда мења континуално од 12,7 mm до 25,4 mm. Наводимо примере патената који описују лабораторијске мешалице сличних карактеристика.

Патент GB1601590A описује мешалицу, код које се брзина обртања и амплитуда подешавају променом дужине одговарајућих полуга. Промена једног параметра условљава промену другог. Уређај не омогућава континуалну промену центрифугалне силе која делује на узорак.

2

Патент **GB2277043A** описује мешалицу са ротором електромотора који је истовремено и носећа плоча узорка који се меша. Орбитално кретање узроковано је променом магнетног поља статора. Сви параметри зависе од оптерећења мотора и позиције узорка. Амплитуда мешања зависи од позиције узорка.

Патент **GB2310147A** орбитално кретање носеће плоче разложено је на два међусобно управна линеарна кретања тако да из кружног може прећи на праволинијско кретање. Амплитуду није могуће подешавати. Учестаност се мења са брзином обртања мотора.

Патент **US8226291B2** амплитуда мешања се подешава у циљу елиминације дејства инерцијалних сила на ротирајуће склопове. Параметри су оптимизовани за продужавање радног века уређаја а не за ефективност мешања. Амплитуда зависи од масе узорка.

Патент **US2015098300A1** описује техничко решење са кретањем помоћу зупчастог преносника. Зупчаници индукују фине вибрације које поспешују мешање. Могуће је подешавање учестаности, али не и амплитуде мешања.

Патент **WO2016020176A1** описује мешалицу са аутоматским динамичким уравнотежавањем, чиме се елиминише дејство инерцијалних сила на ротирајуће делове, што омогућава смањење димензија уређаја. Промена параметара мешања није обухваћена овим патентом.

Патент **WO2016075214A1** описује веома сложени механизам са низом механичких преносника којима се прецизно подешава учестаност мешања узорка. Амплитуда је константна. Зависност центрифугалне силе од мпараметра који се подешава (учестаности) је квадратна функција.

Излагање суштине проналаска

Подесиви механизам лабораторијске мешалице је састављен од склопа руке монтираног на вратило *погонског електромотора 11*, који у себи садржи *осовиницу 4* која се помоћу *вретена 2* може померати дуж *руке 3*. Окретањем *вретена 2* подешава се размак између вратила електромотора и *осовинице 4* која преко *кугличног лежаја 5* у *кућишту 9* помера радну плочу на којој су лабораторијски узорци. Предност овог

проналаска је у могућности подешавања размака између вратила мотора и осовинице тј. амплитуде орбиталног кретања које се изводи директним окретањем вретена потпуно независно од осталих параметара уређаја.

Кратак опис слика нацрта

Проналазак је детаљно описан и приказан на нацрту у коме:

Слика 1 – представља експандирани приказ ексцентричног механизма орбиталне мешалице на коме је представљен проналазак.

Детаљан опис проналаска

Постоји више врста лабораторијских мешалица према кинематици кретања којим врше мешање. Овде приказан подесиви механизам се односи на мешалице са радном плочом која се креће у хоризонталној равни, тј. паралелно основи као што је наведено у примерима који су изнети у делу који се односи на стање технике.

На радну плочу се постављају посуде са агенсима, а то могу бити чаше, ерленмајери, епрувете, итд. које помоћу прилагодних носача леже или су причвршћене за радну плочу.

Захтеви које диктирају материје и њихова обрада, нарочито биолошки агенси које треба мешати, су такви да се мешање врши у хоризонталној равни, одређеном учестаношћу, са одређеном амплитудом, да центрифугалне силе остану у жељеним оквирима како не би разориле ћелијске зидове, мембране, раслојиле материје различитих густина, као и да се формира оптимална контактна површина материје са атмосфером, итд.

У наведеним примерима сви патентирани уређаји имају радну плочу која се креће у хоризонталној равни помоћу механизма заснованог на коленастом вратилу, линеарном механизму, зупчаницима итд. али ни један од наведених не даје могућност промене амплитуде мешања.

Такође, ексцентрични механизам је много једноставнији и поузданији од свих горе наведених примера из стања технике. Механизам се састоји из десет позиција и повезан је са *вратилом погонског електромотора 11* које се обрће око своје осе. Преко *руке 3* која је круто повезана са *вратилом електромотора 11* и за њега учвршћена

преклопном везом – *вијком 6* са *подлошком 8* и *навртком 7,* кретање се преноси до *осовинице 4.* Позиција *осовиница 4* је обликована тако да својим доњим делом лежи у жљебу који се простире дуж *руке 3.* Облик доњег дела *осовинице 4* је такав да не може да се обрће око своје осе, нити да испадне, али може да клиза дуж жљеба. На *осовиници 4* постоји попречни навојни отвор кроз који пролази навојно *вретено 2* за подешавање. *Вретено 2* је на свом супротном крају, који се налази изван *руке 3* снабдевено точкићем који се може окретати руком и ослоњено је на *држач вијка 1. Држач 1* је причвршћен за *руку 3 помоћним вијцима 10.* Окретањем *вретена 2, осовиница 4* се може померати дуж жљеба унутар *руке 3* чиме се мења удаљеност *осовинице 4* од осе обртања *вратила електромотора 11,* и на тај начин се мења амплитуда кружног кретања. Елементи који чине овај склоп су тако димензионисани, да се полупречник, односно амплитуда обртања, може мењати континуално у границама од ½ инча (12,7 mm) до целог инча (25,4 mm) чиме су покривене обе опције амплитуда (12,7 mm и 25,4 mm) лабораторијских мешалица које се тренутно нуде на тржишту. Поред тога се амплитуда мешања, односно центрифугална сила може подешавати за све вредности које се налазе између ове две опције.

На *осовиницу 4* је монтиран стандардни *куглични лежај 5.* Лежај је смештен у *кућиште 9* које је причвршћено за центар радне плоче, са њене доње стране. Како је радна плоча везана за клизаче линеарних вођица које су смештене на телу мешалице, у хоризонталној равни, радна плоча може да се креће орбитално, само у хоризонталној равни. *Лежај 5,* не преноси обртни момент, већ само силу са *осовинице 4* која преко *кућишта 9* покреће радну плочу мешалице и гони је на орбитално кретање.

Вретеном 2 се једноставно мења полупречник, односно амплитуда орбиталног кретања тј. мешања. Потребно је само одвити или завити точкић на *вретену 2.* Број осцилација односно учестаност мешања регулише се бројем обртаја мотора, што није предмет овог малог патента.

Пре пуштања мешалице у рад, на горе описани начин подеси се полупречник, односно амплитуда мешања која остаје стална током рада мешалице. Мешалица са оваквим механизмом замењује мешалице са амплитудом 12,7 mm и 25,4 mm као и све међувеличине.

Патентни захтев

1. Подесиви механизам лабораторијске мешалице, **назначен тиме**, што се састоји из *руке (3)* која је монтирана на вратило погонског *електромотора (11)*, *што је на руци (3) изведен отвор кроз који је провучена осовиница (4) која је* преко *кугличног лежаја (6)* и *кућишта (9) повезана* на радну плочу мешалице, и *што је на руци (3) изведен отвор кроз који пролази вретено (2) за подешавање* и позиционирање *осовинице (4).*
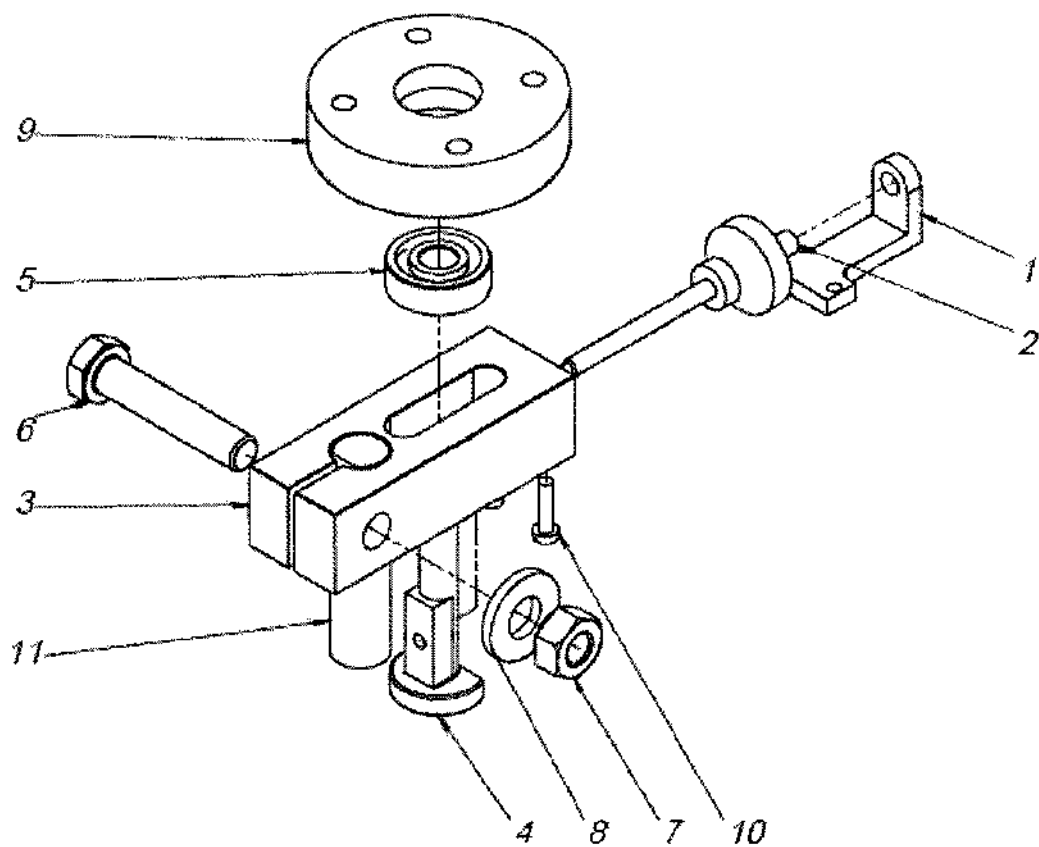
Слика 1

Информативни подаци о малом патенту/пријави малог патента

Регистарски број (Registration number)	1566
Број и датум решења о признању права (Number and date of decision to grant the right)	2018/15457 31.10.2018
Број пријаве (Application number)	МП-2017/0074
Датум пријема пријаве (Reception date)	29.12.2017
Признати датум подношења пријаве (Filing date)	29.12.2017
Статус (Legal status)	Регистрован (Registered)
Датум објављивања и број службеног гласила признатог права (B1) (Publication date and number of gazette of granted right)	(U1) 30.11.2018. 11/2018
Међународна класификација патената (IPC)	F16H 51/02; B01F 15/00
Назив проналаска (Title of invention)	PODESIVI MEHANIZAM LABORATORIJSKE MEŠALICE ADJUSTABLE MECHANISM FOR LABORATORY SHAKER
Подаци о проналазачу (Inventor)	VUKOVIĆ, Đorđe LAZOVIĆ, Saša DIMITRIJEVIĆ, Dragoljub MITROVIĆ-DANKULOV, Marija JOVANOVIĆ, Stevan ĐUKIĆ, VUKOVIĆ, Aleksandra
Подаци о носиоцу права (Owner)	INSTITUT ZA FIZIKU ZEMUN, Pregrevica 118, 11080 Beograd-Zemun, RS