

ИНСТИТУТ ЗА ФИЗИКУ			
ПРИМЛЕНО: 15-03-2018			
Рад.јед.	б р о ј	Арх.шифра	Прилог
ФФ01	351/1		

UNIVERSITY OF BELGRADE
INSTITUTE OF PHYSICS BELGRADE

Pregrevica 118, 11080 Zemun – Belgrade, Serbia
Tel: +381 11 3713000, Fax: +381 11 3162190, www.ipb.ac.rs



Научном већу Института за физику

Предлог за годишњу Студентску награду

Института за физику

Поштовани,

Велико ми је задовољство да предложим др **Јелену Смиљанић** за доделу Студентске награде за најбољу докторску тезу урађену на Институту за физику у Београду. Др Смиљанић је тезу под насловом “Испитивање својстава комплексних мрежа са дискретном динамиком”, одбранила 16.11.2017. на Електротехничком факултету Универзитета у Београду.

Докторска дисертација др Јелене Смиљанић је у области статистичке физике и бави се проучавањем структуре и динамике социјалних група са дискретном динамиком, односно динамиком која је условљена догађајима који су добро локализовани у простору и времену. Дисертација је базирана на два рада [A1, A2] објављена у часопису PLOS ONE и поглављу у књизи [A3]. На радовима [A1, A2] др Смиљанић је први аутор, док је на раду [A3] други аутор. Рад [A1] је до сада цитирани 4 пута без аутоцитата према бази Google Scholar, укључујући и један цитат у часопису *Nature Cell Death Discovery* као пример истраживања чији резултати могу допринети постизању боље кохезије унутар заједнице истраживача. У осталим радовима који цитирају рад [A1] исти је препознат као једино истраживање које се бави динамиком учешћа научника на конференцијама као и како социјални контакти утичу на кохезију научних заједница организованых око једне конференције. Поред радова приказаних у тези др Смиљанић је радила и на више других, врло различитих тема. Резултате у области квантно каскадних ласера [B1] објављени су у часопису *Computer Physics Communications*, док су резултати у области транспорта на телекомуникационим мрежама [B2] објављен у часопису *Physica A: Statistical Mechanics and its Applications*. Ова два рада су према бази Web of Science до сада цитиран 6 пута без аутоцитата.

Дисертација др Јелене Смиљанић се односи на анализу динамичких процеса, структуре мреже интеракција, као и њиховог међусобног утицаја у социјалним системима који подразумевају непосредну, личну комуникацију. У истраживању су коришћени подаци са научних конференција [A1] и Meetup портала. Динамика посматраних социјалних група је дискретна, пошто чланови учествују у колективним активностима које се одржавају на тачно одређеним локацијама у тачно одређеним временским тренуцима.



Циљ докторске дисертације, и радова [A1, A2, A3] на којима је она базирана, био је да се открију и објасне механизми који утичу на динамику учествовања у колективним активностима. У ту сврху коришћена су два приступа: емпиријска анализа и теоријско моделовање. Емпиријска анализа укључивала је анализу образаца учешћа појединца у активностима групе (научника на научним конференцијама и чланова Meetup социјалних група на догађајима који су у оквиру њих организовани), као и анализу структуре и еволуције социјалних мрежа у овим системима. Емпиријска анализа образаца учешћа је показала да су исти изузетно хетерогени, и да се не може користити једноставан модел где појединци насумично бирају догађаје на којима ће учествовати, већ да одлука о учествовању на догађају зависи од претходне активности. Да би се додатно испитао овај механизам учествовања дизајниран је и имплементиран математички модел, заснован на генерализованом бинарном Појином моделу, у коме вероватноћа учествовања зависи од односа броја претходних учествовања и пауза. Оптимални параметри модела одређени су применом нумеричких симулација тако да Кулбак-Лејблерово растојање између емпиријске расподеле и расподеле која се добија применом модела буде минимално. Закључено је да у социјалним групама, чија је динамика дискретна и условљена колективним активностима на којима учествују чланови групе, постоје универзални обрасци понашања који не зависе од сврхе окупљања, величине групе, као ни од учестаности организовања догађаја. Са друге стране, социјална компонента има важан утицај на динамику учествовања појединаца у активностима групе.

Структура социјалне мреже у посматраним системима еволуира кроз колективне догађаје и њеном анализом може се објаснити међусобна зависност између структуре социјалних интеракција и динамике система. Полазећи од бипартитне мреже, у којој дисјунктни подскупови чворова описују чланове социјалне групе и колективне догађаје, одабран је погодан метод за пројекцију на отежињену монопартитну мрежу, где се посматрају само релевантне везе између чланова који су заједно учествовали на догађајима, док су везе између чланова које су последица случајних интеракција занемарене. Утврђено је да са порастом броја догађаја на којима је члан групе учествовао, интеракције са члановима са којима је посматрани појединац повезан имају предност у односу на формирање нових веза у мрежи. Додатно је показано да понављање заједничког учествовања није на нивоу два појединца, већ је карактеристично формирање мањих подзаједница.

Допринос др Јелене Смиљанић у радовима обухваћеним докторском тезом је веома велики и врло разнолик. Колегиница Смиљанић је развила скрипте и методе за сакупљање, пречишћавање и формирање података. У оквиру ове активности треба посебно истаћи алгоритам за дисамбигуацију научника у подацима о научним



конференцијама који је посебно за ову прилику развила. Истраживање у области статистичке физике социјалних система често захтева и дефинисање величина којима се најбоље може описати понашање система. Др Смиљанић је показала да је у стању да одреди ове величине и детаљно изанализира динамику и структуру социјалних група са дискретном динамиком, поштујући притом врло ригорозне статистичке стандарде. Кроз теоријско моделирање социјалних процеса показала је да поседује изузетно знање у области аналитичког и нумеричког решавања ових модела. Све овде наведено указује на то да др Смиљанић поседује изузетне програмерске вештине. Др Смиљанић је такође дала и велики допринос и у писању и атрактивној презентацији резултата у објављеним радовима.

Овде треба истаћи и да је др Смиљанић сплетом околности на половини својих докторских студија променила област истраживања и да своју докторску тезу успела да уради у периоду од три године. У оквиру овог периода она је успешно ушла у нову област, овладала вештинама и знањима потребним за истраживачки рад у овој области, и дала већ сада значајан допринос овој области физике комплексних система. Ово је значајан успех, имајући у виду да се област статистичке физике социјалних система и теорија комплексних мрежа не изучавају на додипломским и мастер студијама ни на једном факултету у Републици Србији.

Докторска дисертација др Јелене Смиљанић представља значајан допринос статистичкој физици социјалних система као и развоју ове области на Институту за физику, и Републици Србији уопште. Стога, имајући у виду све наведено, са задовољством предлажем др Јелену Смиљанић за Студентску награду Института за физику у Београду.

У Београду, 15.03.2018.


др Марија Митровић Данкулов

научни сарадник
Институт за физику Београд

Биографија др Јелене Смиљанић

Јелена Смиљанић је рођена 12. 4. 1987. године у Лазаревцу, где је 2006. године завршила Гимназију као носилац Вукове дипломе. Исте године уписала је основне студије на Електротехничком факултету Универзитета у Београду. Дипломирала је на смеру Телекомуникације и информационе технологије октобра 2010. године са просечном оценом 9.04. Дипломски рад под називом “Сигурност у системима јавних кључева” урадила је под руководством проф. др Предрага Иваниша. Мастер студије је завршила у септембру 2011. године на истом факултету, смер Системско инжењерство и радио комуникације, са просечном оценом 9.83, одбравивши мастер рад на тему “Примена криптоанализе за испитивање сигурности система са јавним кључевима”, такође под руководством проф. др Предрага Иваниша. Јелена Смиљанић је 2011. године уписала докторске студије на смеру Наноелектроника и фотоника на Електротехничком факултету Универзитета у Београду. Докторску дисертацију под називом “Испитивање својстава комплексних мрежа са дискретном динамиком”, урађену под руководством др Марије Митровић Данкулов, одбранила је 16. 11. 2017. на Електротехничком факултету Универзитета у Београду.



Од децембра 2011. године до новембра 2012. године је као стипендиста Министарства просвете, науке и технолошког развоја Републике Србије ангажована у Лабораторији за примену рачунара у науци, у оквиру Националног центра изузетних вредности за изучавање комплексних система Института за физику у Београду, на пројекту основних истраживања ОН171017 “Моделирање и нумеричке симулације сложених вишечестичних система” чији је руководилац др Антун Балаж. У новембру 2012. године Јелена Смиљанић је запослена на Институту за физику у Београду као истраживач приправник на овом пројекту, а руководилац њених докторских студија је била др Марија Митровић Данкулов. Кандидат је звање истраживач сарадник стекла у мају 2013. године.

Јелена Смиљанић је аутор четири рада објављена у међународним часописима, од којих је један рад објављен у међународном часопису изузетних вредности категорије M21a, два рада у врхунском међународном часопису категорије M21 и један рад у истакнутом међународном часопису категорије M22. Такође је аутор и једног поглавља у тематском зборнику радова категорије M13. До сада је учествовала и презентовала своје резултате на више међународних и националних конференција.

Листа публикација др Јелене Смиљанић

Радови укључени у докторску дисертацију

[A1] **J. Smiljanić**, A. Chatterjee, T. Kauppinen, and M. Mitrović Dankulov, *A Theoretical Model for the Associative Nature of Conference Participation*, PLoS ONE 11, e0148528 (2016). (Категорија M21, ИФ=3.234)

[A2] **J. Smiljanić** and M. Mitrović Dankulov, *Associative nature of event participation dynamics: A network theory approach*, PLoS ONE 12, e0171565 (2017). (Категорија M21, ИФ=3.057)

[A3] **M. Mitrović Dankulov** and J. Smiljanić, *The Structure and Dynamics of Meetup Social Networks*, Scientific Computing: Studies and Applications, edited by Caj Erling, Chapter 2, Pages 33-61 (Nova Science Publishers, New York, 2017) ISBN: 978-1-53612-564-1. (Категорија M13)

Радови који нису укључени у докторску дисертацију

[B1] **J. Smiljanić**, M. Žeželj, J. Radovanović, V. Milanović and I. Stanković, *MATLAB-based Program for Optimization of Quantum Cascade Laser Active Region Parameters and Calculation of Output Characteristics in Magnetic Field*, Comput. Phys. Commun. **185**, 998 (2014). (Категорија M21a, ИФ=2.224)

[B2] **J. Smiljanić** and I. Stanković, *Efficient Routing on Small Complex Networks Without Buffers*, Physica A 392, 2294 (2013). (Категорија M22, ИФ=1.722)

Саопштења на интернационалин и националним скуповима

[C1] I. Stanković, M. Žeželj, **J. Smiljanić** and A. Belić, *Modelling of Disaster Spreading Dynamics High-Performance Computing Infrastructure for South East Europe's Research Communities*, Results of the HP-SEE User Forum 2012 Springer Book Series on Modeling and Optimization in Science and Technologies, 2, 31 (2014). (Категорија M33)

[C2] **J. Smiljanić**, M. Žeželj and I. Stanković, *Study of Routing Strategies in the Small Complex Networks*, Telekomunikacije **10**, 54 (2012). (Категорија M53)

[C3] M. Mitrović Dankulov and **J. Smiljanić**, *Associative nature of event-driven social dynamics: a network theory approach*, Complex Networks, 29 November-01 December 2017, Lyon, France. (Категорија M34)

[C4] **J. Smiljanić** and M. Mitrović Dankulov, *Associative nature of conference participation*, MECO, 14-17 February 2016, Vienna, Austria. (Категорија M34)

[C5] **J. Smiljanić** and M. Mitrović Dankulov, *Conference Attendance Patterns*, SFKM, 7-11 September 2015, Belgrade, Serbia. (Категорија M34)

[C6] I. Stanković and **J. Smiljanić**, *Structure of complex networks for minimizing traffic congestion and cost*, DPG, 25-30 March 2012, Berlin, Germany. (Категорија M34)

[C7] M. Mitrović Dankulov and **J. Smiljanić**, Structure and Dynamics of Event-driven Social Groups TINKOS, 9-10 November 2017, Belgrade, Serbia. (Категорија М64)

[C8] **J. Smiljanić** and M. Mitrović Dankulov, Associative nature of conference participation dynamics - an empirical analysis and modeling, TINKOS, 27-28 October 2016, Belgrade, Serbia. (Категорија М64)

[C9] **J. Smiljanić** and P. Ivaniš, Attacks on the RSA cryptosystem using integer factorization, TELFOR, 22-24 November 2011, Belgrade, Serbia. (Категорија М64)

Цитираност

Према бази *Web of Science* радови др Јелене Смиљанић су цитирани укупно 9 пута, од чега 7 пута изузимајући аутоцитате. Према истој бази, Хиршов индекс кандидаткиње је 2. Укупан импакт фактор радова од [А1-А3] је 6.291, док је укупан импакт фактор свих објављених радова 11.125.

Web of Science



Search Search Results My Tools Search History Marked List 5

Citation report for 5 results from Web of Science Core Collection between 1996 and 2018 **Go**

You searched for: From Marked List: [...More](#)

This report reflects citations to source items indexed within Web of Science Core Collection. Perform a Cited Reference Search to include citations to items not indexed within Web of Science Core Collection.

Export Data: Save to Text File

Total Publications

5

1998 2017

h-index

2

Average citations per item

1.8

Sum of Times Cited

9

Without self citations

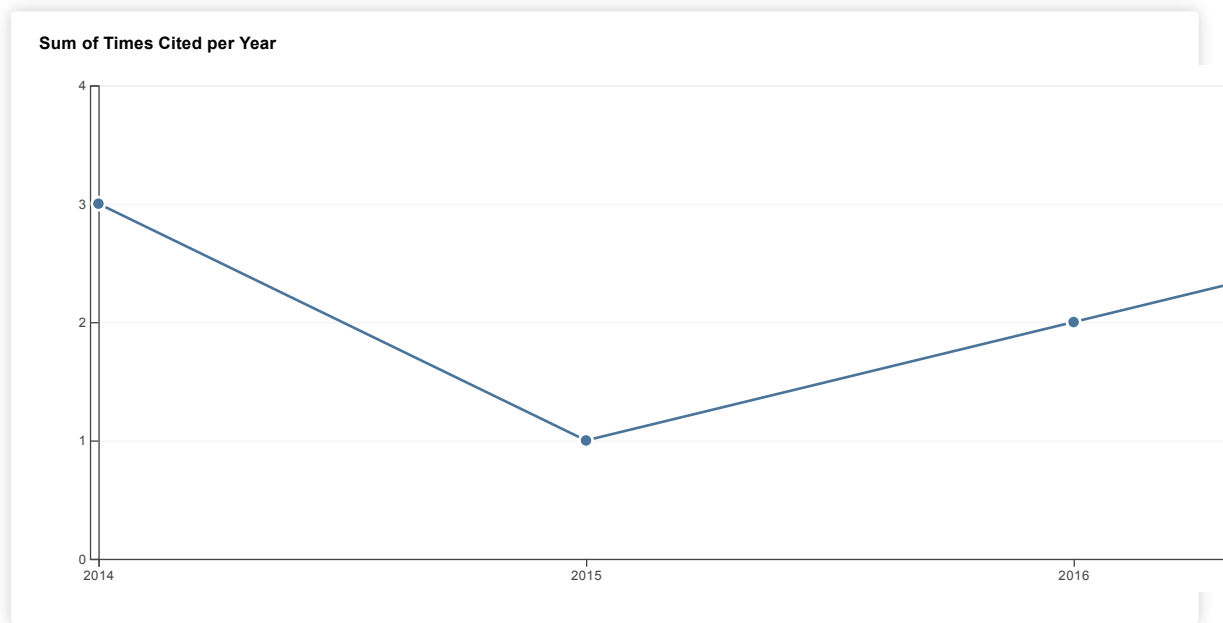
7

Citing articles

9

Without self citations

7



Sort by: Times Cited Date More

Page 1 of 1

	2014	2015	2016	2017	2018	Total	Average Citations per Year
<input type="checkbox"/> 1. MATLAB-based program for optimization of quantum cascade laser active region parameters and calculation of output characteristics in magnetic field	3	1	2	3	0	9	1.80
<input checked="" type="checkbox"/> 1. MATLAB-based program for optimization of quantum cascade laser active region parameters and calculation of output characteristics in magnetic field	2	0	2	0	0	4	0.80

By: Smiljanic, J.; Zezelj, M.; Milanovic, V.; et al.

COMPUTER PHYSICS COMMUNICATIONS Volume: 185 Issue: 3 Pages: 998-1006 Published: MAR 2014

- 2. **Efficient routing on small complex networks without buffers**
 By: Smiljanic, J.; Stankovic, I. 1 1 0 1 0 3 0.50
 PHYSICA A-STATISTICAL MECHANICS AND ITS APPLICATIONS Volume: 392 Issue: 9 Pages: 2294-2301 Published: MAY 1 2013
- 3. **A Theoretical Model for the Associative Nature of Conference Participation**
 By: Smiljanic, Jelena; Chatterjee, Arnab; Kauppinen, Tomi; et al. 0 0 0 2 0 2 0.67
 PLOS ONE Volume: 11 Issue: 2 Article Number: e0148528 Published: FEB 9 2016
- 4. **Associative nature of event participation dynamics: A network theory approach**
 By: Smiljanic, Jelena; Dankulov, Marija Mitrovic 0 0 0 0 0 0 0.00
 PLOS ONE Volume: 12 Issue: 2 Article Number: e0171565 Published: FEB 6 2017
- 5. **Modelling of Disaster Spreading Dynamics**
 By: Stankovic, Igor; Zezelj, Milan; Smiljanic, Jelena; et al. 0 0 0 0 0 0 0.00
 Edited by: Dulea, M; Karaivanova, A; Oulas, A; et al.
 Conference: HP-SEE User Forum Location: Natl Lib Serbia, Belgrade, SERBIA
 Date: OCT 17-19, 2012
 Sponsor(s): Inst Phys Belgrade
 HIGH-PERFORMANCE COMPUTING INFRASTRUCTURE FOR SOUTH EAST EUROPE'S RESEARCH COMMUNITIES: RESULTS OF THE HP-SEE USER FORUM 2012 Book Series: Modeling and Optimization in Science and Technologies Volume: 2 Pages: 31-42 Published: 2014

Select Page |   |

Sort by: Times Cited Date

Page of 1

5 records matched your query of the 40,229,813 in the data limits you selected.

УНИВЕРЗИТЕТ У БЕОГРАДУ
ЕЛЕКТРОТЕХНИЧКИ ФАКУЛТЕТ

Јелена М. Смиљанић

ИСПИТИВАЊЕ СВОЈСТАВА
КОМПЛЕКСНИХ МРЕЖА СА
ДИСКРЕТНОМ ДИНАМИКОМ

докторска дисертација

Београд, 2017

UNIVERSITY OF BELGRADE
SCHOOL OF ELECTRICAL ENGINEERING

Jelena M. Smiljanić

**ANALYSIS OF PROPERTIES OF
COMPLEX NETWORKS WITH
DISCRETE DYNAMICS**

Doctoral Dissertation

Belgrade, 2017

Ментор:

др Марија Митровић Данкулов, научни сарадник
Универзитет у Београду - Институт за физику у Београду

Чланови комисије:

др Јелена Радовановић, редовни професор
Универзитет у Београду - Електротехнички факултет

др Витомир Милановић, професор емеритус
Универзитет у Београду - Електротехнички факултет

др Марија Рашајски, ванредни професор
Универзитет у Београду - Електротехнички факултет

др Антун Балаж, научни саветник
Универзитет у Београду - Институт за физику у Београду

Датум одбране: 16. новембар 2017.

Захвалница

Ова докторска дисертација је урађена у Лабораторији за примену рачунара у науци Института за физику у Београду, под руководством др Марије Митровић Данкулов. Захваљујем се др Митровић Данкулов на указаном поверењу, помоћи око одабира теме истраживања, пренесеном знању, као и на стрпљењу које је уложила у овај рад.

Велику захвалност дугујем др Антуну Балажу на великом подстицају у научноистраживачком раду, константној подршци и корисним саветима.

Желела бих да се захвалим др Игору Станковићу, др Милану Жежељу, проф. др Јелени Радовановић, проф. др Јовану Радуновићу и проф. др Витомиру Милановићу за сарадњу, коауторство и помоћ код објављивања првих радова на почетку моје научноистраживачке каријере.

Колеге из Лабораторије за примену рачунара у науци биле су изузетна подршка и свакодневни заједнички рад био је право задовољство. Овом приликом им се пуно захваљујем.

Посебно се захваљујем својој породици и пријатељима на подршци током мог целокупног школовања.

Овај рад је финансиран у оквиру пројекта ОИ171017 Министарства просвете науке и технолошког развоја Републике Србије.

Испитивање својстава комплексних мрежа са дискретном динамиком

Сажетак

Комплексне мреже су се у току последње две деценије показале као изузетно користан концепт у проучавању карактеристика комплексних система. Потребно је да се прво издвоје појединачни елементи система и интеракције између њих представе помоћу одговарајуће мреже, а затим се проучавање организације система своди на анализу структуре и динамичких процеса на мрежи применом одговарајуће методологије.

Како је разноврсност реалних система за које су доступни подаци који омогућавају увид у структуру мреже интеракција све већа, потребно је непрекидно развијати нове технике и теоријске моделе који могу да објасне понашање специфичних система. У овој дисертацији изучаване су комплексне мреже са дискретном динамиком, на основу података везаних за социјалне системе који су организовани кроз колективне активности. Карактеристике мрежа које описују овакве системе недовољно су проучене. Један од главних разлога за то је доступност емпиријских података. У наведеним системима интеракције између појединаца су директне, због чега је знатно теже доћи до података него у случају социјалних система где појединци комуницирају посредством неког уређаја.

Посебна пажња посвећена је испитивању динамике учествовања појединаца у колективним активностима. На основу резултата статистичке анализе емпиријских података, показано је да учествовање појединаца у колективним активностима није насумично и разматрани су математички модели који могу да објасне активност појединаца на догађајима која је изразито хетерогена. Утврђено је да се применом генерализованог бинарног Појиног модела могу успешно репродуковати посматрани емпиријски резултати.

Користећи ансамбл бипартитних мрежа са задатим параметрима који има максималну ентропију, идентификоване су значајне везе у отежињеној мрежи

које одговарају релевантним социјалним интеракцијама. Како би се добио увид у еволуцију структуре мреже у времену, проучаване су промене локалних структурних параметара након сваког учешћа на догађају. Показало се да чланови система током првих појављивања на колективним догађајима образују нове везе са суседима, а затим, са повећањем броја учешћа на догађајима, интеракције са суседима и учвршћивање постојећих заједница постају доминантни у поређењу са формирањем нових веза у мрежи.

У циљу испитивања утицаја који појединачни догађаји имају на структуру мреже, овде је предложен приступ где се догађаји издвајају на неколико различитих начина и посматрају се промене до којих долази у структури мреже. Добијени резултати показују да су на догађајима са малим бројем учесника карактеристичне интеракције између чланова са јаким везама, за разлику од великих догађаја, где су углавном заступљене интеракције између појединаца са везама које се лако раскидају.

Кључне речи: комплексне мреже, социофизика, еволуција структуре мреже у времену, моделирање динамичких процеса на мрежи

Научна област: Електротехника и рачунарство

Област истраживања: Наноелектроника и фотоника

UDK број: 621.3

Analysis of properties of complex networks with discrete dynamics

Abstract

In the last two decades, complex networks have been proven as very useful concept for examination of properties of complex systems. The first step within this framework is to extract individual elements of the system and to represent interactions between these elements in the form of complex network. After this step, the study of complex system organization is reduced to the analysis of structure and dynamical processes on network with the use of suitable methodology.

The increase in a variety of real systems with available data, which enable insight into the structure of network of interactions, requires constant development of new techniques and theoretical models that could explain behavior of specific systems. In this thesis, we studied complex networks with discrete dynamics using data on event-based social systems. There has been very little previous research on properties of networks representing these systems. One of the main reasons is availability of data. In the given systems individuals interact face-to-face, wherefore it is more difficult to get the data, than in the case of social systems where individuals use some communication device to communicate with each other.

Special attention was paid to examination of activity of individuals in group events. According to the results of statistical analysis of empirical data it has been shown that individuals do not attend events randomly. We analysed mathematical models that can explain member's participation patterns on events which turned out to be strongly heterogeneous. It has been shown that generalized binary Pólya model can reproduce given empirical results successfully.

Using bipartite networks ensemble with maximum entropy, we identified significant connections in weighted network that represent relevant social interactions. In order to get the insight into evolution of the network structure, we analyzed change of local structural parameters after each event attendance. It has been shown that members of the system establish new connections with neighbors during member's

early involvement in the group activities, while later, as number of attended events increase, the interactions with neighbors and strengthening of existing communities become preferred in comparison to forming new connections in network.

In order to analyse the influence that particular event has on network structure, we proposed an approach based on event removal according to different criteria and examination of resulting structural changes in network. The results showed that interactions between individuals with strong connections are dominant on events with small number of members, while during the large events typically individuals with weak connections that could be easily broken interact.

Keywords: complex networks, sociophysics, evolution of network structure in time, modeling of dynamical processes on network

Scientific field: Electrical and Computer Engineering

Research area: Nanoelectronics and Photonics

UDC number: 621.3

Садржај

1	Увод	1
1.1	Физика комплексних система	1
1.2	Комплексне мреже	2
1.3	Социофизика	5
1.4	Предмет и циљ истраживања	7
1.5	Структура рада	8
2	Квантитативне методе	10
2.1	Статистичка анализа	10
2.2	Математички модели	15
2.2.1	Бернулијев процес	17
2.2.2	Хомоген Марковски ланац	18
2.2.3	Временски-нехомоген Марковски ланац	20
2.3	Комплексне мреже	22
2.3.1	Структура комплексне мреже	23
2.3.2	Бипартитна мрежа	28
2.3.3	Филтрирање отежињене мреже добијене пројекцијом бипартитне мреже	29
3	Подаци	33
3.1	Научне конференције	34
3.2	Meetup групе	39
4	Обрасци учешћа	45
4.1	Емпиријска анализа	45
4.1.1	Научне конференције	45

4.1.2	Meetup групе	50
4.2	Генерализовани бинарни Појин модел	54
4.2.1	Научне конференције	55
4.2.2	Meetup групе	60
4.3	Универзалне карактеристике динамике учествовања у колективним активностима	63
5	Структура социјалних мрежа под утицајем учешћа на догађајима	65
5.1	Анализа Meetup социјалних мрежа	66
5.2	Утицај појединачних догађаја на карактеристике социјалне мреже	73
6	Закључак	79
	Додатак А	84
A.1	Аналитичко решавање модела у случају Бернулијевог процеса . . .	84
A.2	Аналитичко решавање модела у случају хомогеног Марковског ланца	85
	Литература	88

1. Увод

1.1 Физика комплексних система

Комплексан систем је појам који се може применити на широку класу система за које је заједничко да се састоје од великог броја елемената између којих постоје комплексне интеракције. У оваквим системима је карактеристично самоорганизовано колективно понашање. То значи да у систему долази до спонтаног формирања структурног уређења и образаца понашања услед локалних интеракција, без претходног планирања и централне контроле. Издвајањем појединачних елемената и проучавањем њихових особина, не може се добити потпуни увид у карактеристике комплексног система. Са једне стране, интеракције доводе до промена стања појединачних елемената, услед чега се мења стање целог система, али истовремено, са променама стања се мењају и својства интеракција. Овакво понашање присутно је, између осталог, у различитим физичким, хемијским, социјалним, биолошким и економским системима [1–4].

Комплексни системи представљају веома актуелну област истраживања [5–7]. Услед сложене међусобне зависности стања појединачних елемената и њихових интеракција, проучавање особина већине комплексних система захтева примену рачунара високих перформанси. Рачунари се користе за решавање и симулације сложених математичких модела којима се описују ови системи. Приликом формулације и испитивања теоријског модела, могу се користити резултати експеримената над реалним системом, при чему се као исход експеримента сматрају и велики скупови података који су данас доступни у различитим електронским форматима. Теоријски модел не може детаљно обухватити сва својства неког реалног система, већ је циљ идентификовати и објаснити најбитније принципе колективног понашања. Код симулација комплексних система,

најчешће је потребно посматрати велики број елемената и испитати утицај различитих параметара, што може бити рачунарски веома захтевно.

Уобичајена подела на системе је на основу врсте њихових градивних елемената (на пример, атоми у физици, протеини у биологији, појединци у социологији), али без обзира на то, у различитим системима могу постојати слични принципи самоорганизације. Како се одређене појаве могу испољити као универзалне макроскопске карактеристике у различитим системима које су независне од детаља на микроскопском нивоу система, закључци који произилазе из проучавања једног система могу се генерализовати. То омогућава нову класификацију комплексних система, на основу универзалних карактеристика, које се објашњавају заједничким механизмима и параметрима.

Увођењем поједностављења где се занемарују специфичности појединачних елемената, комплексан систем се може анализирати применом метода теорије комплексних мрежа, тако да је нагласак на проучавању структуре интеракција између елемената. За разлику од једноставнијих структура мрежа које се користе у неким теоријским моделима система где су елементи распоређени на решетки и могу постојати интеракције само између суседа на одређеној удаљености или случајне мреже у којој се везе између елемената успостављају на насумичан начин, емпиријска анализа је показала да у реалним системима мреже интеракција имају знатно сложенију структуру са хијерархијским уређењем, елементима груписаним у заједнице или неким другим карактеристичним својством [8, 9].

1.2 Комплексне мреже

Комплексна мрежа представља апстракцију неког комплексног система, где су издвојени само основни елементи и интеракције између њих, док су сви остали небитни детаљи занемарени. Елементи се пресликавају на чворове у мрежи, а везе између елемената су представљене као гране између чворова. Код појединих комплексних система, структура одговарајуће мреже не мора бити једнозначно одређена. Приликом описивања мреже, прво је потребно одредити који

елементи система ће бити мапирани на чворове и које интеракције су од интереса како би два чвора била повезана граном. Понекад је уместо експлицитних веза, подесније посматрати функционалну зависност између различитих делова система. На пример, различити делови у мозгу који се пресликавају на чворове у мрежи, могу бити повезани гранама уколико су истовремено активни током извршавања неког задатка [10]. Такође, у мрежи могу постојати гране које не одговарају релевантним везама између елемената, већ су последица шума и потребно их је идентификовати и уклонити [11–13].

Теорија графова, у којој се првенствено изучава статичка структура мреже, представља математичку основу теорије комплексних мрежа. Теорија комплексних мрежа, поред тога додатно подразумева и проучавање динамичких промена структуре мреже, као и динамичких процеса на мрежи у чему важну примену имају методе статистичке физике. Задаци статистичке физике су да на основу структурних промена на нивоу чворова и грана објасни макроскопске особине комплексне мреже и да испита утицај структуре на динамичке процесе. Мреже великих величина које је потребно анализирати захтевају употребу софтверских алата и искуство у обради великих количина података.

Током претходне две деценије, извршена је емпиријска анализа структуре мреже за широк спектар реалних система као што су енергетски, транспортни, економски, биолошки, социјални и бројни други [8–10], што је подразумевало и дефинисање различитих параметара који квантитативно описују структуру мреже. Предложени параметри могу описивати структуру на нивоу чвора или гране, на нивоу групе чворова или на нивоу целе мреже. Пре свега, требало би споменути расподелу степени чворова, коефицијент груписања и дијаметар мреже, као параметре који су међу првима анализирани у реалним мрежама и за које се испоставило да имају слична својства у великом броју различитих реалних мрежа. У раду [14] је прво анализом социјалних и биолошких мрежа показано да се дате мреже одликују високом вредношћу коефицијента груписања и малим дијаметром. Овакво понашање, које је познато као особина малог света, не може се објаснити једноставним теоријским моделима као што су регуларна или случајна мрежа. Поред тога, још једна битна појава јесте хетерогена

расподела повезаности чворова, која је прво анализирана у социјалним и техничким мрежама [15] и бројна истраживања која су затим уследила показала су да је то универзална тополошка карактеристика већине реалних комплексних мрежа [16].

Упоредо са емпиријском анализом структуре мреже, развијени су теоријски модели помоћу којих се може конструисати мрежа са карактеристичним особинама [14, 15, 17–22]. Изучавање ових модела има вишеструку примену. Они доприносе разумевању најбитнијих механизма који доводе до формирања одређених структурних образаца, што омогућује процену како ће се структура мрежа даље развијати. Како се ради о стохастичким моделима, ансамбл генеричких мрежа се може искористити за тестирање значајности тополошких параметара реалне мреже [23]. Поред тога, поређење различитих модела мреже има важну улогу у испитивању утицаја који структура има на динамичке процесе који се одвијају на мрежи [9, 24].

Динамички процеси на мрежи описују промене стања елемената у комплексном систему током времена услед међусобних интеракција. Под стањем елемента се подразумева нека величина придружена елементу која је од интереса у анализи. Представљањем комплексног система помоћу комплексне мреже, сваки појединачни елемент система се пресликава у чвор са придруженом променљивом која је временски зависна и описује стање датог елемента у одговарајућем тренутку, односно може се посматрати као динамички процес на мрежи. Један од највише проучаваних процеса у социјалним, али и у рачунарским мрежама, јесте ширење епидемије [25–30], где у зависности од модела, чвор може бити у здравом, инфицираном или излеченом стању, при чему се стање може променити након интеракције са суседним чворовима. Такође, у социјалним мрежама је детаљно проучавано и формирање одлука и ставова применом модела теорије игара [31–33]. Требало би споменути још и моделе рутирања у телекомуникационим мрежама [34, 35], као и процес синхронизације [36, 37] и размене капитала [38, 39]. Приликом анализе наведених модела динамичких процеса, показало се да је утицај који има топологија мреже изузетно велики.

Нови приступи у проучавању су временске и вишеслојне мреже [40, 41], које

представљају проширење на важне појаве у реалним системима. На развој временских мрежа највише су утицали емпиријски подаци прикупљени из социјалних система, где се везе између појединаца успостављају само у одређеним временским интервалима. Овакво понашање може да има велике последице на динамичке процесе, као што је ширење информација или вируса. Концепт вишеслојних мрежа је развијен како би се у разматрању раздвојиле интеракције које су по својој природи другачије, али се не могу независно анализирати, већ се структурна и динамичка организација посматрају на више слојева. Међу многобројним примерима из стварног света могу се издвојити енергетска и комуникациона мрежа чије је функционисање међусобно условљено [42, 43]. Други типичан пример су односи између појединаца у социјалним системима, који се могу посматрати на нивоу социјалних мрежа и на нивоу физичке комуникације, који међусобно утичу један на други [44].

На основу овог кратког прегледа различитих праваца у проучавању комплексних мрежа, закључује се да оне дају веома користан оквир за анализу најбитнијих својстава комплексних система. Социјални системи спадају међу највише проучаване системе применом метода комплексних мрежа. Иако су неке од појава већ успешно објашњене, услед изузетне сложености људских интеракција, постоје бројна отворена питања везана за основне принципе социјалног понашања и структуру заједница које је потребно размотрити.

1.3 Социофизика

У последње две деценије, применом метода теорије комплексних мрежа у анализи великих скупова података који су прикупљени за различите социјалне системе, квантификоване су бројне особине понашања појединаца, које су пре тога биле описане само квалитативно или уопште нису ни разматране. Као што је раније већ речено, структура социјалних мрежа се одликује хетерогеном расподелом степени чворова, релативно малим растојањима између два случајно изабрана чвора и високом вредношћу коефицијента груписања [14, 15]. Структура социјалних мрежа је својствена и по присуству густо повезаних подмрежа

(заједница) које су са малим бројем грана повезане са остатком мреже [45], из чега даље следи да се гране могу разврстати на оне које имају улогу да повежу различите заједнице и на гране између појединаца које једну заједницу чине стабилнијом [46, 47]. Додатно, показано је да у социјалним групама постоји тенденција успостављања веза између сличних појединаца, особина позната као хомофилија [48].

Развијање математичких модела који би требало да опишу карактеристичне обрасце понашања у социјалним системима значајно је како због разумевања основних принципа социјалне динамике, тако и због важне практичне употребе коју ови модели могу имати у алокацији ресурса, анализи ширења информација или епидемије, али и у оптимизацији успеха у каријери и квалитету животног стандарда уопштено. Моделирање социјалног система је изузетно сложено због утицаја многобројних фактора на активности и интеракције између појединаца, а нарочито због дејства психолошке компоненте. Практично је неизводљиво издвојити и укључити све те факторе код моделирања динамике на микроскопском нивоу. Међутим, у већини случајева то није ни неопходно, јер се показало да постоје универзални механизми који су у основи социјалне динамике и помоћу којих се може објаснити макроскопско понашање [49]. Имајући све ово у виду, код модела се уводе апроксимације и у разматрање се узимају најбитније особине које имају утицај на посматране активности појединаца. Основна идеја је да се на основу микроскопских модела уз примену метода статистичке физике испита колективно социјално понашање које је последица размене информација и других интеракција између чланова. Предложени модел може бити верификован поређењем добијених теоријских резултата са емпиријским подацима.

Код објашњења појединих појава у друштву искоришћени су већ постојећи математички модели који су развијени у физичким, биолошким или неким другим системима. На пример, за моделирање саобраћаја између два града користи се гравитациони модел, у коме као параметри фигуришу величине градова и растојање између њих [50, 51], док су код описа кретања људи у маси примењене методе кинетичке теорије гасова [52]. Осим модела који су преузети, предложе-

ни су и нови модели, као што је модел научне сарадње [21] или модел доношења одлуке на адаптивној социјалној мрежи [53]. Детаљнији преглед модела везаних за опис социјалне динамике може се наћи у [4, 49, 54].

Досадашњи резултати емпиријске анализе и теоријски модели обухватају само одређене социјалне групе. Такође, током времена, долази до већих промена у друштву. Најзначајније промене последњих деценија последица су развоја нових технологија које су утицале на начин комуникације и размене информација између појединаца. Трендови понашања различитих социјалних група у савременом окружењу представљају актуелну тему истраживања у области физике социјалних система.

1.4 Предмет и циљ истраживања

У овој дисертацији се изучавају принципи организације комплексних мрежа код којих структура еволуира у времену, са фокусом на мреже које описују интеракције у социјалним системима. Основни циљ је да се применом метода статистичке физике и теорије комплексних мрежа испита који механизми утичу на понашање појединаца у социјалним групама чија је динамика дискретна и условљена колективним активностима у којима учествују чланови групе. Једна од најважнијих особина ових група је да њени чланови комуницирају лицем у лице. Карактеристике оваквих система су знатно мање проучаване у литератури, у односу на системе који подразумевају комуникацију посредством неког уређаја [55–60]. Разлог за то је доступност емпиријских података. Велике количине података које се генеришу и чувају приликом комуникације где се користе различити електронски уређаји, олакшавају анализу система, док у првом случају није једноставно доћи до података. За потребе ове дисертације прикупљени су и форматирани подаци о учесницима и реализованим догађајима за шест конференција из различитих научних области за период од три деценије и подаци са портала Meetup за четири групе из различитих категорија.

На основу прикупљених података, биће испитана динамика појављивања појединаца на колективним догађајима, што подразумева статистичку анализу

образаца учешћа, детекцију универзалних карактеристика у посматраним групама и развијање теоријског модела који би требало да објасни добијене резултате. Пре свега, важно је показати да ли појединци учествују насумично на догађајима или се могу уочити елементи организације. У најједноставнијем случају, када се претпостави да је учествовање насумично, преласци између стања се могу описати Бернулијевим случајним процесом и расподела броја учествовања има експоненцијални облик. Ипак, статистичке карактеристике у реалним системима, могу бити знатно другачије, што указује на присуство неких спољашњих или унутрашњих утицаја на понашање. То захтева развијање и тестирање нових модела који ће укључити релевантне параметре. Конкретно, овде ће применом генерализованог бинарног Појиног модела бити размотрено како на одлуку да ли ће члан учествовати на догађају утиче његова претходна активност.

Зависност структуре социјалне мреже и динамике система биће емпиријски анализирана посматрањем промена локалних структурних параметара мреже са бројем учешћа чланова на догађајима, чиме ће се истовремено добити информација о утицају који социјална компонента има на колективну динамику учешћа чланова у активностима групе. Специфична организација посматраних социјалних група, где појединци интерагују посредством колективних догађаја у дискретним временским тренуцима, омогућава испитивање улоге појединачних групних активности у еволуцији мреже, што до сада није посматрано. На основу добијених резултата, могу се показати корелације између величине догађаја и интеракција на догађају између појединаца са везама које могу бити слабе или јаке.

1.5 Структура рада

Најпре су у наредном поглављу представљене квантитативне методе које су коришћене у раду. Поред прегледа статистичких техника које се примењују у анализи емпиријских података, дефинисани су математички модели учествовања појединаца у колективним активностима и уведени су појмови из теорије комплексних мрежа који су од интереса за ово истраживање. У трећем поглављу је

описан поступак прикупљања и форматирања емпиријских података и дати су детаљи о њиховом садржају. Четврто поглавље се бави динамичким процесима у посматраним системима. Прво су приказане емпиријске расподеле броја укупног учествовања, броја узастопних учествовања и трајања пауза између учешћа на догађајима у групи, а затим је размотрен генерализовани Појин модел помоћу кога се могу објаснити добијени резултати. Оптимални параметри модела одређени су помоћу нумеричких симулација. На крају поглавља, дискутоване су заједничке карактеристике социјалне динамике у различитим групама које су овде анализирани. У петом поглављу се проучавају отежињене мреже добијене пројекцијом бипартитних мрежа које описују појављивање чланова на догађајима у групи. Помоћу отежињене мреже анализирани су особине социјалних интеракција, при чему је нагласак на испитивању промене локалних структурних параметара мреже са бројем учествовања, као и на испитивању утицаја које време одржавања догађаја и његова величина имају на глобалну структуру мреже. У последњем, шестом поглављу, сумирани су најважнији резултати рада.

2. Квантитативне методе

2.1 Статистичка анализа

Приликом анализе и моделирања појава из стварног света, у већини случајева карактеристично је присуство случајних променљивих које се описују одговарајућом расподелом. Уколико случајна променљива X може узети вредности из дискретног скупа $\Omega = \{x_1, x_2, \dots\}$ на коме је дефинисана вероватноћа P , расподела је дискретна $p(x_n) = P(X = x_n)$. Кумулативна функција расподеле (или краће функција расподеле) дефинисана је са $F(x_n) = P(X \leq x_n)$. У случају када је X континуална случајна променљива, за опис расподеле користи се функција густине вероватноће f , тако да важи $F(x) = \int_{-\infty}^x f(t)dt$. Тада је вероватноћа да случајна променљива X припада интервалу $[a, b]$ дата са $P(a \leq X \leq b) = \int_a^b f(t)dt$.

У применама се анализира емпиријска расподела случајне променљиве X на скупу од N узорака који представљају исходе статистичког експеримента, при чему се ту подразумева и само посматрање неке појаве. Ако је са N_x^c означен број узорака чија је вредност $\leq x$, функција $F_N(x) = \frac{N_x^c}{N}$ представља емпиријску функцију расподеле помоћу које се може апроксимирати теоријска функција расподеле $F(x)$ случајне променљиве X . Потребно је да број узорака N буде довољно велики како би апроксимација била оправдана, тј. важи

$$\lim_{N \rightarrow \infty} F_N(x) = F(x), \quad (2.1)$$

за свако $x \in \mathbf{R}$. За дискретну случајну променљиву X , може се посматрати и емпиријска расподела $p_N(x) = \frac{N_x}{N}$, где N_x представља број узорака који имају вредност x . Емпиријска расподела $p_N(x)$ конвергира ка теоријској расподели

$p(x)$ када $N \rightarrow \infty$.

За задати скуп узорака и одговарајућу емпиријску расподелу, основни задаци у истраживању су математички опис, односно одређивање врсте расподеле, затим оцењивање параметара, као и представљање посматраног реалног система математичким моделом тако да се његовим решавањем могу репродуковати емпиријски резултати. Расподеле које се често јављају приликом моделирања и анализе емпиријских података у комплексним мрежама и социјалним системима и које се користе и за анализу емпиријских података у овој дисертацији су експоненцијална расподела, степена расподела и степена расподела са експоненцијалним репом.

У табели 2.1 дате су дефиниције наведених расподела за дискретну случајну променљиву X , док су у табели 2.2 дате дефиниције континуалних расподела. Код степене расподеле и степене расподеле са експоненцијалним репом, доња гранична вредност x_{min} мора бити већа од нуле, пошто посматране расподеле нису дефинисане за вредности $x \leq 0$, док је код експоненцијалне расподеле услов $x_{min} \geq 0$.

Табела 2.1: Дефиниције дискретних расподела $p(x)$ и вредности одговарајуће нормализационе константе C , где $\zeta(\gamma, x_{min}) = \sum_{n=0}^{\infty} \frac{1}{(x_{min} + n)^\gamma}$ означава Хурвицову зета функцију и $\text{Li}_\alpha(e^{-\beta}) = \sum_{n=1}^{\infty} n^{-\alpha} e^{-\beta n}$ означава полилогаритамску функцију.

	експоненцијална расподела	степенна расподела	степенна расподела са експоненцијалним репом
$p(x)$	$Ce^{-\lambda x}$	$Cx^{-\gamma}$	$Cx^{-\alpha}e^{-\beta x}$
C	$(1 - e^{-\lambda})e^{\lambda x_{min}}$	$\frac{1}{\zeta(\gamma, x_{min})}$	$\frac{1}{\text{Li}_\alpha(e^{-\beta})}$

Табела 2.2: Дефиниције функције густине вероватноће континуалних расподела $f(x)$ и вредности одговарајуће нормализационе константе C , где $\Gamma(1 - \alpha, \beta x_{min}) = \int_{\beta x_{min}}^{\infty} t^{-\alpha} e^{-t} dt$ означава непотпуну гама функцију.

	експоненцијална расподела	степенна расподела	степенна расподела са експоненцијалним репом
$f(x)$	$Ce^{-\lambda x}$	$Cx^{-\gamma}$	$Cx^{-\alpha}e^{-\beta x}$
C	$\lambda e^{\lambda x_{min}}$	$(\gamma - 1)x_{min}^{\gamma-1}$	$\frac{\beta^{1-\alpha}}{\Gamma(1-\alpha, \beta x_{min})}$

Врста расподеле датог скупа података може се претпоставити на основу познавања својства посматране појаве или неком другом методом, а затим се приступа оцени параметара. У случајевима када се не може са сигурношћу претпоставити врста расподеле већ постоје два или више избора, за сваки облик ће се извршити оцена параметара, а затим ће се одговарајућим поступком проценити која од посматраних теоријских расподела најбоље одговара емпиријској расподели [61].

Постоје различите методе за оцену параметара. Овде ће се користити метод максималне веродостојности [61, 62]. Нека су у општем случају непознати параметри расподеле означени са θ . Код експоненцијалне расподеле то је параметар λ , у случају степене расподеле $\theta = \gamma$, док је код степене расподеле са експоненцијалним репом $\theta = (\alpha, \beta)$. У примени метода максималне веродостојности потребно је одредити вредности θ тако да вероватноћа $\mathcal{L}(\theta) = P(X_1, X_2, \dots, X_N)$ реализације независног узорка X_1, X_2, \dots, X_N буде максимална. Вероватноћа $\mathcal{L}(\theta)$ се назива функција веродостојности и у случају дискретне расподеле $p(x|\theta)$ са параметром θ дефинисана је са

$$\mathcal{L}(\theta) = \prod_{n=1}^N p(X_n|\theta) = p(X_1|\theta)p(X_2|\theta) \dots p(X_N|\theta), \quad (2.2)$$

док за континуалну расподелу са функцијом густине $f(x|\theta)$ важи

$$\mathcal{L}(\theta) = \prod_{n=1}^N f(X_n|\theta) = f(X_1|\theta)f(X_2|\theta)\dots f(X_N|\theta). \quad (2.3)$$

У пракси се одређивање максимума функције $\mathcal{L}(\theta)$ може свести на проблем проналажења минимума $-\log \mathcal{L}(\theta) = -\sum_{n=1}^N \log p(X_n|\theta)$, који се за неке расподеле може решити аналитички. Када је аналитичко решавање проблема превише компликовано или није могуће, могу се користити нумеричке методе за минимизацију, као што су, на пример, БФГС (Бројден-Флечер-Голдфарб-Шано) алгоритам, Нелдер-Мид метода (амоеба метод) или Пауел метода [63].

Након оцене параметара, потребно је проверити претпоставку да су емпиријски подаци из посматране врсте расподеле. Један од приступа је тестирање хипотезе о расподели, као што је описано за степену расподелу у [61]. Код овог поступка прво је потребно одредити растојање D између емпиријске F_N и теоријске F кумулативне функције расподеле са оптималним параметром θ , применом метода Колмогорова и Смирнова

$$D = \sqrt{N} \max_{x \geq x_{min}} |F_N(x) - F(x)|. \quad (2.4)$$

Затим се генерише довољно велики број скупова узорака из посматране теоријске расподеле са параметром θ . За сваки генерисани скуп се направи емпиријска расподела и поново изврши оцена параметара $\hat{\theta}$. Када се одреди растојање \hat{D} између емпиријске расподеле добијене из генерисаних података и теоријске расподеле са параметром $\hat{\theta}$, потребно је упоредити D и \hat{D} . Процент генерисаних скупова података код којих је $\hat{D} > D$ представља p -вредност. На крају, уколико је p -вредност мања од изабраног нивоа значајности ϵ , одбацује се хипотеза да подаци потичу из посматране теоријске расподеле. За ниво значајности ϵ најчешће се узимају вредности 0.1, 0.05 или 0.01. Препорука је да се у зависности од изабране вредности ϵ генерише најмање $\frac{1}{4}\epsilon^{-2}$ скупова података. Овај метод тестирања хипотезе о расподели није увек поуздан, нарочито када је број емпиријских података N мали. Метод се може применити на више расподела и

уколико је код две или више расподеле p -вредност већа од ϵ , узима се расподела са већом p -вредности.

Други метод који се може применити је тест количника веродостојности [61]. Овде се пореде две расподеле p_1 и p_2 тако што се за сваку расподелу примени оцена параметара θ_1 и θ_2 , а затим се за оптималне параметре израчунају вредности функција веродостојности $\mathcal{L}_1(\theta_1)$ и $\mathcal{L}_2(\theta_2)$. Ако је логаритам количника функција веродостојности

$$\mathcal{R} = \log \frac{\mathcal{L}_1(\theta_1)}{\mathcal{L}_2(\theta_2)} = \sum_{n=1}^N [\log p_1(X_n) - \log p_2(X_n)] \quad (2.5)$$

позитиван, расподела p_1 представља бољи избор, односно, ако је вредност \mathcal{R} негативна, расподела p_2 је адекватнија, док се за вредности које теже нули не може направити поређење. Ипак, овакав закључак на основу знака може бити непоуздан. Када за различите реализације независних узорака X_1, X_2, \dots, X_N који су из исте расподеле, математичко очекивање логаритма количника веродостојности тежи нули, позитиван или негативан знак може бити последица флукуација, па је потребна додатна провера да ли је вредност $|\mathcal{R}|$ довољно велика у односу на стандардну девијацију. Овде је потребно направити разлику између ситуације када се посматрају две угнеждене расподеле ¹, као што су, на пример, степена расподела и степена расподела са експоненцијалним репом и случаја када расподеле нису угнеждене. Полазећи од претпоставке да се ради о независном узорку X_1, X_2, \dots, X_N , када се посматрају две неугнеждене расподеле и вредности $l_n = \log p_1(X_n) - \log p_2(X_n)$ се могу сматрати независним, па следи да логаритам количника веродостојности \mathcal{R} има нормалну расподелу са варијансом $N\sigma^2$. Варијанса σ^2 се апроксимира са

$$\sigma^2 = \frac{1}{N} \sum_{n=1}^N (l_n - \bar{l})^2, \quad (2.6)$$

где је $\bar{l} = \frac{1}{N} \sum_{n=1}^N l_n$. Вероватноћа да је добијена вредност \mathcal{R} последица флукуа-

¹nested distributions

ција када математичко очекивање тежи нули једнака је

$$\pi = \frac{1}{\sqrt{2\pi N}\sigma} \left[\int_{-\infty}^{-|\mathcal{R}|} e^{-t^2/2N\sigma^2} dt + \int_{|\mathcal{R}|}^{+\infty} e^{-t^2/2N\sigma^2} dt \right]. \quad (2.7)$$

Ако је вероватноћа π мања од одговарајуће вредности ϵ , поређење расподела p_1 и p_2 на основу \mathcal{R} се може прихватити. Када се тест примењује за две угнеждене расподеле, вредности $l_n = \log p_1(X_n) - \log p_2(X_n)$ и одговарајућа варијанса теже нули. Тада статистика \mathcal{R} за довољно велико N има хи квадрат расподелу χ^2 . Уколико је вероватноћа

$$\pi = 1 - F_{\chi^2}(|2\mathcal{R}|, 1) \quad (2.8)$$

мала, расподела p_1 која представља специјалан случај расподеле p_2 неће бити узета у разматрање. У посматраном изразу, $F_{\chi^2}(|2\mathcal{R}|, 1)$ представља кумулативну функцију хи квадрат расподеле.

2.2 Математички модели

Математичко моделирање представља једну од основних метода у научном истраживању, како у области природних и техничких наука, тако и у области друштвених наука. Математички модели се могу користити у комбинацији са експериментом над стварним системом. Када је помоћу експеримента могуће добити податке, математички модели се користе да би се описала и анализира својства система и да би се проценило понашање система у будућности. Како често није могуће извршити експеримент над стварним системом, његово проучавање се може заснивати искључиво на примени математичких модела који су формиран на основу познатих законитости и усвајањем претпоставки.

Нека се посматра систем за који су прикупљени емпиријски подаци и извршена статистичка анализа. Следећи корак у проучавању система је математички опис његовог понашања. Модел би требало да буде дефинисан тако да резултати који се добију помоћу њега у довољно доброј мери репродукују емпиријске податке. Приликом формирања математичког модела потребно је направити компромис између његове општости и прецизности. Битно је дефинисати

најважније особине система које је потребно укључити у модел и разликовати их од неких сувишних детаља који могу имати улогу шума. Компликовани модели са великим бројем параметара могу прецизније поновити емпиријске резултате, али са друге стране они укључују неке особине које су специфичне само за посматрани узорак и ограничавају ширу употребу модела. Модел код кога је број параметара реда као и број података у посматраном узорку нема пуно смисла.

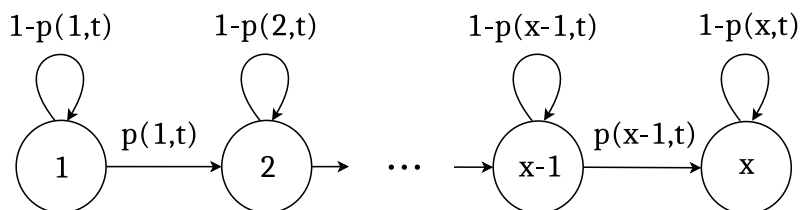
Један од начина за оцену параметара $v = (v_1, v_2, \dots)$ математичког модела јесте поређење емпиријске расподеле $p_N(x)$ и расподеле $p(x|v)$ која се добија када се за опис система користи дати математички модел. Расподела $p(x|v)$ се може добити аналитичким или нумеричким решавањем модела, као и коришћењем симулација. Подударност две расподеле може се квантификовати коришћењем Кулбак-Лејблеровог растојања

$$I = \sum_{n=1}^N p_N(x_n) \log \frac{p_N(x_n)}{p(x_n|v)}, \quad (2.9)$$

где су x_n дозвољене вредности случајне променљиве X тако да важи $0 < p_N(x_n) < 1$ и $0 < p(x_n|v) < 1$. Кулбак-Лејблерово растојање I представља губитак информација када се реалан систем апроксимира моделом са параметрима v [64]. У идеалном случају, када се у потпуности могу репродуковати емпиријски резултати $I = 0$, што се у пракси ретко дешава, пре свега због чињенице да је број мерења увек коначан. Циљ је формирати модел и одабрати параметре тако да се вредност Кулбак-Лејблеровог растојања минимизује.

Приликом моделирања социјалног система, два основна процеса која се посматрају су индивидуално динамичко понашање и интеракције између појединаца. У овој дисертацији од интереса су модели који описују учествовања појединаца у колективним активностима. Посматрани модели су дискретни и стохастички.

Основна идеја приказана је на слици 2.1. У социјалном систему чланови организују колективне активности које у моделу представљају дискретне догађаје. Претпоставка је да се у сваком тренутку времена организује један догађај. По-



Слика 2.1: Модел учествовања појединаца у колективним активностима.

сматра се временска активност за једног члана. У тренутку $t = t_0$ посматрани члан се први пут појављује у систему. Нека је са x означен број учествовања у колективним догађајима. Вредност променљиве x у тренутку t представља стање система. У следећем тренутку $t + 1$ члан може учествовати на догађају и прећи из стања x у стање $x + 1$ или одлучити да не учествује на догађају и остати у истом стању x . Вредности $x(t = t_0), x(t = t_0 + 1), \dots, x(t = t_0 + n)$ описују еволуцију система у времену. У општем случају вероватноћа p прелаза из стања x у стање $x + 1$ у тренутку t зависи од ове две вредности $p = p(x, t)$. Вероватноћа да ће члан остати у истом стању је $q(x, t) = 1 - p(x, t)$. У моделу се претпоставља да ће у одређеном тренутку члан напустити систем. Вероватноћа да се то деси не зависи од временског тренутка t нити од стања x у коме се члан налази, односно то је нека константна вредност r . Потребно је одредити вероватноћу $f(x)$ да ће у тренутку када напусти систем члан имати укупно x учествовања на догађајима.

Овде ће прво бити размотрена два специјална случаја када је вероватноћа прелаза између стања p константна, $p = \mu$ и када ова вероватноћа не зависи од временског тренутка t , већ само од укупног броја учествовања x , $p = p(x)$. Први случај одговара Бернулијевом процесу, а други је врста хомогеног Марковског процеса. За ова два случаја могуће је аналитички одредити облик расподеле $f(x)$. На крају ће бити размотрен временски-нехомоген Марковски ланац када је $p = p(x, t)$.

2.2.1 Бернулијев процес

Нека је почетни тренутак $t_0 = 1$. Када је вероватноћа прелаза из стања x у стање $x + 1$ константна, $p = \mu$, вероватноћа $f(x, t)$ да у тренутку t члан има x

учествовања на догађајима има биномну расподелу

$$f(x, t) = \binom{t}{x} \mu^x (1 - \mu)^{t-x}, \quad t \geq x. \quad (2.10)$$

Овде је претпостављено да учесник са вероватноћом μ узима учешће у првом догађају. Вероватноћа да ће учесник напустити систем у тренутку $t = T + 1$, $T \geq 1$ има геометријску расподелу $(1 - r)^T r$, па је вероватноћа да ће завршити учествовање са укупно x , $x \leq T$ догађаја једнака

$$f(x) = \sum_{T=x}^{\infty} f(x, T) (1 - r)^T r. \quad (2.11)$$

Решавањем се добија да укупан број учествовања има експоненцијалну расподелу

$$f(x) = (1 - e^{-\lambda}) e^{-\lambda x}, \quad (2.12)$$

са параметром

$$\lambda = \ln \left(1 + \frac{r}{\mu(1 - r)} \right). \quad (2.13)$$

Детаљно извођење се може наћи у додатку [A.1](#). Поступак када се стање мења континуално у времену приказан је у [\[65\]](#). Ту се, такође, као решење добија експоненцијална расподела.

2.2.2 Хомоген Марковски ланац

Следећи случај који се разматра јесте када вероватноћа прелаза између стања зависи само од тренутног стања у коме се систем налази, $p = p(x)$. Еволуција система у времену описана је једначином

$$f(x + 1, t + 1) = p(x) f(x, t) + q(x + 1) f(x + 1, t), \quad (2.14)$$

која показује да учесник у тренутку $t + 1$ може имати $x + 1$ учествовање у догађајима, ако је у тренутку t имао x учествовања и активан је у тренутку $t + 1$ или ако у тренутку t има тачно $x + 1$ учествовање и одлучио је да не учествује у следећем догађају.

За почетни тренутак се узима $t_0 = 0$ и претпоставка је да члан учествује на првом догађају, односно $f(x = 1, t = 0) = 1$. Једначина 2.14 се може решити рекурзивним поступком, као што је описано у додатку A.2. Као и у претходном примеру, претпоставља се да је вероватноћа да ће учесник напустити систем константна, па је тражена вероватноћа да ће члан на крају имати укупно x учествовања

$$f(x) = \sum_{T=x-1}^{\infty} f(x, T)(1-r)^T r = \frac{r}{1-r} \frac{p(1) \dots p(x-1)}{\left(\frac{r}{1-r} + p(1)\right) \dots \left(\frac{r}{1-r} + p(x)\right)}. \quad (2.15)$$

Детаљан поступак приказан је у додатку A.2.

Једноставан пример који се може посматрати јесте када вероватноћа $p(x)$ расте линеарно са x , $p(x) = p_0 x$. Потребно је да важи $p_0 \ll 1$ тако да буде задовољен услов $p_0 x < 1$. У посматраном случају вероватноћа $f(x)$ има бета расподелу

$$f(x) = \frac{r}{p_0(1-r)} B\left(x, 1 + \frac{r}{p_0(1-r)}\right), \quad (2.16)$$

која се за велике вредности x може апроксимирати степеном расподелом

$$f(x) \sim x^{-\gamma}, \quad (2.17)$$

са параметром

$$\gamma = 1 + \frac{r}{p_0(1-r)}. \quad (2.18)$$

У [65] су, такође, разматрали случај када $p(x)$ расте са x , али је ова зависност сложенија

$$p(x) = 1 - e^{-(rx)^\gamma}. \quad (2.19)$$

Уз одговарајуће апроксимације, као решење се добија да вероватноћа $f(x)$ опада са степеним законом $f(x) \sim x^{-\gamma}$ за вредности $x < \frac{1}{r}$, док за вредности $x > \frac{1}{r}$ вероватноћа опада експоненцијално $f(x) \sim e^{-rx}$. Овде је важно напоменути да наведене апроксимације важе само за вредности параметра γ које су мање од 1. За $\gamma > 1$ расподела је бимодална и прво за мале вредности x опада, па у једном

тренутку почиње да расте, након чега опет опада експоненцијално.

2.2.3 Временски-нехомоген Марковски ланац

Када вероватноћа прелаза зависи од стања x у коме се систем налази и од временског тренутка t , $p = p(x, t)$, једначина која описује еволуцију система има облик

$$f(x + 1, t + 1) = p(x, t)f(x, t) + (1 - p(x + 1, t))f(x + 1, t). \quad (2.20)$$

Као почетни услов у овом моделу се узима да је $t_0 = 1$ и $x(t_0) = 1$. Матрица прелаза M зависи од временског тренутка t . Елемент $M_{i,j}(t)$ матрице прелаза представља вероватноћу да систем који је у тренутку $t - 1$ у стању i у тренутку t буде у стању j , из чега следи

$$M_{i,j}(t) = \begin{cases} 1 - p(i, t - 1), & j = i \text{ и } i < t, \\ p(i, t - 1), & j = i + 1 \text{ и } i < t, \\ 0, & \text{у осталим случајевима.} \end{cases}$$

У почетном тренутку $t = 1$ важи

$$M_{i,j}(1) = \begin{cases} 1, & j = i = 1, \\ 0, & \text{у свим осталим случајевима.} \end{cases}$$

Ако се посматра производ

$$D(t) = V \cdot (M(1) \cdot M(2) \dots M(t - 1) \cdot M(t)), \quad (2.21)$$

где су M квадратне матрице димензија $t \times t$ и V је матрица врста димензија $1 \times t$ чији су сви елементи јединице, резултујућа матрица $D(t)$ је матрица врста са елементима $D_j(t)$ који представљају вероватноћу да се у тренутку t систем налази у стању j , односно $f(j, t) = D_j(t)$.

Расподела укупног броја учествовања добија се као и у претходна два при-

мера

$$f(x) = \sum_{T=x}^{\infty} f(x, T)(1-r)^T r. \quad (2.22)$$

Приликом нумеричког прорачуна вероватноће $f(x)$, сумирање се може завршити у тренутку T_{max} када $r(1-r)^{T_{max}} \rightarrow 0$.

У овој дисертацији од интереса су модели у којима вероватноћа прелаза расте са бројем догађаја на којима је члан учествовао, али такође може и да опада у времену ако члан изостаје са догађаја, односно остаје у истом стању. Појин модел кутије са куглицама (Pólya urn model) је један пример оваквог модела. Ту је основна идеја да се у једној кутији налазе црне и беле куглице и сваки пут када се из кутије на случајан начин извуче једна куглица, у кутију се убади одређен број нових куглица исте боје као извучена куглица. На тај начин, вероватноћа да ће бити извучена куглица одговарајуће боје зависи од броја претходних извлачења куглице исте боје. Постоје различите генерализације основног модела у којима се могу посматрати куглице са више од две боје, различит број кутија у којима се налазе куглице или различита правила приликом додавања нових куглица у кутију након сваког извлачења [66–70].

Модел који се користи за опис учествовања појединачног члана у колективним активностима у овој дисертацији представља генерализовани бинарни Појин модел, у коме је вероватноћа извлачења куглице пропорционална броју куглица исте боје које се налазе у кутији на неки степен a и након сваког извлачења у кутију додаје још једна куглица исте боје као извучена куглица [71]. Другим речима, члан који у тренутку t има x учествовања на догађајима учествоваће на догађају у тренутку $t + 1$ са вероватноћом

$$p(x, t) = \frac{x^a}{x^a + (y + y_0)^a}, \quad (2.23)$$

где $y = t - x$ представља број догађаја на којима члан није учествовао. Како су у почетном тренутку $t = 1$ вредности $x = 1$ и $y = 0$, уведен је додатни параметар y_0 који може узети позитивну целобројну вредност. У посматраном систему

параметар y_0 се може представити као почетна препрека коју је потребно превазићи, односно број учествовања које би након првог догађаја члан требало да има на наредним догађајима да би вероватноћа учествовања била већа од вероватноће неучествовања.

Понашање функције $p(x, t)$ за различите вредности параметра a разматрамо је у [67, 68]. Када је $a < 1$ након великог броја догађаја, број учествовања и неучествовања ће бити подједнак и вероватноћа $p(x, t)$ ће осциловати око $\frac{1}{2}$. Код овог случаја постоји ефекат негативне повратне спреге. Са друге стране, за $a > 1$ присутан је механизам позитивне повратне спреге. Ако је члан на почетку упоран и достигне y_0 учествовања, вероватноћа учествовања на наредним догађајима ће веома брзо расти. У супротном, вероватноћа учествовања ће бити све мања и мања. Случај када је $a = 1$ одговара класичном Појином моделу.

2.3 Комплексне мреже

Комплексан систем који се састоји од објеката између којих постоје интеракције, може се представити помоћу мреже, у којој чворови одговарају објектима система, док гране мреже које повезују парове чворова означавају интеракције између објеката. Комплексне мреже имају широку примену у изучавању комплексних система, укључујући ту и социјалне системе, где се под појмом интеракције осим физичког контакта и непосредне комуникације између појединаца подразумева и комуникација посредством неког уређаја, као и коауторство и узајамна цитираност на радовима у научним заједницама [9, 54, 72, 73].

Постоје два правца у проучавању комплексних мрежа. Први се односи на структуру мреже, пре свега на алгоритме конструисања мреже и одређивање параметара који показују како је мрежа организована применом теорије графова [8, 9, 74–76]. Други приступ подразумева посматрање динамичких процеса на мрежама, шта је за те процесе специфично, како структура мреже утиче на динамику и како би изгледала оптимална мрежа [9, 75, 77].

2.3.1 Структура комплексне мреже

Структура комплексне мреже може се представити помоћу два скупа $(\mathcal{N}, \mathcal{E})$ таквих да је $\mathcal{N} \neq \emptyset$ и \mathcal{E} представља скуп (не)уређених парова из скупа \mathcal{N} . Елементи из $\mathcal{N} = \{n_1, n_2, \dots, n_N\}$ представљају чворове мреже, док елементи из $\mathcal{E} = \{e_1, e_2, \dots, e_E\}$ представљају гране мреже. Укупан број елемената у посматраним скуповима \mathcal{N} и \mathcal{E} означен је са N и са E .

Величина мреже дефинисана је као број чворова N . Редни број чвора i у скупу \mathcal{N} најчешће се користи као његова ознака. Грана је одређена паром чворова i и j и као ознака се користи (i, j) или e_{ij} и каже се да су грана e_{ij} и чворови i и j инцидентни. Два чвора која су спојена граном зову се суседна. Код неоријентисаних или симетричних мрежа гране нису усмерене, тј. $e_{ij} = e_{ji}$. Ако се сматра да су то различите гране, мрежа се назива оријентисана или антисиметрична.

Веома битну примену у испитивању структурних и динамичких особина комплексне мреже има њихово приказивање помоћу матрице суседства. Матрица суседства A је квадратна $N \times N$ матрица чије врсте, као и колоне, одговарају чворовима мреже, тако да вредност елемента a_{ij} показује да ли између та два чвора постоји грана (1 ако постоји или 0 ако не постоји). Гране чија се оба краја налазе у истом чвору називају се петље. У наставку се посматрају само неоријентисане мреже без петљи ($a_{ii} = 0$). У овим мрежама матрица суседства је симетрична.

Наизменична секвенца чворова и грана, односно секвенца суседних чворова, која почиње са чвором i и завршава се са чвором j представља пут који повезује чворове i и j . Растојање између чворова i и j једнако је најкраћој дужини пута, односно минималном броју грана помоћу којих се i и j могу повезати. Ако се било која два чвора у мрежи могу повезати путем, мрежа је повезана. У супротном, мрежа је неповезана. Подскупови чворова и грана у неповезаној мрежи, такви да пут постоји само између чворова унутар истог подскупа, док се два чвора из различитих подскупова не могу повезати, називају се компоненте повезаности.

Степен чвора i ,

$$k_i = \sum_{j=1}^N a_{ij}, \quad (2.24)$$

дефинише се као број грана које имају један крај у чвору i . Када је за сваки од N чворова у мрежи одређен степен, може се проценити расподела степени чворова $p(k)$ која представља вероватноћу да ће случајно изабрани чвор имати степен k

$$p(k) = \frac{1}{N} \sum_{i=1}^N \delta(k_i - k) = \frac{N_k}{N}, \quad (2.25)$$

где N_k означава број чворова степена k у мрежи. Средњи степен чвора у мрежи једнак је

$$\langle k \rangle = \frac{1}{N} \sum_{i=1}^N k_i = \frac{2E}{N}. \quad (2.26)$$

Степен чвора показује како је посматрани чвор повезан са остатком мреже, док средњи степен чвора пружа информацију о просечној густини повезаности читаве мреже. Помоћу средњег степена чвора могу се поредити мреже различитих величина. Ипак, најдетаљнији увид пружа расподела степени чворова. Познато је да се у случајним мрежама са великим бројем чворова N код којих се чворови насумично повезују [78, 79], расподела степени чворова може апроксимирати Поасоновом расподелом

$$p(k) = e^{-\langle k \rangle} \frac{\langle k \rangle^k}{k!}. \quad (2.27)$$

Ако је вероватноћа да две гране буду повезане ρ , средњи степен чвора је $\langle k \rangle = \rho(N - 1)$. Може се рећи да је оваква мрежа хомогено повезана. Међутим, у реалним системима, везе између објеката се не успостављају насумично и имају другачију расподелу степени чворова. Већина расподела које су генерисане на основу емпиријских података има степену расподелу

$$p(k) \sim k^{-\gamma} \quad (2.28)$$

и у већини мрежа је $2 < \gamma < 3$ [15, 80–83]. Код мрежа са степеном расподелом степени чворова, гране су хетерогено распоређене. Највећи проценат чворова

има мали број суседа, али постоје и чворови са јако великим бројем суседа, тако да је највећи број грана распоређен на свега неколико чворова. Растојања између чворова су услед централизованости мала [84]. Степена расподела је јединствена по томе што након скалирања облик расподеле остаје исти

$$p(ak) = f(a)p(k), \quad (2.29)$$

па се мреже са оваквом расподелом називају још и мреже без скале. У неким реалним мрежама, расподела степени чворова опада по степеном закону до одређене вредности када почиње да опада експоненцијално, односно мрежа има степену расподелу са експоненцијалним репом [85, 86]. Овакво понашање може бити последица недовољно великог броја чворова у мрежи на основу којег се одређује расподела. Осим тога, у неким мрежама, као што су технолошке, инфраструктурне или социјалне мреже, чворови имају ограничене ресурсе па број конекција по чвору не може бити неограничено велики.

Информацију о повезаности суседа чвора пружа коефицијент груписања. Коефицијент груписања представља вероватноћу да су два чвора која имају заједнички суседни чвор такође међусобно повезана. За чвор i степена k_i вредност коефицијента груписања се рачуна као

$$C_i = \frac{\sum_{j=1}^N \sum_{l=1}^N a_{ij} a_{il} a_{jl}}{k_i(k_i - 1)} = \frac{2E_i}{k_i(k_i - 1)} \quad (2.30)$$

где E_i представља број грана између k_i суседа чвора i . Коефицијент C_i може имати вредност између 0 и 1, при чему је $C_i = 0$ уколико не постоји ниједна грана између суседа чвора i , док је $C_i = 1$ када су суседни чворови сви међусобно повезани. Средњи степен груписања у мрежи одређен је средњим коефицијентом груписања $\langle C \rangle$, који се добија усредњавањем по свим чворовима

$$\langle C \rangle = \frac{1}{N} \sum_{i=1}^N C_i. \quad (2.31)$$

Вредност средњег коефицијента груписања може се интерпретирати као просечна вероватноћа да ће за неки чвор у мрежи два произвољно изабрана суседа

бити повезана.

Реалне мреже имају већи средњи коефицијент груписања у односу на случајне мреже које имају сличну густину повезаности [14, 87–89]. Ово својство реалних мрежа објашњава се постојањем локалног груписања на нивоу чвора које подразумева да ако је чвор i повезан са чвором j и j је истовремено повезан са l , онда ће са великом вероватноћом i такође бити повезан са l . Како је расподела степени чворова у мрежи хетерогена, да би се добио детаљнији увид требало би посматрати средњи коефицијент груписања $\bar{C}(k)$, који се одбија усредњавањем по свим чворовима истог степена k

$$\bar{C}(k) = \frac{\sum_{i=1}^N C_i \delta(k_i - k)}{\sum_{i=1}^N \delta(k_i - k)}. \quad (2.32)$$

У мрежама код којих коефицијент груписања опада са порастом степена чвора [90, 91], постоји хијерархијска организација где чворови малог степена са својим суседима формирају густо повезану групу, док су код чворова великог степена суседи повезани са малом вероватноћом, односно чворови великог степена са својим суседима не образују једну групу већ повезују више различитих група чворова.

Понекад је поред податка да ли су чвор i и чвор j повезани граном e_{ij} , битан и интензитет интеракције. На пример, да ли два научника имају више од једног заједничког рада или колико је трајао разговор између две особе. У том случају, грани e_{ij} се додељује тежина $w_{ij} \geq 0$ и каже се да је посматрана мрежа отежињена. Приликом визуелизације отежињене мреже, обично се користе различите дебљине грана или различите боје, како би се представиле њихове тежине. У отежињеној матрици W , елементи w_{ij} имају вредности које одговарају тежини гране између посматрана два чвора, уколико су они повезани. Када чвор i и чвор j нису повезани, вредност w_{ij} је једнака 0.

Код отежињене мреже, могу се дефинисати додатне тополошке карактеристике. Снага чвора i

$$s_i = \sum_{j=1}^N w_{ij}, \quad (2.33)$$

представља збир тежина свих грана које су инцидентне чвору i . Ова величина садржи информације о учешћу које чвор има у функционисању мреже. Тако у рачунарској мрежи снага чвора може представљати количину саобраћаја који тај чвор може да опслужи, у мрежи аеродрома укупан број летова са посматраног аеродрома или у социјалној мрежи где појединци комуницирају помоћу електронске поште, снага чвора може да представља укупан број мејлова које је појединац разменио са другима. Код два чвора која имају исту снагу, један може имати неколико пута мањи степен, па приликом проучавања отежињене мреже, снагу и степен чвора би требало узети заједно у разматрање. Слично као код коефицијента груписања, могу се посматрати корелације између степена и снаге чвора помоћу усредњене вредности $\bar{s}(k)$. Када су тежине грана хомогено распоређене, између снаге и степена чвора важи релација

$$\bar{s}(k) = \langle w \rangle k, \quad (2.34)$$

где је $\langle w \rangle$ усредњена тежина по свим гранама у мрежи. Када постоје корелације између степена чвора и тежине његових грана, ова зависност неће бити линеарна.

Отежињени коефицијент груписања

$$C_i^w = \frac{1}{s_i(k_i - 1)} \sum_{j=1}^N \sum_{l=1}^N \frac{w_{ij} + w_{il}}{2} a_{ij} a_{il} a_{jl} \quad (2.35)$$

осим информације о повезаности суседа чвора, носи и информацију о средњој тежини тих веза [92, 93]. Након нормирања збира са $s_i(k_i - 1)$, вредности отежињеног коефицијента груписања могу имати вредности између 0 и 1. Ако се упореде вредности C_i^w и C_i , може се показати да ли постоји правилност приликом груписања. Када је $C_i^w = C_i$, тежине грана немају улогу приликом груписања чворова, за разлику од случаја $C_i^w > C_i$ где су груписани чворови којима одговарају велике тежине грана или случаја $C_i^w < C_i$, који указује на груписање између чворова са малим тежинама грана.

2.3.2 Бипартитна мрежа

Мрежа код које је у скупу чворова \mathcal{N} могуће идентификовати два дисјунктна подскупа \mathcal{V} и \mathcal{U} тако да гране мреже повезују само чворове из дисјунктних подскупова, што значи да чворови унутар истог подскупа не могу бити повезани, назива се бипартитна мрежа. Матрица суседства се у случају бипартитне мреже дефинише на мало другачији начин. Нека је са V и U означен број елемената у подскуповима \mathcal{V} и \mathcal{U} . Матрица суседства B је матрица димензија $V \times U$ чији врсте представљају чворове подскупа \mathcal{V} , док колоне одговарају чворовима из \mathcal{U} . Елемент матрице b_{ij} има вредност 1, ако су чвор i и чвор j повезани граном. У супротном је $b_{ij} = 0$. Степен чвора i из подскупа \mathcal{V} једнак је броју суседа у подскупу \mathcal{U}

$$k_i = \sum_{j=1}^U b_{ij}. \quad (2.36)$$

За чвор j из подскупа \mathcal{U} важи обрнуто.

Реални комплексни системи којима се може придружити структура бипартитне мреже су бројни. Гени и поремећаји које узрокују [94], корисници и услуге, односно производи [95–99], мрежа научника и публикација [100, 101] су само неки од примера. Реални системи, који се проучавају у овој дисертацији и чији су ентитети учесници и догађаји, могу се, такође, посматрати као примери бипартитне мреже [102].

Како су у бипартитној мрежи гране између чворова који припадају истом подскупу забрањене, да би се анализирале карактеристике односа између ових чворова, бипартитна мрежа се може пројектовати на две монопартитне мреже. У једној пројекцији бипартитне мреже чворови ће припадати скупу \mathcal{V} , а у другој скупу \mathcal{U} . У монопартитној мрежи чворови i и j који припадају скупу \mathcal{V} биће повезани граном, ако у бипартитној мрежи имају бар један заједнички суседни чвор l из скупа \mathcal{U} . Тежина гране (i, j) једнака је броју заједничких суседа у бипартитној мрежи.

2.3.3 Филтрирање отежињене мреже добијене пројекцијом бипартитне мреже

Један од проблема који се може јавити након пројекције бипартитне мреже на један подскуп чворова јесу гране у отежињеној монопартитној мрежи које представљају "лажне" (или случајне) интеракције између чворова. Чворови из истог подскупа који у бипартитној мрежи имају велики степен са великом вероватноћом ће бити повезани граном у монопартитној мрежи. Такође, у монопартитној мрежи требало би направити разлику између грана које имају једнаке тежине, када су у бипартитној мрежи њихови чворови повезани са заједничким суседом који има велики степен и ситуације када су повезани са суседом малог степена. У првом случају њихова веза у монопартитној мрежи може бити од мањег значаја у односу на други случај.

У циљу детектовања релевантних веза у отежињеној монопартитној мрежи, у овој дисертацији се користи статистички тест као што је предложено у [13, 103]. За статистику теста узима се вероватноћа π_{ij} да чворови i и j имају w_{ij} или више заједничких суседа у одговарајућем ансамблу бипартитних мрежа. Ако је статистика теста већа од задатог нивоа значајности p_c , претпоставка да је грана (i, j) последица случајних интеракција не може се одбацити и веза између чворова i и j у монопартитној мрежи биће занемарена. Ову претпоставку треба одбацити када је $\pi_{ij} \leq p_c$ и тада се веза између чворова i и j може сматрати релевантном.

Као нулти (референтни) модели бипартитних мрежа са којима се пореди оригинална мрежа изабране су мреже са истом расподелом степени чворова у подскуповима \mathcal{V} и \mathcal{U} као у оригиналној мрежи [13, 99, 103, 104]. Вероватноћа π_{ij} се може одредити ако се бипартитна мрежа прво рандомизује превезивањем грана тако да се не мења расподела степени чворова, а затим се пројектује на монопартитну мрежу за изабрани подскуп чворова. Када се конструише довољно велики број рандомизованих мрежа и њихових пројекција, удео отежињених монопартитних мрежа код којих је тежина гране између два чвора већа или једнака тежини гране између та два чвора у оригиналној мрежи, представља

вероватноћу π_{ij} . Описани поступак за израчунавање вероватноће π_{ij} помоћу симулација предложен је у [105]. Недавно је у радовима [13, 103] приказан супериорнији метод помоћу кога се вероватноћа π_{ij} може одредити директно. Детаљан опис метода дат је у наставку.

Прво ће бити размотрена основна идеја за општи случај [106]. Нека се посматра ансамбл мрежа \mathcal{G} са задатим тополошким карактеристикама $\{x_i\}$, при чему се ту могу подразумевати број чворова и грана у мрежи, расподела степени чворова, корелације између степени чворова и слично. Ако је $P(G)$ вероватноћа неке мреже G у ансамблу \mathcal{G} , ентропија ансамбла је

$$S = - \sum_{\mathcal{G}} P(G) \ln P(G). \quad (2.37)$$

Ентропија ансамбла биће максимална када је вероватноћа $P(G)$ једнака

$$P(G) = \frac{1}{Z} e^{-\sum_i \lambda_i x_i}, \quad (2.38)$$

где су λ_i Лагранжови множиоци, док је статистичка сума дефинисана са

$$Z = \sum_G e^{-\sum_i \lambda_i x_i}. \quad (2.39)$$

Средња вредност параметра мреже x_i у ансамблу може се изразити помоћу

$$\langle x_i \rangle = \sum_G x_i(G) P(G) = - \frac{\partial}{\partial \lambda_i} \ln Z. \quad (2.40)$$

Пошто су вредности параметара $\{x_i\}$ унапред задате, решавањем система једначина 2.40 могу се одредити вредности Лагранжових множилаца λ_i .

Нека се сада посматра конкретан случај, када је задата бипартитна мрежа са $\{k_1, k_2, \dots, k_V\}$ вредностима степена чворова из подскупа \mathcal{V} и $\{d_1, d_2, \dots, d_U\}$ вредностима степена чворова из подскупа \mathcal{U} . Заменом задатих услова за степене чворова у израз 2.39 добија се

$$Z = \sum_G e^{-\sum_i \alpha_i k_i - \sum_l \beta_l d_l} = \sum_G e^{-\sum_{il} (\alpha_i + \beta_l) B_{il}} = \prod_{il} (1 + e^{-(\alpha_i + \beta_l)}). \quad (2.41)$$

Лагранжови множиоци α_i и β_l се добијају решавањем система нелинеарних једначина

$$k_i = -\frac{\partial}{\partial \alpha_i} \ln Z = \sum_{l=1}^M \frac{e^{-\alpha_i - \beta_l}}{1 + e^{-\alpha_i - \beta_l}}, \quad (2.42)$$

$$d_l = -\frac{\partial}{\partial \beta_l} \ln Z = \sum_{i=1}^N \frac{e^{-\alpha_i - \beta_l}}{1 + e^{-\alpha_i - \beta_l}}. \quad (2.43)$$

Посматрани систем неједначина може се решити једноставно итеративним нумеричким поступком [103]. Након тога, у следећем кораку потребно је одредити вероватноћу P_{il} појаве гране између чворова i и l у ансамблу бипартитних мрежа. Ако је $\lambda_{il} = \alpha_i + \beta_l$, важи

$$Z = \sum_G e^{-\sum_{il} \lambda_{il} B_{il}} = \prod_{il} (1 + e^{-\lambda_{il}}), \quad (2.44)$$

из чега даље следи

$$P_{il} = \langle B_{il} \rangle = -\frac{\partial}{\partial \lambda_{il}} \ln Z = \frac{e^{-\lambda_{il}}}{1 + e^{-\lambda_{il}}} = \frac{e^{-\alpha_i - \beta_l}}{1 + e^{-\alpha_i - \beta_l}}. \quad (2.45)$$

Када је вероватноћа P_{il} позната, може се одредити и вероватноћа P_{ijl} да чвор i и j из подскупа \mathcal{V} имају заједнички суседни чвор l из подскупа \mathcal{U} . Претпоставка је да не постоје корелације између чворова из истог подскупа, па важи $P_{ijl} = P_{il}P_{jl}$. На крају, вероватноћа $P_{ij}(w)$ да чвор i и j имају w заједничких суседа, тј. вероватноћа да је грана између ова два чвора у монопартитној мрежи тежине w , има Поасонову биномну расподелу

$$P_{ij}(w) = \sum_{\mathcal{U}_w} \prod_{l \in \mathcal{U}_w} P_{ijl} \prod_{\bar{l} \notin \mathcal{U}_w} (1 - P_{ij\bar{l}}), \quad (2.46)$$

где је \mathcal{U}_w подскуп од w чворова који су изабрани из скупа \mathcal{U} [13, 103, 107]. За одређивање Поасонове биномне расподеле у овој дисертацији користи се DFT-CF метод (Discrete Fourier Transform of characteristic function) који је предложен у [108]. У специјалним случајевима, расподела $P_{ij}(w)$ се може апроксимирати Поасоновом или Гаусовом расподелом [107].

Тражена вероватноћа π_{ij} једнака је збиру вероватноћа свих грана (i, j) у

ансамблу код којих је тежина w већа од тежине w_{ij} у оригиналној мрежи

$$\pi_{ij} = \sum_{w \geq w_{ij}} P_{ij}(w). \quad (2.47)$$

Када је вероватноћа π_{ij} да ће грана (i, j) у ансамблу имати тежину већу од w_{ij} мала, може се сматрати да ова грана носи важну информацију у оригиналној мрежи.

3. Подаци

Напредак у проучавању комплексних социјалних система у узајамној је вези са напретком у области информационо комуникационих технологија. Употребом мобилних телефона и разних сервиса на Интернету, свакодневно се генеришу и чувају велике количине података о међусобној комуникацији, географском положају и разним другим активностима појединаца. Могућност систематског сакупљања и статистичке анализе великих количина података коришћењем одговарајућих софтверских пакета утицали су на напредак у развоју модела који могу описати понашање и интеракције између појединаца, а самим тим и на боље разумевање и предвиђање будућег понашања система.

Доступност података који подразумевају комуникацију посредством неког уређаја је знатно већа у односу на податке који подразумевају непосредну, личну комуникацију између појединаца, па су карактеристике система са посредном комуникацијом детаљније проучаване. Да би се добиле информације о интеракцијама лицем у лице могу се користити различите сензорске технологије [109–111]. Овакви системи су релативно мали, са неколико десетина или стотина чланова, период посматрања је неколико дана до неколико месеци и интеракције између више од два члана су најчешће занемарљиве. Ова дисертација се фокусира на социјалне групе које могу имати и више хиљада чланова и чија је организација заснована на колективним активностима које подразумевају непосредну комуникацију већег броја појединаца [71, 102]. Прикупљени су подаци за два различита система: 1) подаци о учествовању на научним конференцијама и 2) подаци о учествовању на Meetup догађајима.

3.1 Научне конференције

Социјална динамика на научним скуповима проучавана је користећи податке са шест конференција из различитих области. Изабране су две конференције које организује америчко друштво физичара "American Physical Society March Meeting"(APSMM) и "American Physical Society April Meeting"(APSAM). Ужа научна област APSMM конференције је физика чврстог стања, док на APSAM конференцији учествују научници чије истраживање припада области физике високих енергија. Поред тога, посматра се конференција из области примењене математике "Society for Industrial and Applied Mathematics Annual Meetings"(SIAM), конференција из области неуралних мрежа "Neural Information Processing Systems Conference"(NIPS), конференција из области суперрачунара "International Conference on Supercomputing"(ICS) и једна конференција из молекуларне биологије "International Conference on Research in Computational Molecular Biology"(RECOMB).

Заједничко за свих шест конференција је, поред њихове дуговечности, да се одржавају сваке године, али се осим по области истраживања разликују и по броју учесника, укупном броју одржаних конференција, да ли су уско специјализоване или покривају различите теме, као и по томе да ли су организоване у једној или више држава. APSMM и APSAM конференције се одржавају у различитим градовима широм Сједињених Америчких Држава, док су остале четири конференције током година биле организоване у различитим државама. APSMM, APSAM и SIAM су конференције са дугом традицијом, али подаци о учесницима су доступни почевши од 1994. године. За NIPS и ICS, које су први пут организоване 1988. године и за RECOMB, која је први пут организована 1997. године, прикупљени су комплетни подаци. Информације везане за године одржавања конференције за које су прикупљени подаци и укупан број учесника током тог периода приказане су у табели 3.1, док су детаљне информације о броју учесника на појединачним конференцијима дате у табели 3.2.

Табела 3.1: Преглед научних конференција које су изабране за емпиријску анализу. Друга и трећа колона садрже информацију о почетној (Y_0) и крајњој (Y_f) години временског интервала за који су подаци прикупљени, док је у четвртој колони приказан укупан број различитих учесника на конференцији током посматраног временског интервала.

Конференција	Y_0	Y_f	Број учесника
APSM	1994	2014	78544
APSAM *	1994	2014	16264
SIAM **	1994	2014	8879
NIPS	1988	2014	6902
ICS	1988	2014	2504
RECOMB	1997	2014	1564

* Подаци нису доступни за 1999. годину.

** Подаци нису доступни за 2007. и 2011. годину

Подаци о прихваћеним радовима (презентације и постери) за посматрану годину доступни су на веб сајту конференције и прикупљени су уз помоћ скрипти написаних у програмском језику python које се извршавају у Selenium WebDriver-у. Сваки појединачни узорак садржи наслов рада, имена аутора и њихове афилијације. Сматра се да је научник који је потписан као један од аутора на раду учествовао на конференцији те године. На тај начин може се направити база податка која садржи информације о научницима и години учествовања на конференцији. Проблем који се том приликом јавља јесте да научници користе различите начине потписивања са пуним првим и средњим именом или само са првим именом или само иницијале, што може узроковати погрешну статистику. На великим конференцијама, као што су APSM и APSAM, додатни проблем је ситуација када два научника имају исто име и презиме. Због тога је пре него што се приступи статистичкој анализи, прво потребно очистити прикупљене податке.

Да би се направила прецизна идентификација различитих научника, могу се искористити информације о њиховим афилијацијама и коауторима и сваком

Табела 3.2: Број учесника на конференцији по годинама.

	APSMM	APSAM	SIAM	NIPS	ICS	RECOMB
1988	-	-	-	214	132	-
1989	-	-	-	205	121	-
1990	-	-	-	297	123	-
1991	-	-	-	302	116	-
1992	-	-	-	270	112	-
1993	-	-	-	301	114	-
1994	9660	3309	540	270	114	-
1995	9897	1947	425	292	144	-
1996	9991	2356	279	289	127	-
1997	9191	3388	579	289	109	111
1998	10924	2301	456	298	158	120
1999	20426	-	367	296	172	121
2000	10816	1744	403	307	105	150
2001	12401	1818	823	396	146	101
2002	11944	2446	1115	432	118	98
2003	13548	2127	642	469	103	95
2004	14595	1668	767	492	102	136
2005	14673	1140	792	515	165	141
2006	16484	1008	945	479	124	154
2007	16655	943	-	530	96	123
2008	16441	1473	1053	633	132	142
2009	16775	1630	1054	654	242	127
2010	17790	1342	1166	733	127	157
2011	18368	1088	-	746	171	167
2012	22343	1480	1223	938	133	148
2013	21510	1430	1205	884	210	125
2014	22789	1704	1407	1064	147	137

научнику доделити јединствена лабела, као што је предложено у [112]. Прво је потребно за сваки потпис издвојити почетно слово имена и презиме и то представља текстуални део лабеле. Затим се групишу сви потписи са истим текстом лабеле и одабере се један потпис коме се дода редни број 1, као други, нумерички део лабеле. Лабели ће као атрибути бити придружени скуп потписа, скуп коаутора и скуп афилијација. У почетном тренутку ови атрибути ће сарджати информације из само једног узорка из кога је узет почетни потпис. Онда се претражују сви груписани узорци у којима један од аутора има исти текстуални део лабеле. Прво се пореде коаутори из изабраног узорка са скупом коаутора који је придружен лабели. Ако постоји поклапање, сматра се да је у питању исти научник и његов потпис, коаутори и афилијације се додају у атрибуте придружене лабели, након чега се потпис из овог узорка сматра класификованим и неће се даље разматрати. Ако не постоји поклапање коаутора, пореде се афилијације и понавља се поступак. У случају када не постоји поклапање ни код коаутора, ни код афилијација, не може се закључити да се ради о истој особи и потпис из овог узорка се сматра неklasификованим. Описани поступак се понавља док се за све потписе из узорака из исте групе не утврди да ли им се може доделити иста лабела или се ради о различитим научницима који имају исто почетно слово имена и презиме. Након тога се изабере један од неklasификованих потписа из исте групе и на текстуални део лабеле се дода редни број 2 и понавља се претрага. Када се заврши класификација свих аутора из посматране групе додајући редне бројеве на текстуални део лабеле, поступак се понавља из почетка за следећу групу аутора. На крају, када су сви аутори идентификовани, лабеле које садрже текстуални део због лакше обраде података биће замењене редним бројевима од 1 до укупног броја различитих научника на конференцији.

Да би описани поступак чистијења података био у потпуности одређен, требало би објаснити и како су упоређени коаутори и афилијације. Код поређења коаутора, посматрају се почетно слово имена и презиме и потребно је да постоји потпуно поклапање. Приликом поређења афилијација, пошто их научници потписују на различите начине, могућа су два приступа. Прва опција која је раз-

матрана била је да се искористи Google геокодер како би се за задату афилијацију као повратна информација добили град и држава. Да би Google геокодер вратио тачне резултате, потребно је да афилијација садржи довољно података, што није увек случај. Овај проблем је решен употребом scikit-learn библиотеке за машинско учење у python-у, која је омогућила да се афилијације за које су одређени градови и државе искористе као скуп података за тренинг помоћу кога ће се алгоритам за класификацију научити и затим осталим афилијацијама доделити градове и државе. Ипак, овакав поступак се испоставио као непоуздан, пошто су у великим градовима као што су Пекинг, Њујорк и слично, сви научници са истим почетним словом имена и презименом класификовани као исте особе, при чему се често радило о научницима из различитих институција. Уместо тога, код другог приступа, афилијације су се поредиле директно, поређењем њихових текстуалних записа помоћу функције `token_set_ratio` из библиотеке `fuzzywuzzy` у python-у, при чему услов да поклапање буде веће од 50%. Класификација применом овог поступка се испоставила као прецизнија.

Код конференција APSMM, APSAM и SIAM, код којих подаци нису доступни од самог почетка одржавања конференције, већ као што је претходно речено тек од $Y_0 = 1994$. године, потребно је изоставити научнике који су учествовали на конференцијама пре тог периода. За ове научнике величине које се посматрају, број укупних учествовања, број узастопних учествовања без паузе и трајање паузе између два учествовања, неће бити тачне. Како не постоји прецизан начин да се ови научници детектују, посматра се вероватноћа њиховог појављивања на конференцијама пре датог периода. За сваку од ове три конференције одређено је просечно трајање паузе између два појављивања $\langle \tau \rangle$, тако што је прво за сваког учесника израчуната просечна пауза, а затим је ова величина усредњена по свим учесницима. При томе се трајање паузе између два доласка на конференцију мери у годинама, односно у броју конференција на којима научник није учествовао. Научници који су се први пут појавили на конференцији пре $Y_0 + \langle \tau \rangle$ године, изузети су из даље анализе, пошто се претпоставља да су учествовали на конференцији пре Y_0 године и да подаци о њиховој активности нису комплетни. Конкретно, након одређивања просечне

паузе, код све три конференције не посматрају се научници који су учествовали на конференцијама пре 1997. године. Процент научника за које се сматра да не постоје комплетни подаци износи 10% за APSMM и SIAM, док је за APSAM тај проценат виши и износи 25%.

Описани метод за процену да ли су подаци комплетни, разликује се од поступка који се користио у [113], где је за сваког појединца i одређено просечно трајање паузе τ_i и уколико је његова активност регистрована пре тренутка $Y_0 + \tau_i$, сматра се да су подаци непотпуни. Разлог зашто је овде посматрана просечно трајање паузе усредњено по свим појединцима $\langle \tau \rangle$, представља претпоставка да се вероватноћа учествовања код појединаца мења током времена, а самим тим и дужине пауза, па би непоузданост процене била већа када би се уместо $\langle \tau \rangle$ посматрало τ_i .

Потребно је напоменути, да за APSAM конференцију, која није одржана 1999. године, ова година се не узима као пауза и учествовања 1998. и 2000. године се посматрају као узастопна. Исто важи за 2007. и 2011. годину код SIAM конференције.

3.2 Meetup групе

Meetup је онлајн сервис који пружа могућност људима који деле заједничка интересовања да организују окупљања у офлајн простору. Корисник који се региструје на Meetup порталу може основати нову групу или се учланити у неку од већ постојећих група. Чланство у групи је бесплатно и добровољно. Постоји укупно 33 различите категорије које покривају различите области интересовања и свака Meetup група је придружена одговарајућој категорији. У табели 3.3 дат је списак свих категорија. Укупно има више од 240000 различитих група које су регистроване у 181 држави на различитим континентима.

Један или више чланова групе могу предложити и организовати окупљање са дефинисаним планом и програмом у одговарајућем термину и на одговарајућем месту, док би остали чланови групе, уколико су заинтересовани и у могућности да учествују, требало да потврде свој долазак. Догађај може бити

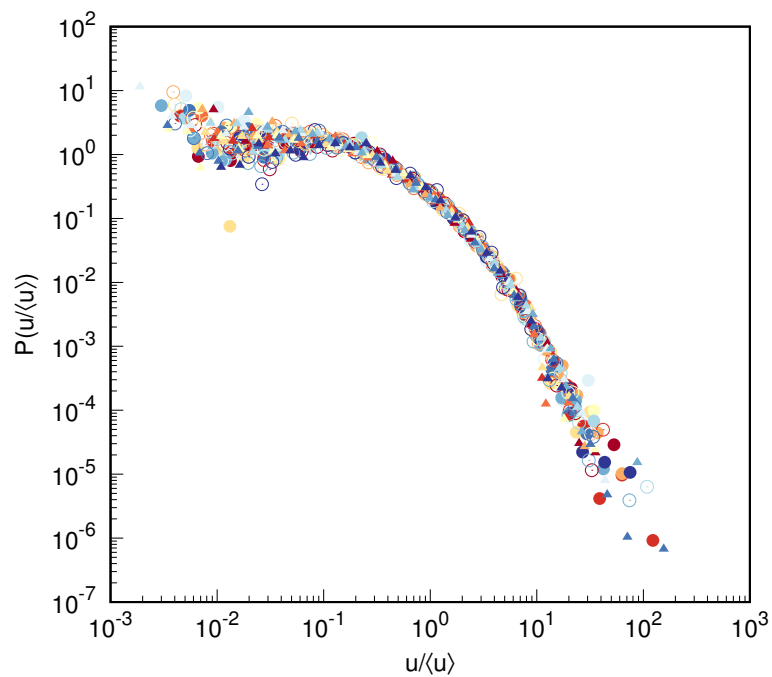
Табела 3.3: Преглед Meetup категорија.

Музика	Родитељи и породица	Хоби и занат
Политика	Религија	Аутомобили
Образовање	Паранормалне појаве	Ново доба и духовност
Здравље	Књижевност	Животна средина
Игрице	Жене	Мода
Фитнес	Кућни љубимци	Подршка
Каријера	Спорт	Животни стил
Научна фантастика	ЛГБТ	Технологија
Уметност и култура	Плес	Храна и пиће
Језик и етничке групе	Самци	Филм
Активности напољу	Дружење	Фотографија

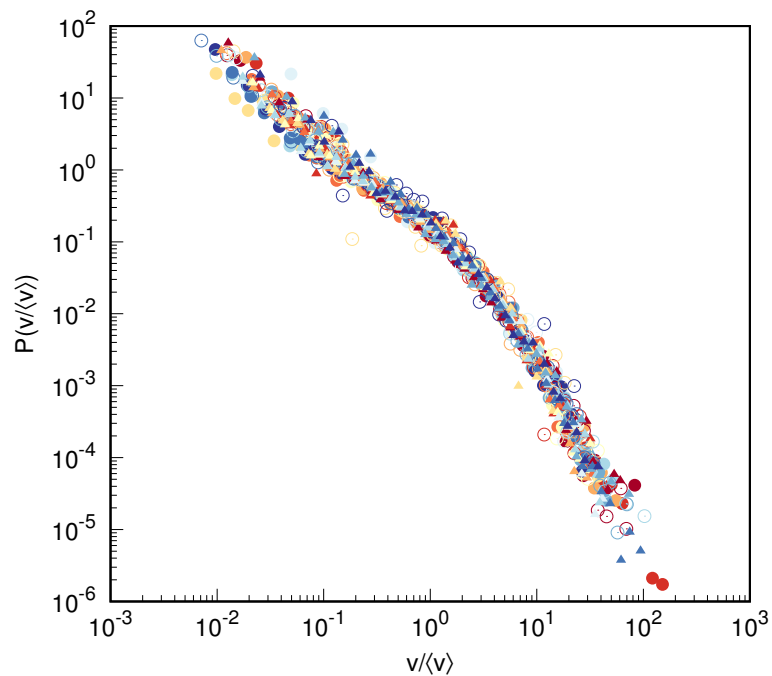
веома популаран ако већина чланова одлучи да учествује, али исто тако постоји могућност да нико од чланова не потврди учествовање на неком догађају. Број чланова и учестаност догађаја се мењају током времена и разликују се од групе до групе. Приликом избора групе, веома битну улогу има географска локација. Нема пуно смисла да се корисник учлани у групу где се догађаји организују на великој физичкој удаљености на којима не може да учествује. На слици 3.1 се може видети да је расподела величине група у оквиру исте категорије изузетно хетерогена. Такође, може се закључити да након скалирања са просечном величином групе у категорији $\langle u \rangle$, расподеле за све различите категорије имају исти облик. Слично важи и за расподелу броја организованих догађаја у групи, која је приказана на слици 3.2.

Подаци о групама у оквиру сваке категорије, члановима и догађајима организованим у групи чувају се у бази података и доступни су преко Интернета. Приступ подацима је бесплатан и омогућен је путем Meetup API-ја. Код слања упита, потребно је навести који подаци се очекују као одговор. Подразумевани формат података у коме ће бити одговор је JSON. Како су подаци структурирани, није потребно њихово чишћење, као у случају научних конференција. Сваком регистрованом кориснику додељена је јединствена лабела.

На почетку прикупљања података, прво је као одговор на упит добијена



Слика 3.1: Расподела величина група u скалираних са просечном величином групе $\langle u \rangle$ у датој Meetup категорији. Расподеле су одређене за сваку од 33 различите категорије. Величина групе једнака је броју регистрованих корисника.



Слика 3.2: Расподела броја организованих догађаја у групи v скалираних са просечним бројем организованих догађаја $\langle v \rangle$ у групама из посматране Meetup категорије. Приказане су расподеле за сваку од различитих категорија.

листа свих категорија, која се може видети у табели 3.3. Затим је за сваку категорију затражен списак група, информација о броју чланова групе и броју организованих догађаја у оквиру групе у том тренутку, као и информација да ли су подаци о појединачним члановима и догађајима јавно доступни, пошто администратори групе имају право да имена чланова и појединости везане за догађаје не буду видљиви. На крају су за четири групе из различитих категорија (погледати табелу 3.4), код којих је омогућен приступ свим подацима, прикупљени подаци о терминима сваког појединачног догађаја и члановима који су потврдили учешће. Изабране групе су велике и активне, са неколико хиљада чланова и организованих догађаја, као што се може видети из табеле 3.5. Из резултата приказаних на слици 3.3 се види да се у датим групама између одржавања два догађаја протекне свега неколико сати, док је на слици 3.4 приказано колико се чланова месечно учлани у групу од тренутка њеног оснивања.

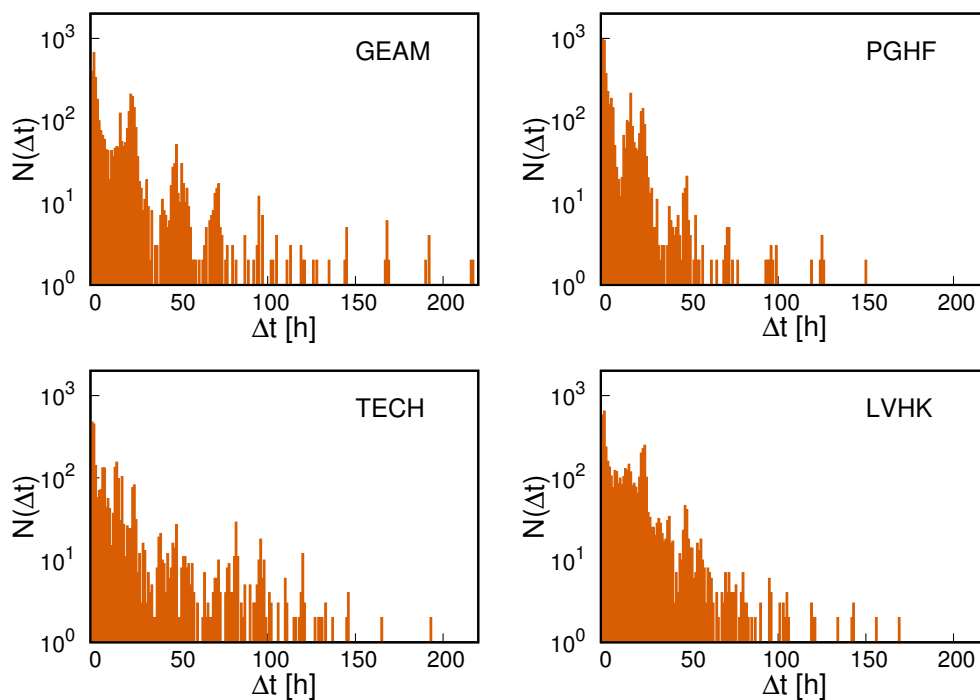
Табела 3.4: Приказ четири изабране Meetup групе.

Назив групе	Скраћеница	Категорија
geamclt	GEAM	Храна и пиће
pittsburgh-free	PGHF	Дружење
techlifecolumbus	TECH	Технологија
VegasHikers	LVHK	Активности напољу

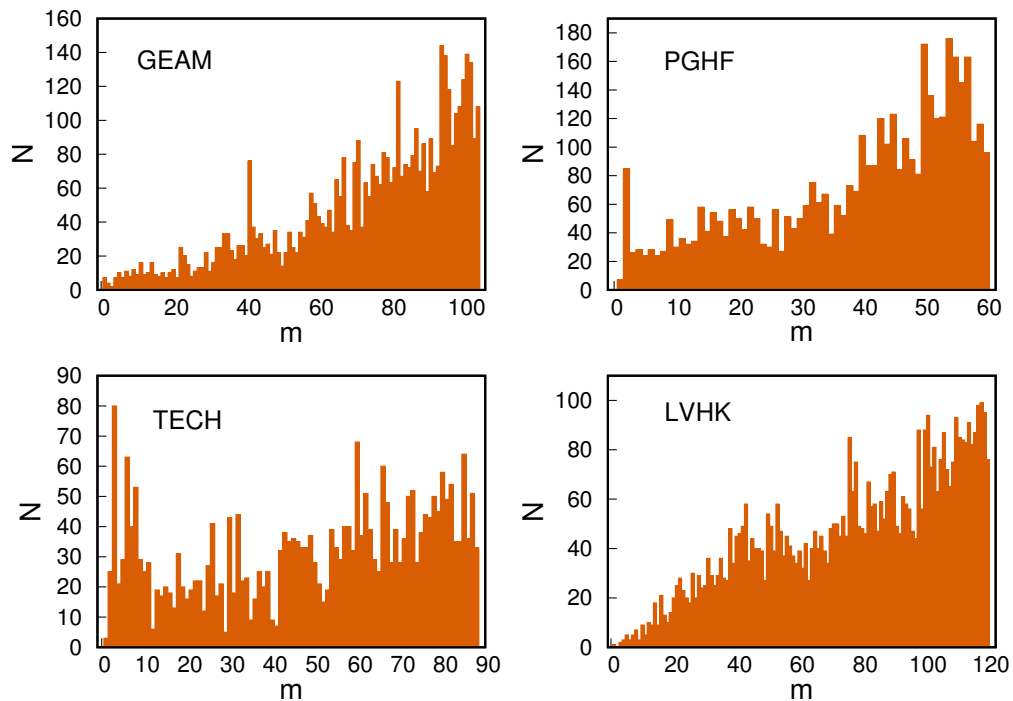
Група geamclt (GEAM) регистрована је у Шарлоту у Сједињеним Америчким државама. Њени чланови се окупљају у ресторанима и баровима како би пробали различита јела и коктеле. У групи pittsburgh-free (PGHF), чланови заједно иду на дешавања у Питсбургу, као што су концерти, фестивали, обиласци музеја и код којих је улаз бесплатан или јефтин. За разлику од ових група, чији је циљ дружење и забава, у techlifecolumbus групи (TECH) идеја је да се стручњаци из области информационих технологија упознају, размењују искуства и можда започну нову сарадњу. Седиште ове групе је у граду Коламбус

Табела 3.5: Преглед података за посматране Meetup групе. Број чланова и организованих догађаја се узима закључно са 18. мартом 2016. године.

Назив групе	Датум оснивања	Број чланова	Број догађаја
geamclt	01.04.2007.	5377	3986
pittsburgh-free	31.12.2010.	4995	4617
techlifecolumbus	28.08.2008.	3217	3162
VegasHikers	18.12.2005.	6061	5096



Слика 3.3: Хистограм трајања пауза између одржавања два догађаја. На x-оси, Δt представља време мерено у сатима. На y-оси је дат број пауза које трају Δt .



Слика 3.4: Број нових чланова N који су се регистровали у току сваког месеца m од тренутка оснивања групе ($m = 1$).

у Охају. Четврта изабрана група VegasHikers (LVHK) окупља појединце из Лас Вегаса који су заинтересовани за заједничке шетње и камповање у природи.

За ове четири групе сваком догађају је на основу тренутка одржавања додељен редни број као лабела почевши од 1 па до укупног броја догађаја. Чланови групе имају јединствене лабеле које представљају комбинацију слова и бројева, али су им због једноставније анализе додељене нове лабеле које, слично као код догађаја, узимају вредности од 1 до N , где N представља укупан број чланова.

4. Обрасци учешћа

У овом поглављу ће се разматрати динамика учествовања појединачних чланова у активностима различитих социјалних група. Како би се квантификовали обрасци учешћа, посматрају се расподеле укупног броја учествовања, расподела броја узастопног учешћа (без пауза), као и расподела трајања паузе између два учествовања. Прво је за емпиријске расподеле које су одређене на основу прикупљених података, испитано помоћу којих теоријских расподела се могу описати и извршена је оцена параметара, а затим је приказано како се применом математичког модела могу репродуковати резултати у реалним системима.

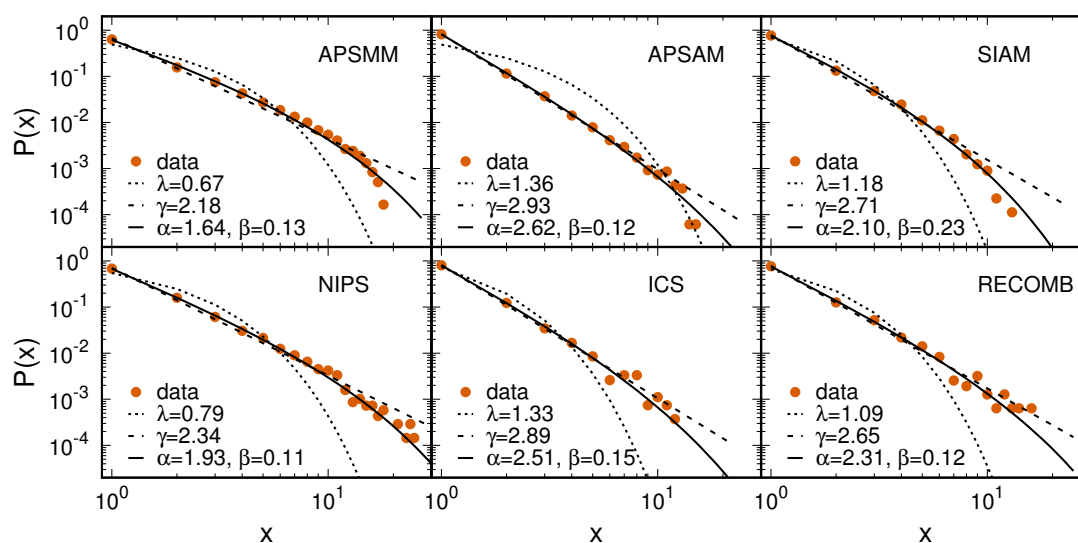
4.1 Емпиријска анализа

4.1.1 Научне конференције

За сваку од шест конференција које су изабране за анализу идентификовани су различити научници који су учествовали на конференцији током посматраног временског периода и сваком појединачно су придружени подаци о годинама на којима је учествовао на конференцији, као што је описано у претходном поглављу [3.1](#).

На основу прикупљених података, прво је анализирано колико је пута неки научник учествовао на конференцији. На слици [4.1](#) емпиријска расподела укупног броја учествовања, која је приказана наранџастим тачкама, показује да највећи проценат научника представи своје резултате на датој конференцији само једном. Код конференција APSAM и SIAM тај проценат је највиши и износи око 80%, док на APSMM и NIPS конференцијама око 60% научника има само једно учествовање. Процент научника који имају укупно x учествовања опада како x расте. На појединим конференцијама постоје и учесници који немају ни

један изостанак током посматраног периода. Њихов број је изузетно мали.



Слика 4.1: Расподела укупног броја учествовања на научним конференцијама. Наранџастим тачкама приказана је емпиријска расподела. Тачкаста линија одговара експоненцијалној расподели, испрекидана линија степеној расподели и пуна линија одговара степеној расподели са експоненцијалним репом.

Поред емпиријске расподеле, на истој слици су приказани и резултати добијени након фитовања података са три теоријске расподеле, при чему је оцена параметара извршена применом метода максималне веродостојности, код које је коришћена Нелдер-Мид метода за минимизацију. Тачкастом линијом приказана је експоненцијална расподела са оптималним параметром λ . Може се закључити да емпиријска расподела опада знатно спорије од експоненцијалне расподеле, на основу чега даље следи да се учествовање на конференцији не може описати једноставним математичким моделом где појединци долазе на догађаје са неком константном вероватноћом. Испрекидана линија одговара степеној расподели са параметром γ , а пуна линија степеној расподели са експоненцијалним репом и параметрима α и β . Да би се одговорило на питање која од ове две расподеле најбоље описује емпиријске резултате, примењен је тест количника веродостојности (погледати поглавље 2.1). У табели 4.1 се може видети да је вредност логаритма количника функција веродостојности \mathcal{R} негативна за свих шест конференција, што показује да је степена расподела са експоненцијалним репом бољи избор. Додатно је тестирана хипотеза да ли је овај резултат последица

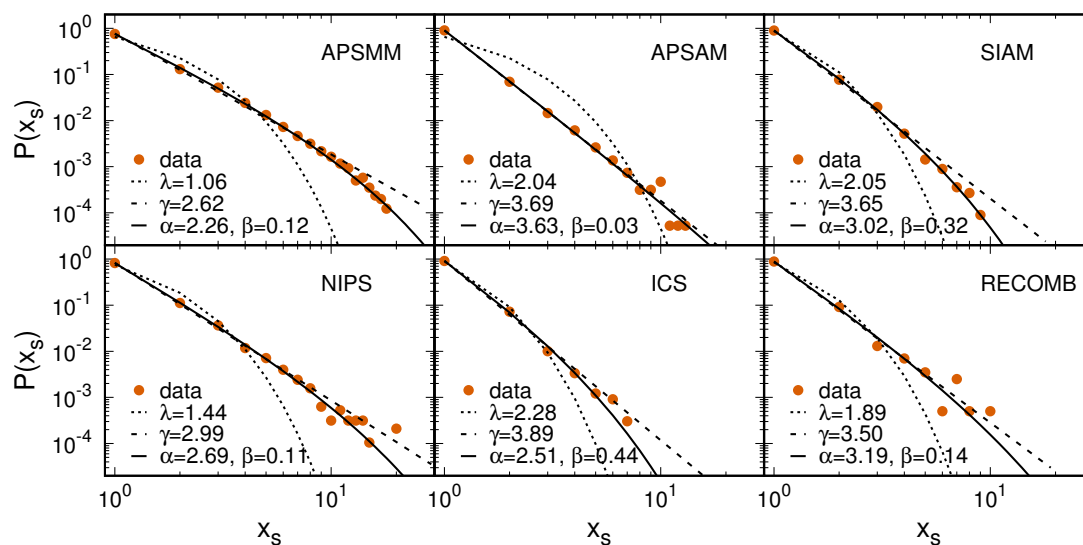
флуктуација, као што је објашњено у поглављу 2. Ова хипотеза је одбачена, пошто је вероватноћа $\pi \simeq 0$, као што је приказано у табели 4.1. Из приказаних резултата, може се закључити да вероватноћа да учесник има x учествовања на почетку опада по степеном закону са параметром $\alpha > 1.5$, а затим за веће вредности x опада експоненцијално.

Табела 4.1: Вредности параметара \mathcal{R} и π након примене метода количника веродостојности за степену расподелу и степену расподелу са експоненцијалним репом када се посматрају узорци који одговарају укупном броју учествовања на конференцији.

Конференција	\mathcal{R}	π
APSMM	-1758.44	0.0
APSAM	-36.89	0.0
SIAM	-75.26	0.0
NIPS	-76.64	0.0
ICS	-8.54	3.60e-05
RECOMB	-7.22	1.45e-04

Код научника који имају једнак број учествовања, доласци могу бити непрекидни или са повременим кратким одсуствовањем, али исто тако могу бити спорадични са дужим паузама. Да би се добио детаљнији увид у динамику учествовања, осим укупног броја појављивања, потребно је узети у разматрање како је оно организовано у времену. У скупу података у коме су задате године на којима је један од научника учествовао на конференцији, издвојене су секвенце узастопних година током којих је научник учествовао на конференцији без изостанака. Емпиријска расподела броја узастопних долазака приказана је наранџастим тачкама на слици 4.2. Подаци за APSMM и NIPS конференције показују да постоје учесници који током 20 година нису изостали ни са једне конференције. Са друге стране, код SIAM, ICS и RECOMB конференције не постоји учесник који је имао више од 10 узастопних учествовања.

Понављањем истог поступка као када је разматран укупан број појављивања и овде се показало да подаци имају облик степене расподеле са експоненцијалним репом, која је приказана пуном линијом. Оптималне вредности параметара



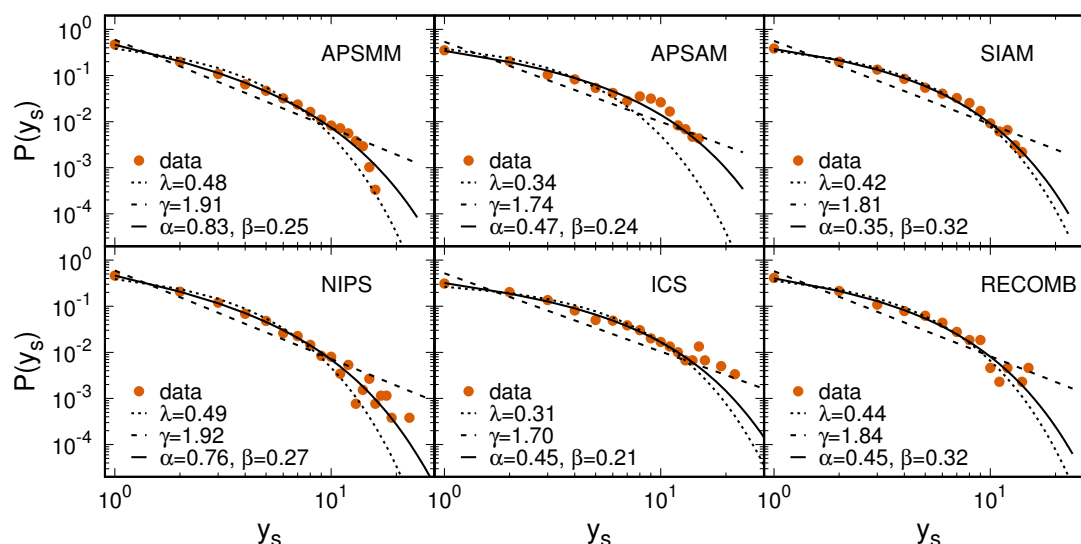
Слика 4.2: Расподела броја узастопних учествовања без пауза на научним конференцијама. Наранцастим тачкама приказана је емпиријска расподела. Тачкаста линија одговара експоненцијалној расподели, испрекидана линија степеној расподели и пуна линија одговара степеној расподели са експоненцијалним репом.

расподеле α и β могу се видети на слици 4.2, док су резултати добијени применом теста количника веродостојности за степену расподелу и степену расподелу са експоненцијалним репом дати у табели 4.2. Ако се упореде параметри расподеле укупног броја појављивања и расподеле броја узастопних појављивања, може се видети да расподела броја узастопних појављивања опада доста брже. То значи да код већине учесника између њиховог првог и последњег учествовања које је забележено на конференцији постоје паузе.

Да би се даље квантификовала динамика учествовања на конференцији, одређена је и расподела трајања паузе између два појављивања. На слици 4.3 наранцастим тачкама је приказана расподела добијена из прикупљених података, док су различитим линијама приказане три теоријске расподеле. Овде је очигледно да се степена расподела, која је приказана испрекиданом линијом, може искључити из даљег разматрања и потребно је испитати да ли је бољи избор за опис емпиријских података експоненцијална расподела или степена расподела са експоненцијалним репом. Након примене теста количника веродостојности, чији су резултати приказани у табели 4.3, утврђено је да треба

Табела 4.2: Вредности параметара \mathcal{R} и π након примене метода количника веродостојности за степену расподелу и степену расподелу са експоненцијалним репом када се посматрају узорци који одговарају броју узастопних учествовања без паузе на конференцији.

Конференција	\mathcal{R}	π
APSMM	-554.05	0.0
APSAM	-0.77	0.21
SIAM	-17.98	2.01e-09
NIPS	-17.52	3.24e-09
ICS	-4.99	1.57e-03
RECOMB	-1.48	0.09



Слика 4.3: Расподела дужине паузе између два учествовања на научним конференцијама. Наранџастим тачкама приказана је емпиријска расподела. Тачкаста линија одговара експоненцијалној расподели, испрекидана линија степеној расподели и пуна линија одговара степеној расподели са експоненцијалним репом.

користити степену расподелу са експоненцијалним репом. Резултати показују да највећи проценат пауза које научници направе између два појављивања на конференцији траје само једну годину. Ако научник одсуствује са конференције неколико година, вероватноћа да ће поново учествовати на конференцију опада веома брзо.

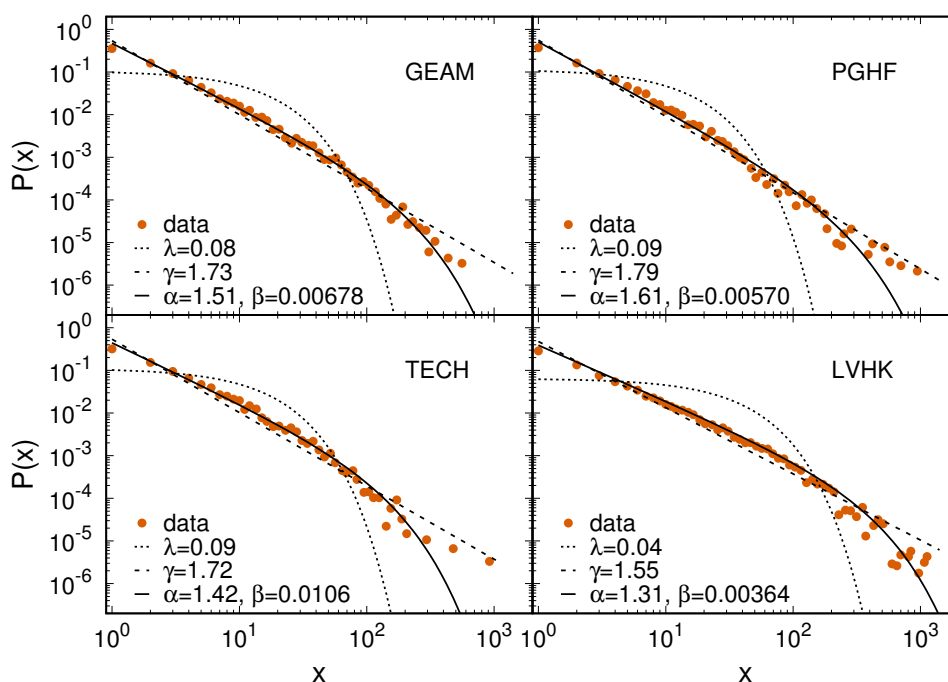
Табела 4.3: Вредности параметара \mathcal{R} и π након примене метода количника веродостојности за експоненцијалну расподелу и степену расподелу са експоненцијалним репом када се посматрају узорци који одговарају трајању паузе између два учествовања на конференцији.

Конференција	\mathcal{R}	π
APSM	-756.91	0.0
APSAM	-34.59	1.11e-16
SIAM	-11.54	1.55e-06
NIPS	-58.22	0.0
ICS	-7.64	9.24e-05
RECOMB	-3.60	7.27e-03

4.1.2 Meetup групе

Емпиријске расподеле су показале интересантне обрасце учествовања на научним конференцијама. За свих шест научних конференција, све три посматране расподеле, а то су расподела укупног броја учествовања, расподела броја узастопних учествовања и расподела трајања паузе, имају облик степене расподеле са експоненцијалним репом. То указује на изузетно хетерогену активност појединаца. Код највећег процента је забележено само једно учествовање, али са друге стране, постоје и они који су активни током читавог периода посматрања не направивши ни један изостанак. Такође, изостанци између два учествовања су најчешће кратки, али ту су и примери где су учесници били неактивни више од десет година, а затим су поново имали рад на истој конференцији. Да би се испитало да ли је овакво понашање карактеристично само за научну заједницу или је у питању општи тренд присутан и у другим социјалним групама, неопходно је анализирати и податке о другим заједницама. У наставку је анализирана активност чланова Meetup група које у односу на научну заједницу представљају потпуно другачији социјални систем.

За разлику од посматраних научних конференција које се организују једном годишње, време које протекне између два Meetup догађаја је знатно краће, случајна променљива која представља укупан број учествовања овде може имати доста веће вредности у поређењу са конференцијама. Ипак, ова два система



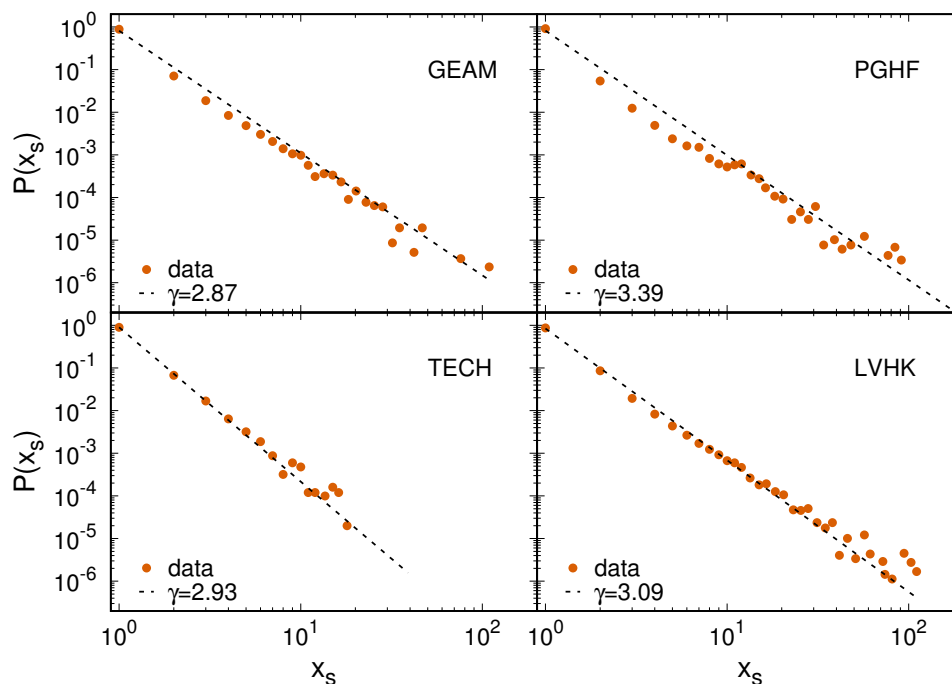
Слика 4.4: Расподела укупног броја учествовања на Meetup догађајима. Наранцастим тачкама приказана је емпиријска расподела. Тачкаста линија одговара експоненцијалној расподели, испрекидана линија степеној расподели и пуна линија одговара степеној расподели са експоненцијалним репом.

испољавају слично понашање. Као што се може видети на основу слике 4.4 и табеле 4.4, степена расподела са експоненцијалним репом најбоље фитује податке о укупном броју учествовања на Meetup догађајима, као и код конференција. Као последица великог броја организованих догађаја на којима чланови могу да учествују током посматраног временског периода и чињенице да тај период није дужи од 10 година, тако да се може сматрати да ће већина учесника задржати активан статус, вредност параметра β код степене расподеле са емпиријским репом на слици 4.4 је неколико десетина пута мања него код конференција. Активност чланова је изразито неравномерно распоређена и не може се груписати око неке средње вредности. Током више од 3000 организованих догађаја, највећи проценат чланова групе је дошао на само један догађај, након чега више нису били заинтересовани за учествовање, али може се уочити и да постоје малобројни, али веома активни чланови са неколико стотина, па чак и преко хиљаду учествовања.

Табела 4.4: Вредности параметара \mathcal{R} и π након примене метода количника веродостојности за степену расподелу и степену расподелу са експоненцијалним репом када се посматрају узорци који одговарају укупном броју учествовања на Meetup догађајима.

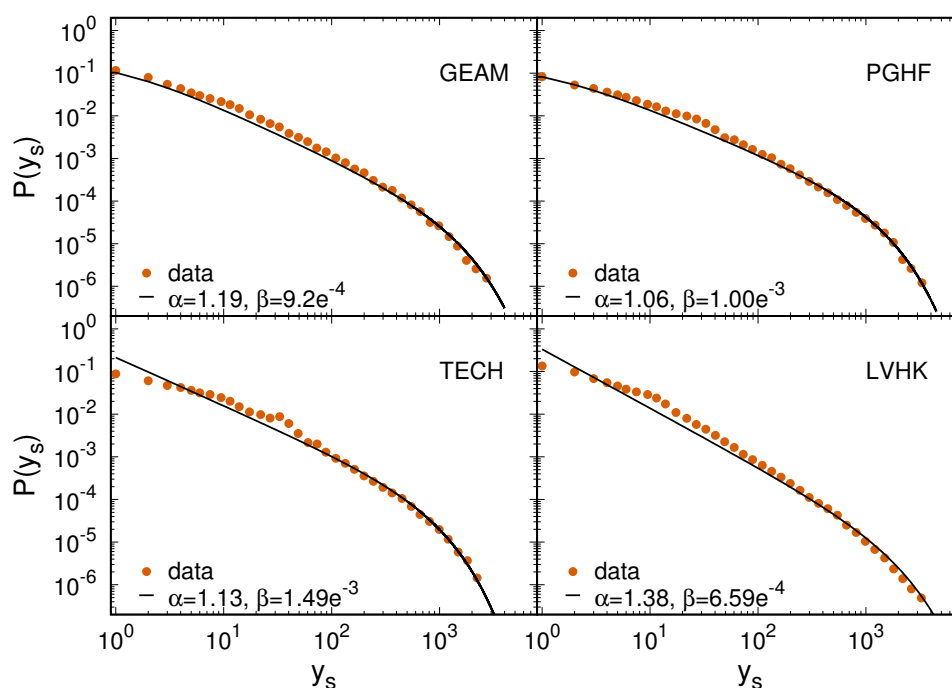
	\mathcal{R}	π
GEAM	-97.70	0.0
PGHF	-58.04	0.0
TECH	-90.84	0.0
LVHK	-236.52	0.0

Када се погледа расподела броја узастопних догађаја без изостанака на слици 4.5, може се видети да она опада по степену закону са параметром γ чија је вредност између 2.8 и 3.4. Максималан број узастопних долазака је око 100, осим код групе TECH где не постоји члан који је дошао на више од 20 догађаја не направивши притом ниједан изостанак.



Слика 4.5: Расподела броја узастопних учествовања без пауза на Meetup догађајима. Наранџастим тачкама приказана је емпиријска расподела, док пуна линија одговара степенуј расподелу.

У склопу анализе динамике учествовања на Meetup догађајима, одређена је

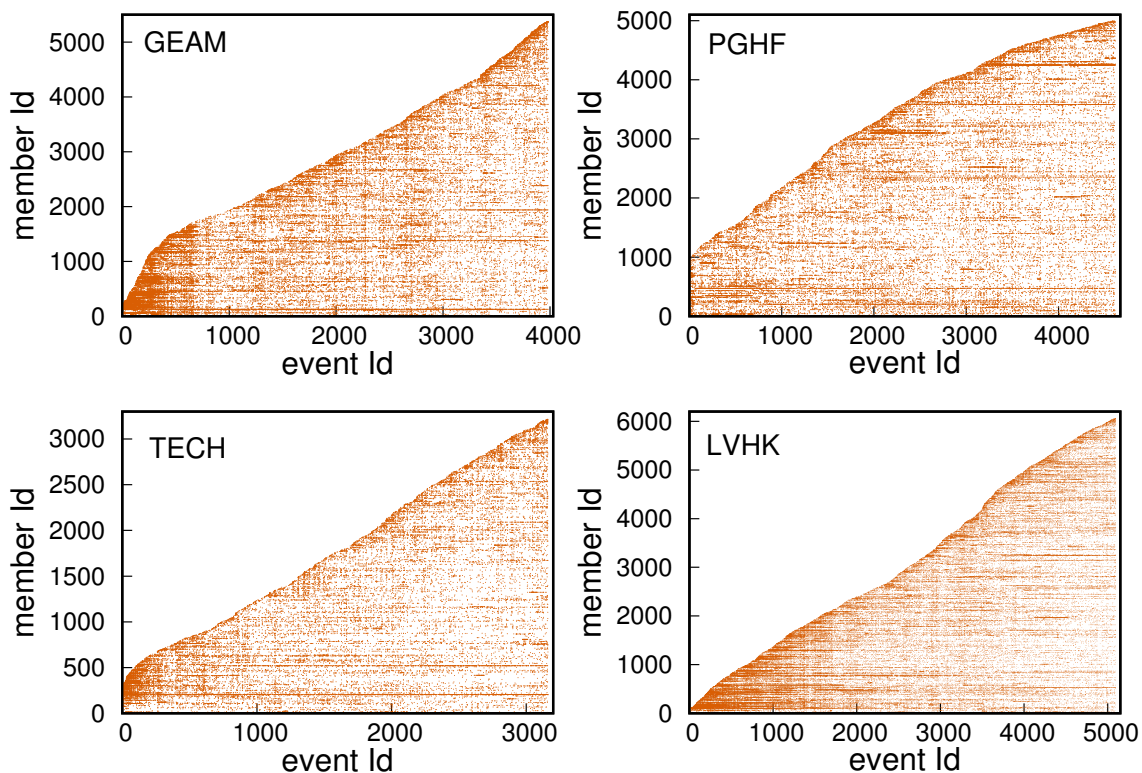


Слика 4.6: Расподела дужине паузе између два учествовања на Meetup догађајима. Наранџастим тачкама приказана је емпиријска расподела, док пуна линија одговара степеној расподели са експоненцијалним репом.

и емпиријска расподела трајања паузе између два учествовања (слика 4.6). Ова расподела има облик степене расподеле са експоненцијалним репом, као и расподеле које су одређене на основу података за конференције, али постоји значајна разлика између њихових параметара. Код Meetup догађаја вероватноћа да се појединац неће пријавити на y_s узастопних догађаја, а затим потврдити долазак на следећем догађају, опада знатно спорије него код конференција. Тако проценат пауза које подразумевају изостанак са једног догађаја износи између 10% и 20%, док је вероватноћа да ће неко изостати са неколико хиљада догађаја, а затим одлучити да поново учествује веома мала, али емпиријска расподела показује да су забележени и такви случајеви.

Да би се употпунили резултати који описују активност у Meetup групама, на слици 4.7 је дат визуелни приказ образаца учествовања. На x -оси су означени идентификатори догађаја који су им додељени на основу времена одржавања. На y -оси су дати идентификатори чланова, тако да слика илуструје активност чланова на догађајима. Са слике се може видети да постоји неравномерна рас-

подела учествовања и пауза између чланова. Највећи број чланова активан је само повремено, али постоје и изузеци који се издвајају по учесталим доласцима, што је у складу са претходним резултатима. Овде се додатно види како изгледа долазак нових чланова на догађаје и да тренутак првог појављивања на догађају у групи нема значајнији утицај на каснију активност члана.



Слика 4.7: Илустрација матрице која описује учествовање на Meetup догађајима. Свака врста ове матрице одговара једном члану Meetup групе, док су у колонама представљени догађаји по редоследу одржавања. Наранџасте тачке на слици означавају догађаје на којима је члан учествовао.

4.2 Генерализовани бинарни Појин модел

Један од најбитнијих закључака који следи из резултата који су до сада приказани јесте да ниједна од посматраних расподела не опада по експоненцијалном закону. То значи да се динамика учествовања у активностима једне групе не може посматрати као Бернулијев процес где појединац учествује на догађајима са неком константном вероватноћом (погледати поглавље [2.2.1](#)). Познати су

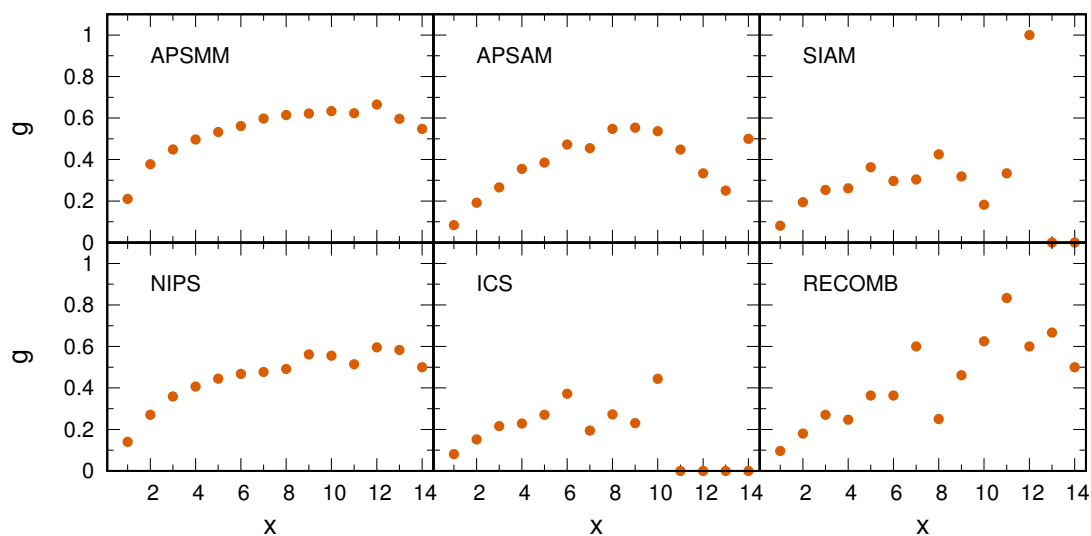
многобројни примери из реалног света где постоје изразито неједнаке расподеле код којих је одступање од средње вредности веома велико, као што су степена расподела и степена расподела са експоненцијалним репом [114], што је случај и код података који су овде посматрани. Ова појава се објашњава Матејевим ефектом, који је добио назив због стиха из јеванђеља по Матеју "ко има даће му се" и односи се на кумулативност успеха. Често се користи и изрека "богат ће постати богатији" (на енглеском "the rich get richer"). Узимајући у обзир дато објашњење, за опис активности члана у групним догађајима тестиран је генерализовани бинарни Појин модел предложен у поглављу 2.2.3, где вероватноћа учествовања расте са бројем претходних појављивања и опада ако појединац прави изостанке.

4.2.1 Научне конференције

Код научних конференција прво је проверено да ли се вероватноћа учествовања заиста мења са бројем претходних долазака и недолазака. На слици 4.8, x означава која је то конференција по реду на којој један научник учествује, а на y -оси је приказана вероватноћа да ће учествовати на конференцији следеће године. Дата вероватноћа је одређена на основу емпиријских података као однос броја научника који долазе на наредну конференцију након x долазака. Приказани резултати потврђују претпоставку да вероватноћа учествовања расте са бројем појављивања. Одступања која постоје за веће вредности x , последица су ограниченог броја учествовања које један научник може да оствари. Како расподела укупног броја учествовања за веће вредности x експоненцијално опада (слика 4.1), вредност g је одређена на основу веома малог узорка у тој области.

Показало се да број изостанака, такође, има утицај на вероватноћу појављивања. Одвојено су посматрани научници који имају 1, 2 и 3 учествовања, за које је на слици 4.9 приказано колико њих ће након n везаних изостанака направити и $n + 1$ изостанак на следећој конференцији и види се да тај проценат расте са сваким новим изостанком.

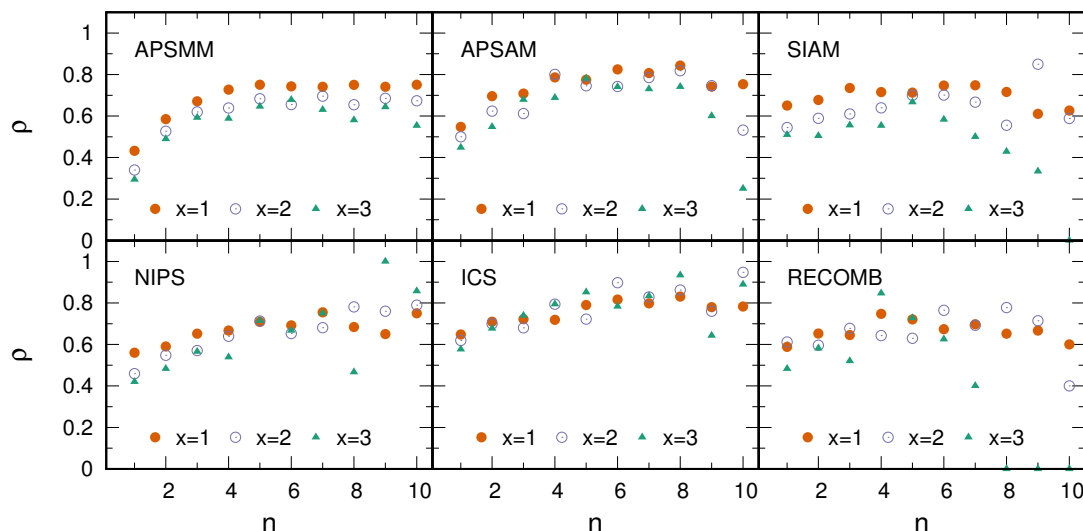
Пошто је утврђено да је претпоставка да вероватноћа појављивања на конференцији зависи од претходних долазака оправдана, размотрен је математич-



Слика 4.8: Процент научника g који након x учествовања долазе на наредну конференцију.

ки модел учествовања појединаца у колективним активностима. Подсетимо се да по моделу описаном у поглављу 2.2, начин на који су појављивања и паузе организовани у времену зависи од вероватноће прелаза између стања $p(x, t)$ и вероватноће да ће члан напустити систем r . Ако је вероватноћа прелаза дата генерализованим бинарним Појиним моделом $p(x, t) = \frac{x^a}{x^a + (y + y_0)^a}$, циљ је показати да се одговарајућим избором параметара a , y_0 и r , могу описати обрасци учешћа на научним конференцијама.

Расподела укупног броја појављивања може се одредити нумеричким решавањем једначина 2.20 и 2.22 или симулацијама учествовања великог броја појединаца, код којих се као резултат на излазу генеришу подаци о појављивању и изостанцима чланова, на основу којих се осим расподеле укупног броја појављивања, могу додатно одредити и расподела броја узастопних учествовања без пауза, као и расподела трајања паузе. Конкретно, у једној симулацији се генерише 10^5 различитих узорака, на основу којих се одређују расподеле. Сваки узорак представља низ нула и јединица, где 1 означава да је члан дошао на догађај, а 0 да се није појавио. Дужина низа једнака је броју догађаја до тренутка када је посматрани члан напустио систем. За сваког члана, као почетни услов се узима $x = 1$, $y = 0$ и $t = 1$. Симулације су извршене за различите скупове параметара (a, y_0, r) да би се одредиле оптималне вредности у односу



Слика 4.9: Процент научника ρ који су након x учествовања изостали са n наредних конференција и који ће изостати и са $n + 1$ конференције, али ће поново учествовати на некој од следећих конференција.

на емпиријске резултате са научних конференција. За сваки скуп параметара, симулација је поновљена 100 пута, а затим је израчунато Кулбак-Лејблерово растојање између расподеле одређене на основу прикупљених података са конференција и усредњене расподеле која се добије из експеримента над моделом. Вредности a , y_0 и r за које је Кулбак-Лејблерово растојање минимално изабране су као оптимални параметри.

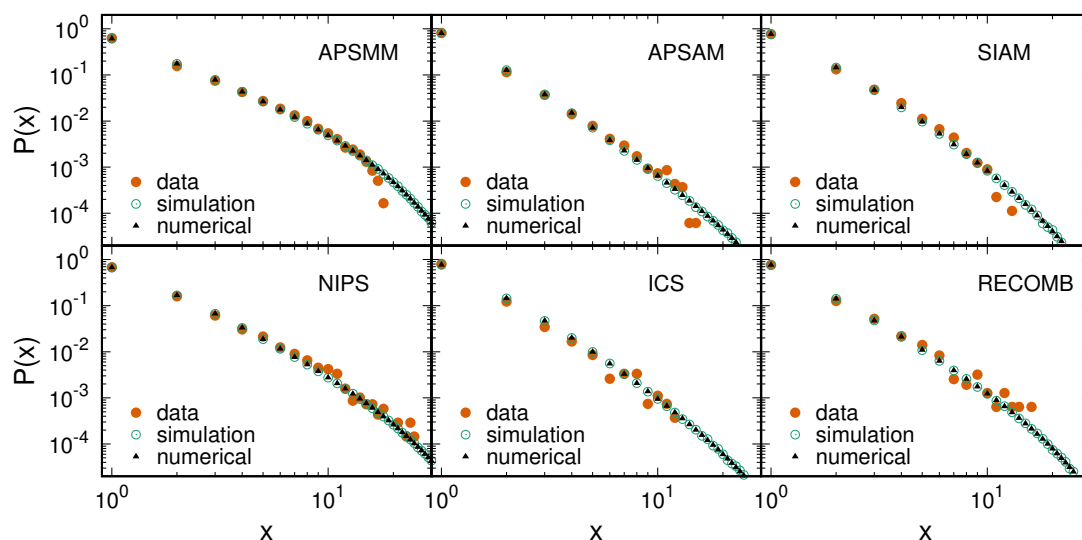
Оптималне вредности параметара за шест посматраних конференција дате су у табели 4.5. Као што се може видети, свакој конференцији одговарају различити параметри. Треба приметити да је вредност параметра $a > 1.5$, што значи да вероватноћа учествовања брзо расте када учесник редовно долази на конференцију, али исто тако брзо и опада, ако прави изостанке. Вредности параметра y_0 показују да је након првог појављивања вероватноћа доласка на следећу конференцију мала, односно да на почетку постоји препрека коју је потребно превазићи редовним учествовањем. Све ово указује на битну улогу коју има социјална компонента. Када се први пут појави на конференцији, потребно је да научник образује познанства са другим учесницима. Уколико је упоран и настави да долази на конференцију, успоставиће стабилне везе у социјалној мрежи и постаће део заједнице. Изостајањем са конференције, везе са другим

учесницима ће ослабити. У моделу је подразумевано да научник може одустати од учествовања на конференцији са неком вероватноћом r услед промене области истраживања, одлуке да напусти академску заједницу, пензионисања или неког другог разлога. На основу резултата који су се добили као оптималне вредности параметара модела, проценат научника који одустане од учествовања на конференцији сваке године је висок и износи преко 13%.

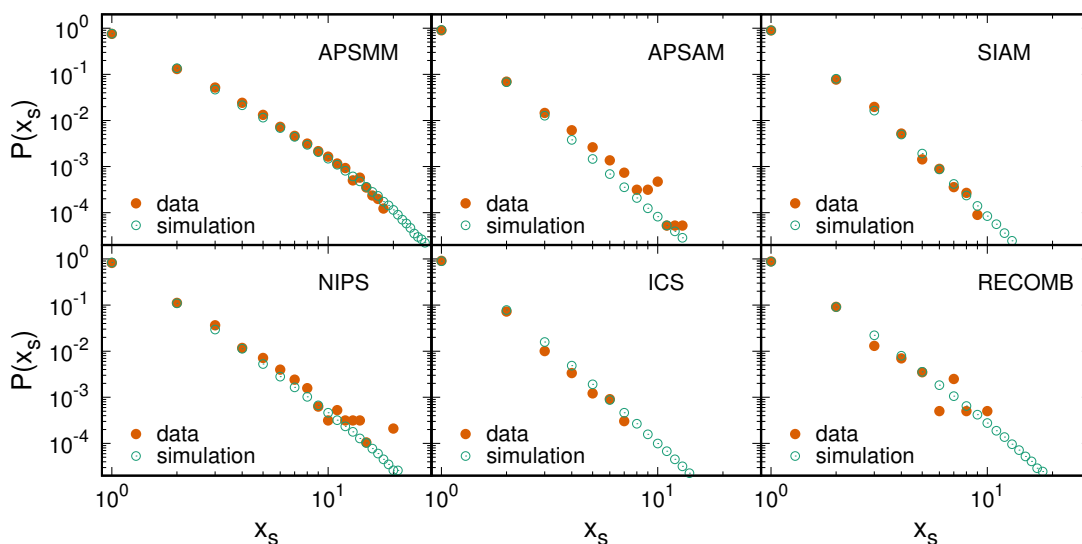
Табела 4.5: Оптималне вредности параметара модела за сваку од посматраних конференција.

Конференција	a	y_0	r
APSM	1.550	2	0.165
APSAM	1.700	4	0.135
SIAM	1.525	4	0.155
NIPS	1.525	3	0.130
ICS	1.575	4	0.135
RECOMB	1.675	3	0.175

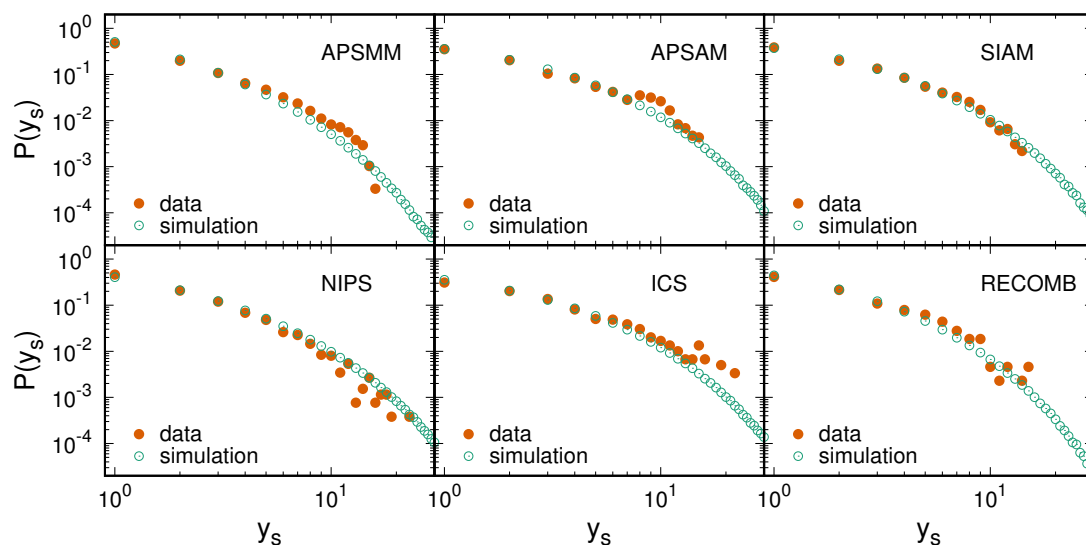
Са слике 4.10 се види да су расподела укупног броја учествовања на конференцији на основу прикупљених емпиријских података и расподела применом математичког модела са оптималним параметрима веома сличне. Исто важи и за расподелу броја узастопних учествовања без пауза, која је приказана на слици 4.11, као и за расподелу трајања паузе, дату на слици 4.12. Овим је потврђено да се одабиром одговарајућих параметара помоћу предложеног математичког модела може веродостојно описати динамика учествовања на научним конференцијама.



Слика 4.10: Расподела укупног броја учествовања на научним конференцијама. Наранџастим тачкама приказана је расподела на основу података прикупљених са конференција, зеленим кружићима на основу резултата симулација и црним троугловима су приказани резултати добијени нумеричким решавањем модела.



Слика 4.11: Расподела броја узастопних учествовања без пауза на научним конференцијама. Наранџастим тачкама приказана је расподела на основу података прикупљених са конференција, а зеленим кружићима на основу резултата симулација.



Слика 4.12: Расподела дужине паузе између два учествовања на научним конференцијама. Наранџастим тачкама приказана је расподела на основу података прикупљених са конференција, а зеленим кружићима на основу резултата симулација.

4.2.2 Meetup групе

Пре него што се приступи тестирању модела код Meetup група, потребно је направити измену која се односи на период активности неког појединца у групи. Код моделирања учествовања на конференцијама, претпостављено је да научник може одустати од даљег учествовања са неком константном вероватноћом, исто као и у [65], где је моделирано напредовање у професионалној каријери, која не мора бити искључиво научна. Како сви корисници који су регистровани на Meetup порталу имају активан статус, без обзира на то колико дуго нису узимали учешће у групним активностима, претпоставка да се чланство може завршити у неком случајном тренутку није оправдана. То додатно потврђују и резултати приказани на слици 4.6, где се може видети да постоје чланови који након учествовања праве паузе од преко 3000 догађаја, а затим поново учествују на неком догађају у групи. У моделу који је коришћен код Meetup група, претпостављено је да се на сваком догађају појави 20 нових учесника и да они остају активни на свим наредним догађајима. Вероватноћа да ће се након свог првог учествовања, члан појавити на неком следећем догађају, описана је гене-

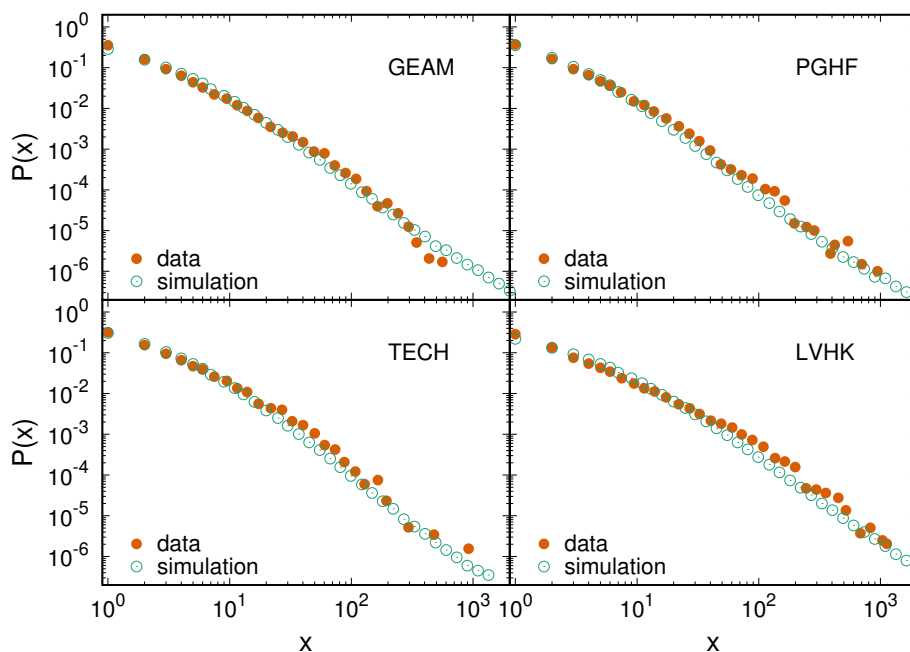
рализованим бинарним Појиним моделом. Параметри модела a и y_0 , одређени су на потпуно исти начин као и код научних конференција. У свакој симулацији посматрано је 5000 догађаја и укупно 10^5 различитих учесника. Оптималне вредности параметара, које су добијене као резултат, приказане су у табели 4.6.

Табела 4.6: Оптималне вредности параметара модела за сваку од посматраних Meetup група.

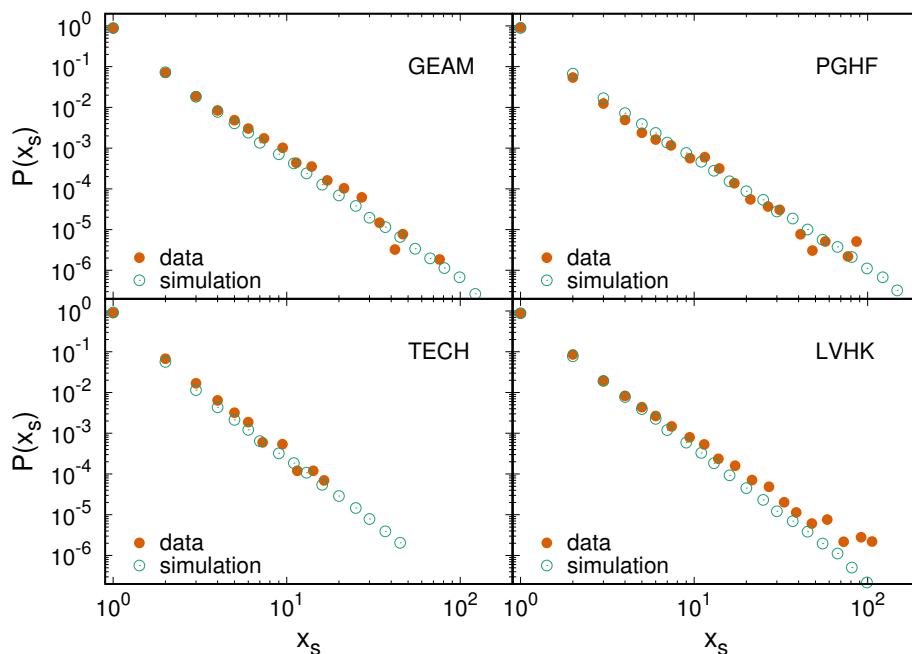
Конференција	a	y_0
GEAM	1.35	7
PGHF	1.40	7
TECH	1.35	8
LVNK	1.30	7

Вредности параметра y_0 знатно се разликују у односу на конференције, што је и очекивано с обзиром на мањи временски размак између Meetup догађаја, па је тако вероватноћа паузе након првог учествовања већа. Додатни утицај на овакве вредности параметра y_0 има и број учесника Meetup догађаја, који је мањи него код конференција, због чега је теже образовати стабилне везе у социјалној мрежи. Са друге стране, утицај изостанака на почетку на будуће учествовање је мањи, пошто параметар a има вредност између 1.3 и 1.4, за разлику од конференција код којих је $a > 1.5$, тако да код Meetup догађаја вероватноћа учествовања спорије опада када члан прави паузе и исто тако спорије расте са бројем долазака. Интересантно је да су вредности параметара модела a и y_0 за четири посматране Meetup групе веома сличне, што није случај код конференција.

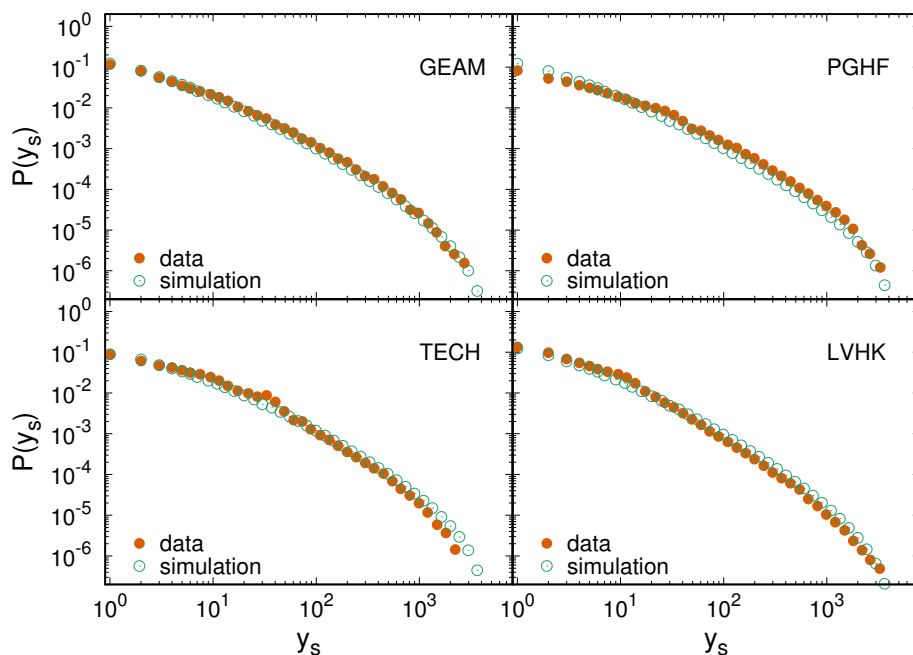
Статистичка анализа узорака који су се добили након симулације учествовања појединаца на догађајима, користећи генерализовани бинарни Појин модел са параметрима који су дати у табели 4.6, показала је да се помоћу овог модела могу објаснити обрасци учествовања на Meetup догађајима. На слици 4.13 приказана је расподела укупног броја учествовања, на 4.14 расподела броја узастопних учествовања без пауза, док се на 4.15 може видети расподела трајања паузе између два појављивања. На овим сликама, наранџастим тачкама



Слика 4.13: Расподела укупног броја учествовања на Meetup догађајима. Наранџастим тачкама приказана је расподела на основу података прикупљених са Meetup портала, а зеленим кружићима на основу резултата симулација.



Слика 4.14: Расподела броја узастопних учествовања без пауза на Meetup догађајима. Наранџастим тачкама приказана је расподела на основу података прикупљених са Meetup портала, а зеленим кружићима на основу резултата симулација.



Слика 4.15: Расподела дужине паузе између два учествовања на Meetup догађајима. Наранџастим тачкама приказана је расподела на основу података прикупљених са Meetup портала, а зеленим кружићима на основу резултата симулација.

су приказани емпиријски резултати, а зеленим круговима резултати симулација. Код све четири посматране Meetup групе, постоји добро поклапање између емпиријских резултата и модела.

4.3 Универзалне карактеристике динамике учествовања у колективним активностима

На основу резултата који су представљени у оквиру овог поглавља, могу се извести генерални закључци о обрасцима учествовања у колективним активностима и направити поређења између различитих социјалних група. И поред тога што између посматраних научних конференција, као што је већ речено у поглављу 3, постоје битне разлике као што су област истраживања, број учесника или место одржавања, емпиријска анализа је показала да њихови учесници имају сличан тренд понашања када су доласци на конференцију у питању. Расподеле броја појављивања и пауза које опадају по степеном закону до одређене

вредности када почињу да опадају експоненцијално, представљају заједничку карактеристику код свих шест конференција. Иако се параметри ових расподела разликују, оне указују на постојање механизма који узрокују изузетно хетерогену активност између научника на одређеној конференцији, која се не може објаснити тривијалним моделом, као што је Бернулијев процес. Математички модел учествовања на догађајима где се вероватноћа прелаза између стања мења у времену као у генерализованом бинарном Појином моделу са позитивном повратном спрегом, може успешно описати динамику учествовања на различитим научним конференцијама. Јасно је да се у разматрање могу узети и додатне карактеристике система и да се модел може даље унапредити, али то овде није циљ. Показало се да се помоћу посматраног модела може објаснити и учествовање на Meetup догађајима. Meetup групе представљају потпуно другачију врсту социјалног система са дискретном динамиком у односу на научне конференције. Пре свега, научне конференције подразумевају професионалана окупљања, за разлику од догађаја на којима могу да учествују чланови Meetup група. Битна разлика је и учестаност догађаја која је код конференција на годину дана, док је код Meetup група знатно већа. Без обзира на ове разлике, резултати на основу прикупљених емпиријских података, као и чињеница да предложени математички модел може успешно објаснити ове резултате, указују на универзалне механизме који се налазе у основи динамичких процеса у посматраним системима.

5. Структура социјалних мрежа под утицајем учешћа на догађајима

Примери реалних мрежа код којих структура остаје непромењена током времена су веома ретки. Код већине мрежа, постојеће везе се могу раскинути, а исто тако се између чворова који нису повезани могу успоставити нове везе. Структурне промене се не морају односити искључиво на гране, већ се и број чворова може повећавати или смањивати. Приликом проучавања мрежа чија се структура мења током времена, једна могућност је да се изабере тренутак у коме ће се посматрати структура применом метода који се користе код анализе статичке мреже, занемарујући на тај начин динамичке промене. Овакав приступ је углавном био коришћен у прошлости, док су у новијим истраживањима промене структуре узете у разматрање и посвећено је доста пажње како би се разумела њихова својства, при чему методи и мотивација проучавања могу бити различити. Требало би споменути истраживање које се односи на процену стабилности мреже приликом каскадних отказа грана и чворова проузрокованих кваровима, злонамерним нападима или неким другим разлозима [115–118], као и истраживања на тему организације структуре мреже у времену која може бити спонтана или под утицајем различитих фактора [119–122]. Недавно је уведен појам временских мрежа и дефинисани су одговарајући параметри који поред тополошких имају и додатну временску димензију [40, 123].

Мрежа која описује социјалне интеракције између појединаца је карактеристичан пример мреже чија се структура мења у времену. У социјалним мрежама које су овде проучаване, интеракције између појединаца одвијају се кроз учествовање на колективним догађајима, а самим тим и мрежа еволуира посредством колективних догађаја. Како су временска резолуција и број одржаних догађаја у изабраним Meetup групама велики, а исто тако у овим групама по-

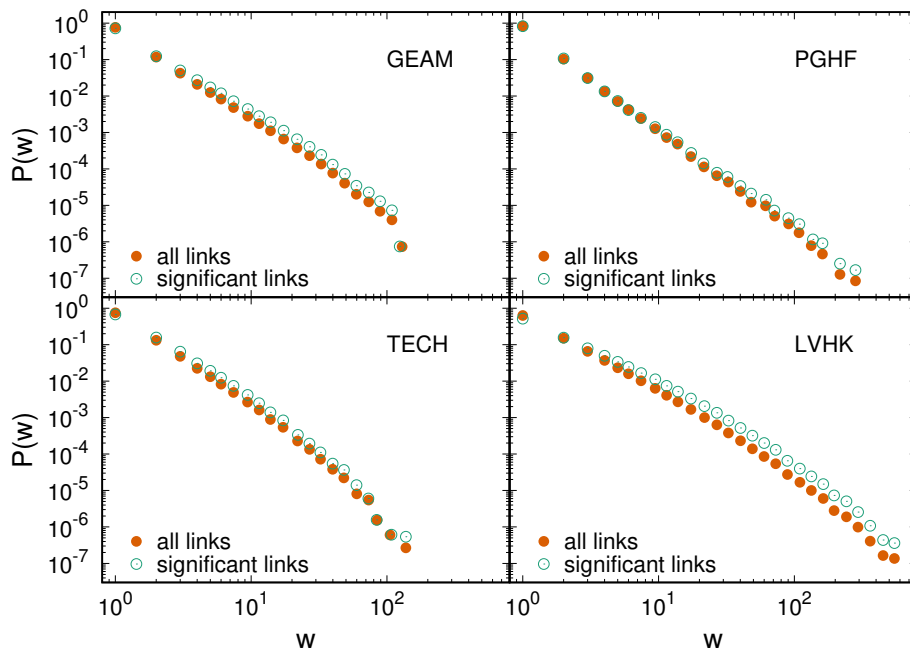
5. Структура социјалних мрежа под утицајем учешћа на догађајима

стоје подаци о члановима који су потврдили долазак на догађај, за разлику од конференција, где појављивање појединих научника као коаутора на раду не мора да значи и њихов долазак на конференцију, Meetup групе су знатно погодније за анализу од научних конференција. Због наведених разлога, у наставку овог поглавља разматрана је еволуција социјалне мреже између чланова Meetup група. Након одржавања сваког појединачног догађаја долази до промене структуре мреже. Те промене могу подразумевати: 1) долазак нових чланова који ће формирати нове везе међусобно, као и са старим члановима, 2) формирање нових веза између старих чланова који пре посматраног окупљања нису остварили међусобни контакт или 3) везе између старих чланова који су раније већ учествовали заједно на једном или више догађаја ће постати јаче. Прво је посматрано како се локални тополошки параметри појединачних чланова мењају са бројем учествовања на догађајима и дискутована је међусобна зависност структурних и динамичких карактеристика, а затим је испитан утицај који појединачни догађаји имају у организацији мреже.

5.1 Анализа Meetup социјалних мрежа

Прикупљени подаци о учествовању чланова Meetup групе у њеним активностима, не могу се директно представити у виду мреже социјалних интеракција. Идеја је да се прво конструише бипартитна мрежа чији подскупови чворова одговарају члановима и догађајима, а затим ће се као пројекција бипартитне мреже добити отежињена монопартитна мрежа у којој чворови и гране представљају чланове и интеракције између чланова. У отежињеној мрежи, два чвора ће бити повезана граном, уколико су одговарајући чланови заједно учествовали на неком од догађаја, док је тежина гране једнака укупном броју њиховог заједничког учешћа. Овако пројектована мрежа биће густо повезана. Недостатак описаног поступка за конструисање социјалне мреже на основу активности чланова Meetup групе јесте појава грана које не представљају релевантне интеракције између чланова. Као пример може се узети заједничко учествовање на великим догађајима или заједничко учествовање веома активних чланова.

5. Структура социјалних мрежа под утицајем учешћа на догађајима

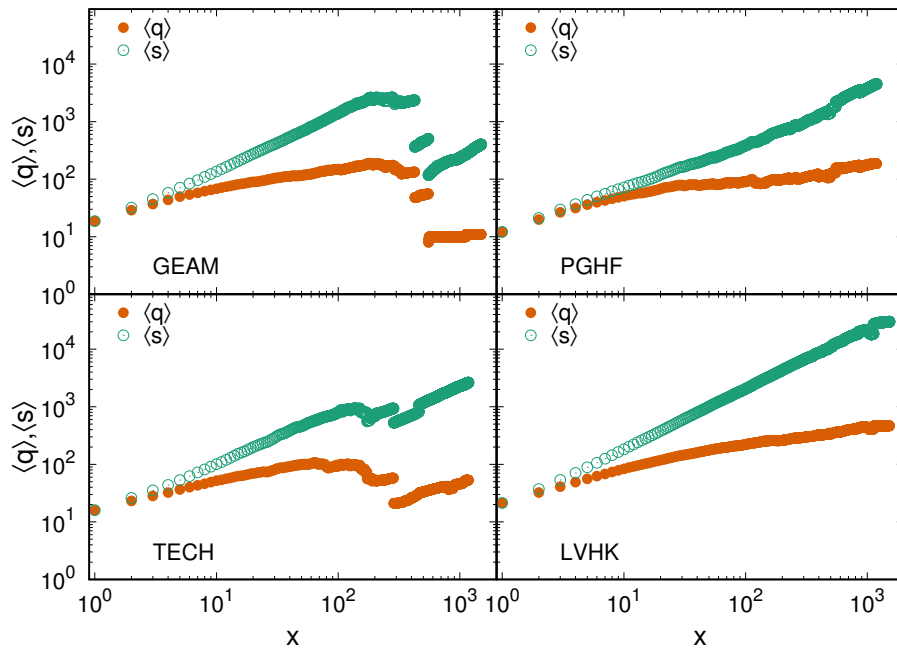


Слика 5.1: Расподела тежина грана у пројектованој монопартитној мрежи пре филтрирања (наранџасте тачке) и након филтрирања (зелени кружићи).

У наведеним примерима истовремена појава два члана на догађају може бити последица случајности и не мора указивати да је између њих постојала интеракција. Да би се издвојиле битне од случајних интеракција, мрежа је филтрирана као што је описано у поглављу 2.3.3. Сматра се да све гране чија је вероватноћа у ансамблу бипартитних мрежа са задатом расподелом степени чворова већа од $p_c = 0.05$ представљају интеракције које нису релевантне и које су последица случајности. Након филтрирања, густина мреже ће се смањити, при чему међу уклоњеним гранама већи удео чине гране са малим тежинама, што утиче на расподелу тежина грана, као што се може видети на слици 5.1, где је расподела приказана за случај пре и после филтрирања.

У оквиру истраживања како се у четири различите Meetup групе развијају структуре социјалних мрежа након појединачних догађаја, посматрана је промена степена и снаге чвора са бројем учествовања. За сваког учесника су одређени укупан број значајних интеракција и укупна снага тих интеракција након сваког догађаја на којем је учествовао, а затим су за учествовање са редним бројем x вредности степена и снаге усредњени по свим учесницима. Резултати су приказани на слици 5.2. Када се учесник први пут појави на догађају, може

5. Структура социјалних мрежа под утицајем учешћа на догађајима



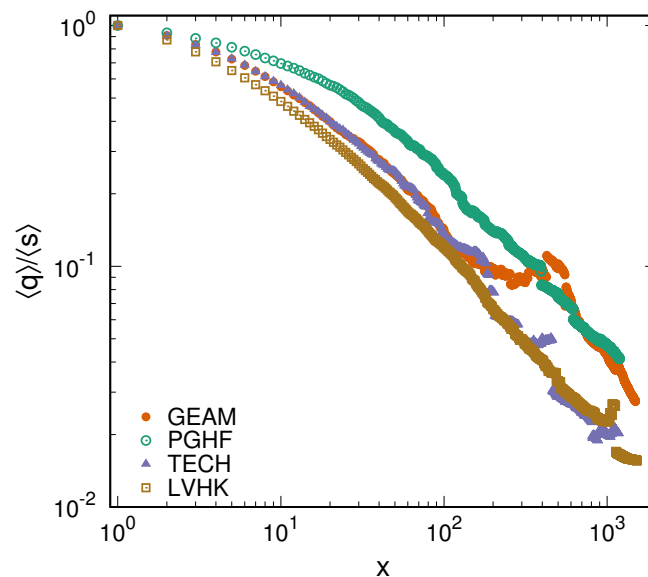
Слика 5.2: Зависност средњег степена $\langle q \rangle$ и средње снаге чвора $\langle s \rangle$ од броја учествовања x .

успоставити везе са присутним учесницима тежине 1, тако да у том тренутку степен и снага имају једнаке вредности. После само неколико учествовања снага чвора ће имати знатно већу вредност у односу на његов степен и та разлика ће наставити да се повећава са бројем учествовања код све четири посматране групе. У групама GEAM и TECH, јављају се дисконтинуитети након неколико стотина учествовања, који су последица лоше статистике, односно средња вредност се рачуна на основу малог узорка који садржи изузетно мали број чланова који имају 300 или више учествовања.

Прегледнији увид у промену односа између степена и снаге чвора са бројем долазака на догађаје може се добити са слике 5.3. У тренутку стотог учествовања на догађају, просечна вредност снаге чвора је и до десет пута већа од просечне вредности степена. Из посматраних резултата следи закључак да постоји изражен тренд понављања интеракција између чворова који су повезани граном, док је процес успостављања нових веза заступљен, али у мањој мери.

На слици 5.4 је приказана зависност између степена и средње снаге чвора у комплетној мрежи након одржавања последњег међу посматраним догађајима. Сви чворови који имају једнаке вредности степена q су груписани и одређена

5. Структура социјалних мрежа под утицајем учешћа на догађајима

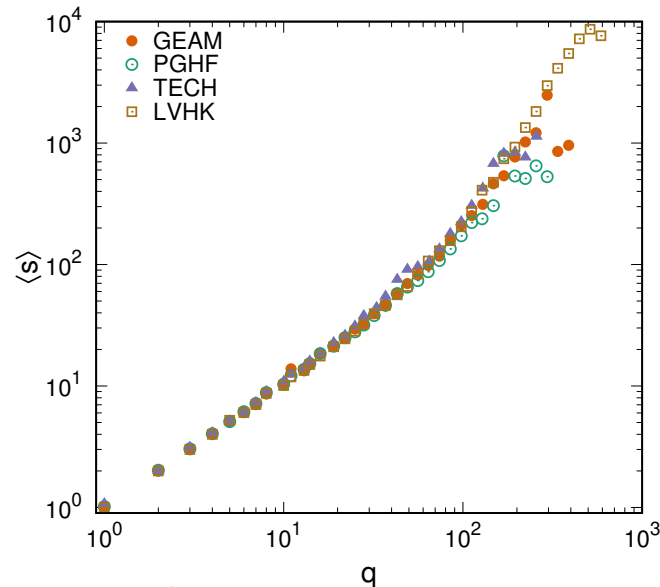


Слика 5.3: Промена односа између средњег степена $\langle q \rangle$ и средње снаге чвора $\langle s \rangle$ са бројем учествовања x .

је средња вредност снаге $\langle s \rangle$. Код чворова са малим степеном ($q < 50$), снага чвора расте линеарно, односно снага и степен чвора имају приближно једнаку вредност. Овакво понашање је карактеристично за чланове који имају једно или свега неколико учествовања на догађајима, тако да они образују мали број веза са суседима и те везе имају мале тежине. Са порастом вредности степена, снага чвора расте веома брзо, па чворови који имају највећи број успостављених веза са суседима, имају и највећу просечну вредност тежина грана. Важно је приметити да код све четири групе, постоји поклапање резултата, без обзира на разлике у величини мреже, броју одржаних догађаја, али и разлике између брзине промене односа између средњег степена и средње снаге чвора након сваког новог учествовања (слика 5.3).

Ако се ови резултати упореде са резултатима из претходног поглавља где је разматрана динамика учествовања на догађајима и утврђено да вероватноћа учествовања расте са бројем долазака, може се закључити да социјална компонента има важну улогу када је појављивање на догађајима у питању. Учествовањем на догађајима, чланови образују нове везе, али пре свега они повећавају тежине веза које су већ успостављене на претходним догађајима и на тај начин постају стабилније повезани са остатком заједнице у социјалној мрежи.

5. Структура социјалних мрежа под утицајем учешћа на догађајима



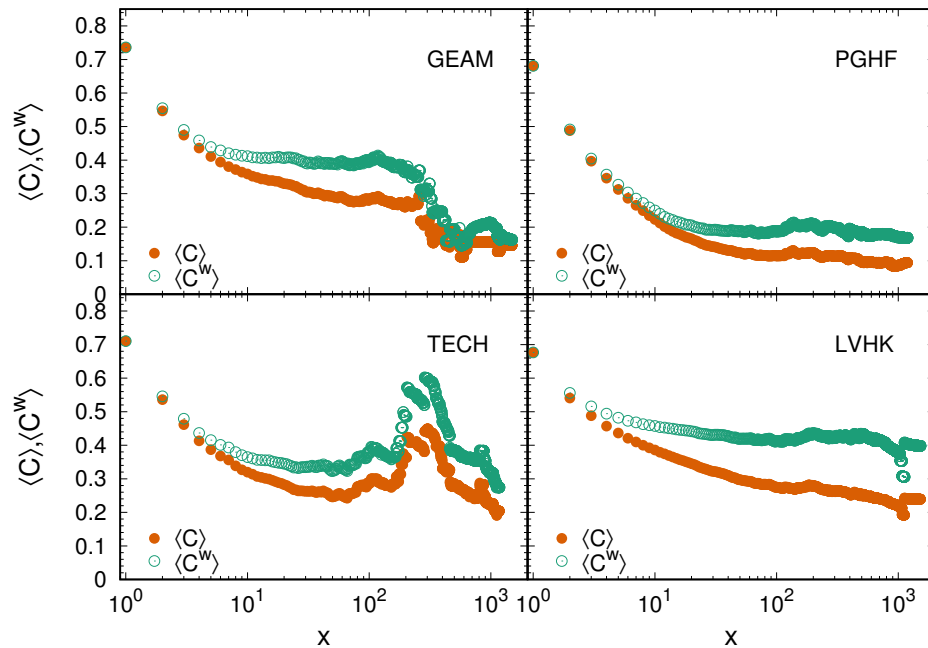
Слика 5.4: Зависност између степена q и средње снаге чвора $\langle s \rangle$ у комплетној мрежи након одржавања последњег посматраног догађаја.

После првог учествовања на догађају, везе корисника у социјалној мрежи су слабе и вероватноћа учествовања на следећим догађајима је мала. Активним учествовањем корисник ће истовремено учврстити своје везе у мрежи и повећати изгледе за долазак на следеће окупљање, при чему та одлука зависи и од одлуке чланова који су му суседи у социјалној мрежи да ли ће узети учешће на догађају.

Поред степена и снаге чвора, посматрана је и промена усредњеног коефицијента груписања $\langle C \rangle$ након сваког новог учествовања на Meetup догађају. Резултати приказани на слици 5.5 показују да је вероватноћа да су чланови са којима учесник образује гране на свом првом учествовању такође међусобно повезани већа од 60%. Ова вероватноћа ће се смањити како члан буде образовао нове везе у социјалној мрежи са члановима које ће упознати на наредним догађајима. На почетку ће средња вредност коефицијента груписања опадати брзо, али ће већ након 10 појављивања промене постати веома мале. Одступања која постоје за веће вредности x у групама GEAM и TECH, као што је већ објашњено, последица су лоше статистике.

На истој слици, зеленим кружићима приказана је и промена вредности усредњеног отежињеног коефицијента груписања $\langle C^w \rangle$, који поред податка о груписању

5. Структура социјалних мрежа под утицајем учешћа на догађајима



Слика 5.5: Зависност средњег коефицијента груписања $\langle C \rangle$ и средњег отежињеног коефицијента груписања $\langle C^w \rangle$ од броја учествовања x .

у локалне заједнице на нивоу три чвора, додатно укључује и информацију о тежини грана (погледати поглавље 2.3.1). Поређењем вредности $\langle C \rangle$ и $\langle C^w \rangle$ може се утврдити да ли се формирају заједнице између чворова чије гране имају велике тежине. У тренутку првог појављивања, вредности $\langle C \rangle$ и $\langle C^w \rangle$ су једнаке, пошто гране имају јединичне тежине. Слично као и код неотезињеног коефицијента груписања и овде ће вредност опадати са новим учествовањима и образовањем нових веза, на почетку брже, а затим спорије. Ипак, важно је приметити да отежињени коефицијент груписања има већу вредност, што показује да се локалне заједнице формирају са суседима са којима чвор има јаке везе. Ова разлика између вредности $\langle C \rangle$ и $\langle C^w \rangle$ мање је изражена током првих учествовања када члан тек успоставља везе са својим суседима у мрежи. Како број догађаја на којима се члан појавио расте, формирају се нове везе, али исто тако тежине неких од већ постојећих веза постају све веће. Тада и разлика између неотезињеног и отежињеног коефицијента груписања, такође, постаје све већа. Ово указује на битну улогу које груписање суседа са којима постоје везе великих тежина има у организацији мреже током времена.

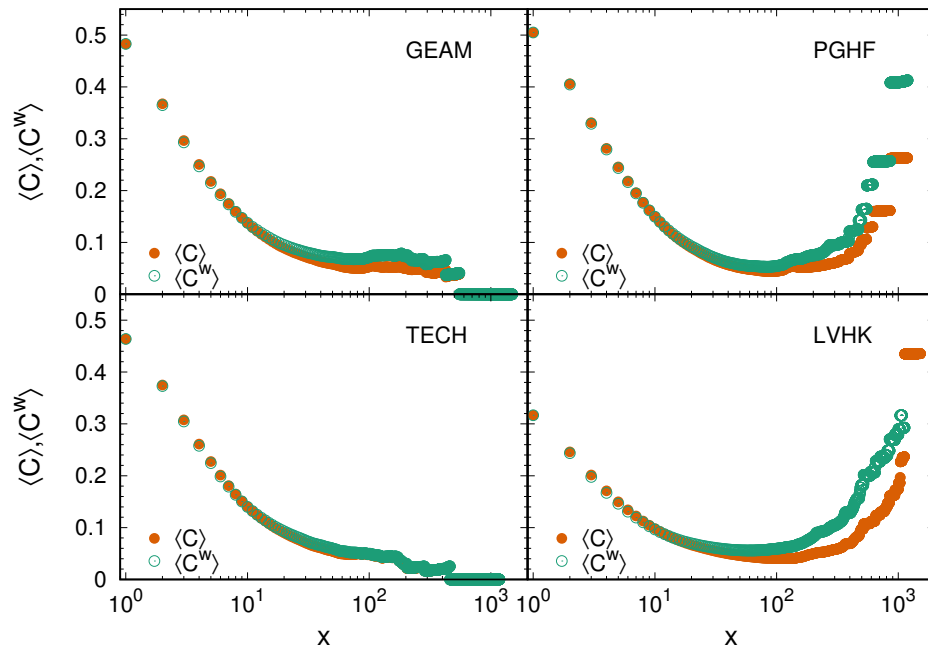
Да би се додатно истакао значај резултата који су добијени за вредности

5. Структура социјалних мрежа под утицајем учешћа на догађајима

отежињеног и неотежињеног коефицијента груписања, посматрано је како би ови параметри изгледали у случају отежињене мреже која се добија пројекцијом рандомизоване бипартитне мреже. Бипартитна мрежа је рандомизована тако што је извршено превезивање постојећих грана између чланова и догађаја, при чему су вредности степена чворова остале непромењене. Код процеса превезивања, прво се изабери на случајан начин два чвора i и j који представљају чланове Meetup групе, а затим се изабери два догађаја l^i и l^j , такође насумично, на којима су изабрани чланови учествовали. Уколико је $l^i \neq l^j$ и у бипартитној мрежи не постоји веза између чвора i и чвора l^j , као ни између чвора j и l^i , ове две везе ће бити успостављене, а постојећа веза између чворова i и l^i и веза између j и l^j , биће раскинуте. Описани поступак ће бити поновљен $10 \times$ број чланова \times број догађаја пута. Након рандомизације бипартитне мреже, пројектована отежињена мрежа имаће промењену структуру. Величина неког догађаја ће остати иста, али ће доћи до промене учесника, што ће узроковати повећање тежине појединих грана, али исто тако, долази и до смањења тежина неких других грана. Укупан збир тежина свих грана у мрежи ће остати исти. За сваку од четири Meetup групе, генерисано је 100 рандомизованих отежињених мрежа у којима је извршено филтрирање веза за које је процењено да нису релевантне.

Вредности неотежињеног и отежињеног коефицијента груписања у рандомизованој мрежи након x учествовања, прво су усредњене по члановима, а затим су усредњене за сваку од 100 рандомизованих мрежа. Резултати су приказани на слици 5.6. За разлику од оригиналне мреже, где вредност отежињеног коефицијента груписања већ након првих неколико учествовања има већу вредност у односу на неотежињени коефицијент груписања и та разлика се са сваким новим доласком на догађај све више повећава, у рандомизованој мрежи та разлика готово и да не постоји. То показује да у рандомизованој мрежи приликом груписања три чвора тежине грана немају утицај. Истовремено, додатно је потврђено да груписање чворова између којих постоје јаке везе представља једну од важних карактеристика у реалној мрежи.

5. Структура социјалних мрежа под утицајем учешћа на догађајима



Слика 5.6: Зависност средњег коефицијента груписања $\langle C \rangle$ и средњег отежињеног коефицијента груписања $\langle C^w \rangle$ од броја учествовања x у рандомизованој мрежи.

5.2 Утицај појединачних догађаја на карактеристике социјалне мреже

Еволуција структуре мреже до сада је у овој дисертацији посматрана само са становишта промене локалних структурних параметара на нивоу појединачног чвора. Подједнако интересно, али и важно је размотрити промене које се јављају на макроскопском нивоу након сваког догађаја у групи.

Поступак који је примењен у испитивању улоге коју појединачни догађаји имају у организацији социјалне мреже, подразумева анализу отежињене моно-партитне мреже која се добија пројекцијом бипартитне мреже након уклањања догађаја одговарајућим редоследом. На почетку су, за комплетну мрежу, идентификоване релевантне везе, док су оне за које је утврђено да су последица случајности занемарене. Уклањање догађаја из бипартитне мреже може узроковати смањење тежина грана или чак и њихово раскидање. Уколико између два учесника догађаја постоји релевантна веза, тежина гране ће бити умањена за један, па уколико је грана имала јединичну тежину, ова веза ће бити

5. Структура социјалних мрежа под утицајем учешћа на догађајима

раскинута.

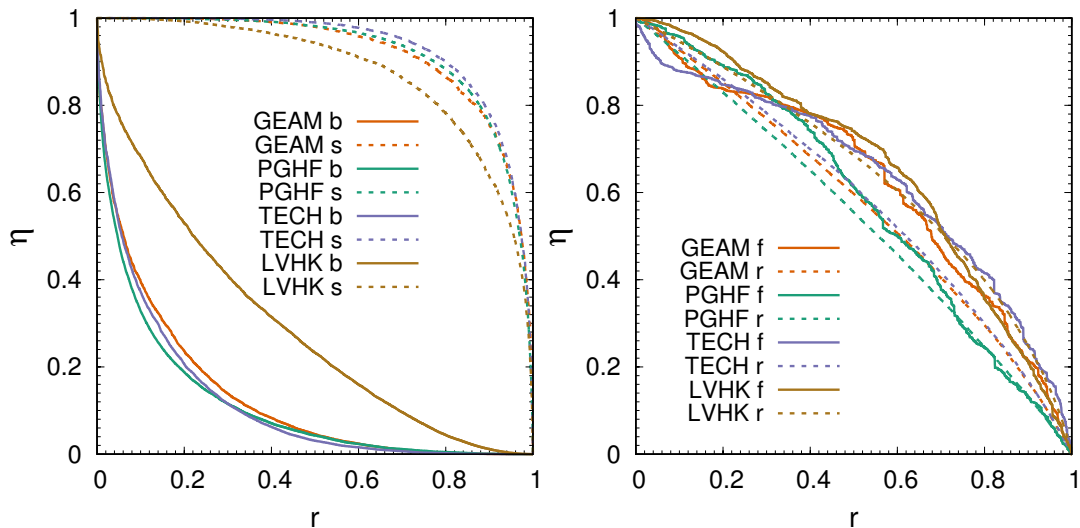
Посматрана су два параметра. Први се односи на проценат грана у мрежи η , које би постојале и да поједини догађаји нису организовани. Овај параметар је дефинисан као однос броја грана у мрежи E^r који преостане када се одстрани удео од r догађаја и укупног броја грана у комплетној мрежи E

$$\eta(r) = \frac{E^r}{E}. \quad (5.1)$$

Јасно је да расподела тежина грана значајно утиче на вредност датог параметра, пошто се гране којима су придружене мање тежине могу лакше раскинути. Поред промене густине повезаности у мрежи, посматрана је и промена средњег отежињеног коефицијента груписања \bar{C}^w до које долази након издвајања фракције r догађаја. Уколико у заједницама од три међусобно повезана чвора, постоји бар једна слаба веза, вредност отежињеног коефицијента груписања ће брзо опадати.

Два најбитнија параметра која одређују догађај су време његовог одржавања и број учесника који су потврдили свој долазак. Да би се испитао њихов утицај на структурно уређење мреже, на повезаност чворова и њихово груписање, за анализу су изабрана четири различита редоследа по којима се вршило издвајање догађаја. Прво су догађаји сортирани по величини и онда су уклањани од најбројнијег ка догађају са најмањим бројем учесника, а потом и на супротан начин, од најмањег ка највећем. Затим је поступак поновљен за редослед којим су се догађаји одржавали, од првог ка последњем. На крају је посматран случај када су догађаји који ће бити одстрањени бирани насумично све док не би преостао ни један догађај у бипартитној мрежи. Поступак насумичног уклањања поновљен је 100 пута, а затим су добијени резултати усредњени. Поређењем резултата који се добију применом различитих метода изостављања догађаја, требало би да се покаже да ли су карактеристике интеракција на великим и малим окупљањима другачије. Такође, очекује се одговор на питање да ли постоји разлика између стабилности веза које се остваре на почетним и каснијим догађајима, као и да ли је одлика поновљених заједничких учествовања

5. Структура социјалних мрежа под утицајем учешћа на догађајима

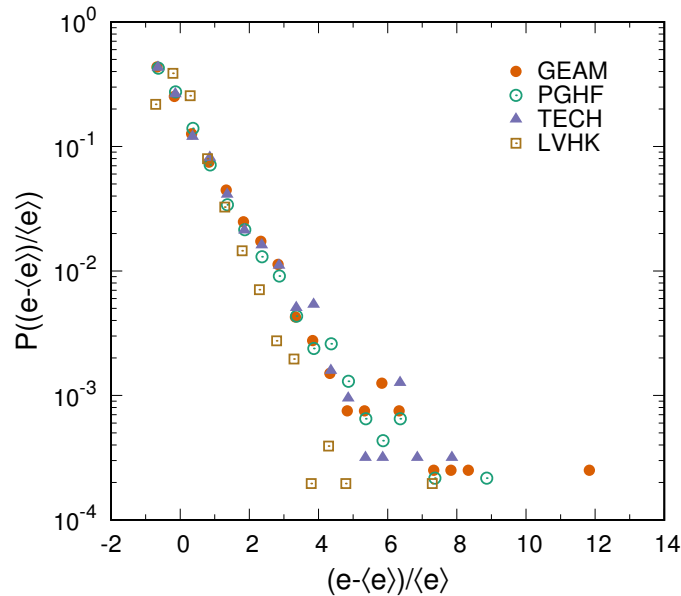


Слика 5.7: Утицај појединачних догађаја на међусобну повезаност чворова у мрежи. Слика лево: догађаји су уклањани по величини, од највећег ка најмањем (b) и обрнуто, од најмањег ка највећем (s). Слика десно: догађаји су уклањани по редоследу одржавања (f) и насумично (r).

локализованост у одређеном временском интервалу.

Резултати на слици 5.7, приказују колико веза у мрежи у односу на укупан број веза би постојало и када неки од догађаја не би били одржани. На слици лево је критеријум по којем су догађаји издвајани била њихова величина. Пуном линијом је приказано како вредност параметра η опада када се догађаји уклањају од највећег ка најмањем, а испрекидана линија одговара супротном редоследу, од најмањег ка највећем. Може се видети да је у три од четири посматране групе, довољно уклонити само 20% догађаја са највећом бројем учесника како би око 80% веза било раскинуто. Овај податак указује да код највећег броја парова чланова између којих долази до интеракција на неком од највећих догађаја, интеракције нису постојале пре посматраног догађаја и неће се поновити на догађајима који ће уследити. Са становишта структуре комплексне мреже, то значи да између два чвора која су придружена посматраним члановима постоји грана мале тежине. Са друге стране, ако се уклони 80% догађаја, бирајући прво оне са најмањим бројем учесника, проценат раскинутих веза је веома низак, око 10% у групама GEAM, PGHF и TECH и око 20% у групи LVHK. За разлику од веза које су последица повремених интеракција

5. Структура социјалних мрежа под утицајем учешћа на догађајима



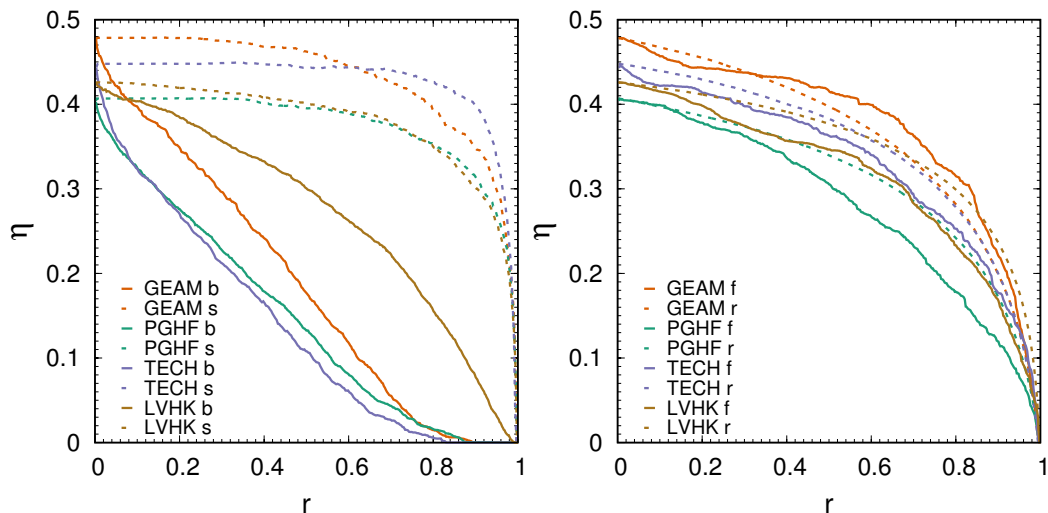
Слика 5.8: Расподела релативних флукуација величина догађаја $\frac{e-\langle e \rangle}{\langle e \rangle}$, где је са e означена величина појединачних догађаја, а са $\langle e \rangle$ просечна величина догађаја у групи.

на највећим догађајима и имају мале тежине, на мањим догађајима учествују чланови који имају велики број заједничких појављивања и њихове везе се не могу лако раскинути. Могућа објашњења за различито понашање које се јавља код групе LVHK су другачија расподела тежина грана (слика 5.1) тако да је просечна вредност већа у односу на остале три групе, али и мање варијације између величина догађаја у овој групи (слика 5.8).

На слици 5.7 десно, може се видети како се вредност параметра η смањује када се догађаји уклањају по редоследу одржавања, од првог ка последњем (пуна линија) и уколико се догађаји бирају на случајан начин (испрекидана линија). Код група PGHF и LVHK, сличност између резултата који су се добили применом ова два поступка, показује да време одржавања догађаја на којима је појединац учествовао, не утиче на стабилност његових веза. У датим групама не постоји тенденција да ће се заједничко учествовање на почетним догађајима поновити са већом вероватноћом у односу на учествовање на каснијим догађајима или обрнуто. Такође, овде се не могу идентификовати ограничени временски интервали када су учесници између којих постоје везе већих тежина активни, јер би у том случају вредност η требало да опада спорије када се редослед ели-

5. Структура социјалних мрежа под утицајем учешћа на догађајима

минације догађаја рандомизује. Код група GEAM и TECH, када се посматра хронолошки редослед, постоји благо изражени тренд да се вредност η на почетку смањује брже, а затим спорије. Овакво понашање може бити узроковано формирањем већег броја слабих веза на почетним окупљањима, које се неће учврстити на каснијим догађајима, када ће се у групи појавити нови чланови и када ће појединци тежити да успоставе стабилније везе са својим суседима.



Слика 5.9: Утицај појединачних догађаја на средњи отежињени коефицијент груписања \bar{C}^w . Слика лево: догађаји су уклањањани по величини, од највећег ка најмањем (b) и обрнуто, од најмањег ка највећем (s). Слика десно: догађаји су уклањањани по редоследу одржавања (f) и насумично (r).

Уколико се погледа промена средњег отежињеног коефицијента груписања на слици 5.9, у случају где тренутак када ће догађај бити избрисан из бипарититне мреже зависи од његове величине (слика 5.9 лево), важиће слични закључци као и када је посматран релативан број грана η . Уклањањем догађаја по растућем поретку њихових величина, вредност \bar{C}^w ће остати готово непромењена и после уклањања 80% догађаја. То додатно потврђује да се на догађајима који немају изразито велики број учесника окупљају учесници између којих постоје јаке везе у социјалној мрежи, при чему учесник ту не тежи да успостави јаке везе само са једним суседом, већ постоје заједнице од три или више чворова повезаних гранама које се не могу лако раскинути. Када се посматра обрнут редослед, где се прво бирају највећи догађаји, вредност отежињеног коефицијента груписања усредњена по свим чворовима опада спорије него што опада

5. Структура социјалних мрежа под утицајем учешћа на догађајима

број грана у мрежи. Суседи са којима чвор има интеракције на највећим догађајима не морају бити међусобно повезани, тако да раскидање великог броја грана на овим догађајима има мање последице на вредност отежињеног коефицијента груписања него што је можда очекивано. Ипак, вредност \bar{C}^w опада постепено, како се смањују тежине грана и раскидају најслабије веза у групама од три међусобно повезана чвора.

Са слике 5.9 десно, види се слично понашање када се догађаји уклањају по времену њиховог одржавања и случајним редоследом. Ипак, интересантно је да са насумичним брисањем догађаја, број грана опада спорије него када се догађаји уклањају по редоследу одржавања, док је у случају средњег отежињеног коефицијента груписања ситуација обрнута, са изузетком групе GEAM. Ова појава може указивати, да се насумичном елиминацијом догађаја раскине већи проценат слабих веза између чворова који нису груписани у заједнице са истим суседима.

6. Закључак

Приликом проучавања колективног понашања комплексног система, неопходно је узети у разматрање интеракције између његових елемената. Све до пред крај 20. века, за моделирање мреже интеракција коришћени су једноставнији модели, као што су регуларна мрежа или случајна мрежа. Крајем 20. века, све већа примена рачунара у истраживањима, омогућила је прикупљање и обраду великих количина података о карактеристикама реалних система. Емпиријском анализом података утврђено је да реалне мреже имају сложену структуру која се не може описати једноставним моделима који су до тада посматрани. Такође, анализа је показала да су поједине особине реалних мрежа универзалне у различитим системима. Све то је мотивисало различита истраживања на тему комплексних мрежа и довело до великог напретка у овој области у последње две деценије.

У овој дисертацији, разматрана је еволуција структуре комплексних мрежа у времену и динамички процеси на мрежама. Анализа је заснована на мрежама које описују реалне социјалне системе чија је динамика дискретна и у којима појединци имају непосредне интеракције на колективним догађајима. У досадашњим радовима из области физике социјалних система, својства система који подразумевају директну комуникацију између појединаца су недовољно проучена, пре свега због тежег приступа подацима него у случају комуникације посредством неког уређаја. Осим тога, специфична организација система који су овде посматрани, где у комуникацији истовремено учествује већи број чланова и то само у одређеним временским тренуцима, чини ове системе посебно интересантним за изучавање.

Својства посматраних социјалних мрежа у овој дисертацији су анализирана применом статистичких метода, метода теорије комплексних мрежа, мате-

матичког моделирања и нумеричких симулација. На почетку поглавља 2, у најкраћим цртама су представљене методе које су коришћене у статистичкој обради емпиријских података. Ту су дате дефиниције релевантних расподела, описан је метод максималне веродостојности за оцену параметара расподеле и приказана су два метода, тестирање хипотезе о расподели и тест количника веродостојности, који се користе за проверу да ли је оправдано користити посматрану теоријску расподелу за опис емпиријских података.

У посматраним системима, након првог учествовања на неком од организованих догађаја, појединац постаје део система, после чега може учествовати са или без пауза на наредним догађајима, али постоји и могућност да у неком тренутку напусти систем. У оквиру поглавља 2, размотрено је и моделирање динамике учествовања у колективним активностима. Прво је за најједноставнији случај, када је вероватноћа прелаза између стања константна, односно када је учествовање насумично, показано да расподела укупног броја учествовања има експоненцијални облик. Затим су размотрена два сложенија модела, прво када вероватноћа прелаза зависи само од тренутног стања у коме се систем налази, а затим када ова вероватноћа зависи и од стања у коме се систем налази али и од тренутка посматрања. У овом другом, временски-нехомогеном случају, акценат је стављен на генерализовани бинарни Појин модел, где вероватноћа прелаза расте са бројем учествовања и опада са паузама.

Социјални системи чија је организација заснована на учествовању чланова у колективним активностима, могу се представити помоћу мреже тако што се чланови пресликавају на чворове, а интеракције између чланова на колективним догађајима су представљене као гране између чворова којима су придружене тежине једнаке укупном броју заједничког учествовања на догађајима. Поголавље 2 је комплетирано прегледом метода теорије комплексних мрежа које су интензивно коришћене у раду. Дефинисани су параметри који се користе за опис структуре мреже и представљен је поступак за испитивање значајности линкова у отежињеној мрежи.

Било је потребно обухватити већи број социјалних група у емпиријској анализи, како би се утврдило да ли је нека карактеристика опште присутна у по-

сматраним системима или се појављује само у одређеној групи. У ту сврху, прикупљени су подаци о учесницима на шест научних конференција из различитих научних области током периода од три деценије и подаци о члановима четири различите Meetup групе и њиховом учествовању на догађајима. На порталу Meetup, сваком догађају у групи и сваком члану додељени су јединствени идентификатори, па није било потребе за додатним чишћењем података, за разлику од научних конференција где је поступак прикупљања и чишћења података био знатно сложенији. Код конференција су подаци о учесницима добијени из зборника радова, где су научници током година користили различите начине потписивања. Додатни проблем се јављао код конференција са великим бројем учесника на којима су били присутни различити научници са истим именом. Како би учесници конференција били једнозначно идентификовани, поред списка коаутора на пријављеним радовима, искоришћене су и њихове афилијације. Поступак прикупљања, чишћења и форматирања података детаљно је приказан у поглављу 3.

На основу резултата емпиријске анализе образаца учествовања на догађајима, који су приказани у поглављу 4, показано је да учествовање чланова није насумично и да постоје слични трендови понашања у различитим социјалним групама које су овде посматране. Затим је тестиран теоријски модел који обухвата основне механизме који утичу на понашање појединаца када је учествовање у колективним активностима у питању. Показано је да је применом генерализованог бинарног Појиног модела и одабиром одговарајућих параметара могуће успешно описати обрасце понашања у различитим групама без обзира на састав њихових чланова, врсту окупљања или на учестаност догађаја, што указује на присуство универзалних принципа понашања. Модел је тестиран применом нумеричких симулација, при чему су параметри модела одређени на основу Кулбак-Лејблеровог растојања између емпиријске расподеле и расподеле која се добија симулацијом модела. Вредности оптималних параметара се разликују између различитих социјалних група, с тим да је вредност експонента у генерализованом бинарном Појином моделу увек већа од један. Ово представља веома интересантан резултат, који показује да активно учествовање

у колективним активностима утиче на повећање вероватноће доласка на будуће догађаје, али исто тако ова вероватноћа брзо опада уколико је појединац неактиван. Стога, може се закључити да је изразита хетерогеност међу члановима када је активност на догађајима у питању узрокована дејством позитивне повратне спреге.

Да би се спровела квантитативна анализа еволуције социјалне мреже у времену, бипартитна мрежа, у којој дисјунктни подскупови одговарају члановима и догађајима у једној социјалној групи, пројектована је на отежињену монопартитну мрежу. Приликом тог поступака процењено је које интеракције на догађајима су последица случајности и оне су занемарене у даљој анализи. За податке са портала Meetup, у поглављу 5 конструисане су монопартитне мреже са релевантним везама применом метода за филтрирање описаног у поглављу 2. Испитивањем промене локалних структурних параметара чворова са бројем учесћа на догађајима, показано је да постоји тенденција заједничког учествовања на догађајима чланова између којих постоје везе у социјалној мрежи, док је образовање нових веза од мањег интереса. Овај резултат се међусобно допуњује са моделом предложеним у поглављу 4 и показује зависност која постоји између структуре мреже и социјалне динамике. Закључак који следи је да појединац активним учествовањем на догађајима тежи да учврсти везе са суседима, а са друге стране, што су везе са суседима у социјалној мрежи стабилније, вероватноћа појављивања на следећем догађају расте. Додатно је показано да у зависности од броја учесника, догађаји имају различит утицај на структуру мреже. На великим догађајима доминантне су интеракције између чланова који нису повезани или су њихове гране малих тежина. Са друге стране, ако на догађају учествује свега неколико појединаца, карактеристично је да између њих постоје јаке везе.

Методологија представљена у овом раду може се искористити за проучавање ширег скупа проблема везаних за организацију комплексних мрежа са дискретном динамиком. У наставку истраживања требало би више пажње посветити детекцији заједница у оваквим мрежама и развијању теоријског модела који би објаснио како долази до њиховог формирања. Било би интересантно за

проучавање посматрати и промене величине социјалне групе у времену. Том приликом, требало би испитати корелације између структуре социјалне мреже, учестаности догађаја и појаве нових чланова у систему.

Додатак А

А.1 Аналитичко решавање модела у случају Бернулијевог процеса

Вероватноћа да ће у тренутку t члан учествовати на x догађаја је

$$f(x, t) = \binom{t}{x} \mu^x (1 - \mu)^{t-x}, \quad t \geq x. \quad (\text{A.1})$$

Члан може напустити систем у било ком тренутку са вероватноћом r . Вероватноћа да ће посматрани члан имати укупно x учествовања може се одредити једноставно у неколико корака

$$\begin{aligned} f(x) &= \sum_{T=x}^{\infty} f(x, T) (1-r)^T r \\ &= r \left(\frac{\mu}{1-\mu} \right)^x \sum_{T=x}^{\infty} \binom{T}{x} [(1-\mu)(1-r)]^T \\ &= r \left(\frac{\mu}{1-\mu} \right)^x \frac{[(1-\mu)(1-r)]^x}{[1-(1-\mu)(1-r)]^{x+1}} \\ &= \frac{r}{r+\mu-r\mu} \left[\frac{\mu(1-r)}{r+\mu-r\mu} \right]^x \\ &= \left(1 - \frac{\mu-r\mu}{r+\mu-r\mu} \right) e^{x \ln \frac{\mu-r\mu}{r+\mu-r\mu}}. \end{aligned} \quad (\text{A.2})$$

А.2 Аналитичко решавање модела у случају хомогеног Марковског ланца

Да би систем у тренутку $t \geq 1$ био у стању x , $x \leq t + 1$, потребно је да дође до $x - 1$ прелаза. Ако се искористи овај закључак, може се написати

$$f(x, t) = p(1)p(2) \dots p(x - 1)F(x, t). \quad (\text{A.3})$$

Заменом израза [A.3](#) у једначини

$$f(x + 1, t + 1) = p(x)f(x, t) + q(x + 1)f(x + 1, t) \quad (\text{A.4})$$

добија се

$$F(x + 1, t + 1) = F(x, t) + q(x + 1)F(x + 1, t). \quad (\text{A.5})$$

Решавање ове једначине може се поједноставити, као што је предложено у [\[124\]](#), ако се уместо функције $F(x, t)$ посматра њена карактеристична функција $\varphi(x, s)$ која се дефинише са

$$\varphi(x, s) = \sum_{t=x-1}^{\infty} F(x, t)s^t. \quad (\text{A.6})$$

Прво је потребно помножити леву и десну страну једначине [A.5](#) са s^t , а затим сумирати по $t \geq x - 1$

$$\sum_{t=x-1}^{\infty} F(x + 1, t + 1)s^t = \sum_{t=x-1}^{\infty} F(x, t)s^t + q(x + 1) \sum_{t=x-1}^{\infty} F(x + 1, t)s^t, \quad (\text{A.7})$$

из чега даље следи

$$\frac{1}{s}\varphi(x + 1, s) = \varphi(x, s) + q(x + 1)\varphi(x + 1, s). \quad (\text{A.8})$$

Како је

$$\varphi(1, s) = \sum_{t=0}^{\infty} q(1)^t s^t = \frac{1}{1 - q(1)s}, \quad (\text{A.9})$$

из рекурзивне формуле

$$\varphi(x+1, s) = \frac{s}{1 - q(x+1)s} \varphi(x, s) \quad (\text{A.10})$$

се добија

$$\varphi(x, s) = \frac{s^{x-1}}{(1 - q(x)s) \dots (1 - q(1)s)}. \quad (\text{A.11})$$

Применом метода неодређених коефицијената

$$\prod_{n=1}^N \frac{1}{1 - a_n s} = \sum_{n=1}^N \frac{A_n}{1 - a_n s}, \quad (\text{A.12})$$

где је

$$A_n = \frac{1}{(1 - \frac{a_1}{a_n})(1 - \frac{a_2}{a_n}) \dots (1 - \frac{a_{n-1}}{a_n})(1 - \frac{a_{n+1}}{a_n}) \dots (1 - \frac{a_N}{a_n})} = \prod_{m=1}^N \frac{1 - a_n/a}{1 - a_m/a} \Big|_{a=a_n}, \quad (\text{A.13})$$

једначина A.11 се може развити у

$$\varphi(x, s) = s^{x-1} \sum_{n=1}^x \frac{1}{(1 - \frac{q(1)}{q(n)}) \dots (1 - \frac{q(n-1)}{q(n)})(1 - \frac{q(n+1)}{q(n)}) \dots (1 - \frac{q(x)}{q(n)})} \frac{1}{1 - q(n)s}, \quad (\text{A.14})$$

одакле следи

$$\varphi(x, s) = \sum_{n=1}^x s^{x-1} q(n)^{x-1} Q(n) \sum_{\tau=0}^{\infty} q(n)^{\tau} s^{\tau}, \quad (\text{A.15})$$

где је

$$Q(n) = \frac{1}{(q(n) - q(1)) \dots (q(n) - q(n-1))(q(n) - q(n+1)) \dots (q(n) - q(x))}. \quad (\text{A.16})$$

Овде је згодно увести смену $t = \tau + x - 1$, да би се из

$$\varphi(x, s) = \sum_{t=x-1}^{\infty} \sum_{n=1}^x Q(n) q(n)^t s^t \quad (\text{A.17})$$

прво добило решење

$$F(x, t) = \sum_{n=1}^x Q(n) q(n)^t, \quad (\text{A.18})$$

а затим и вероватноћа

$$f(x, t) = p(1)p(2) \dots p(n-1) \sum_{n=1}^x Q(n)q(n)^t. \quad (\text{A.19})$$

Као што је већ познато, вероватноћа да ће учесник напустити систем је константна. У тренутку када напусти систем, вероватноћа да ће забележити учествовање на x догађаја је

$$\begin{aligned} f(x) &= \sum_{T=x-1}^{\infty} f(x, T)(1-r)^T r \\ &= r(1-q(1)) \dots (1-q(x-1)) \sum_{n=1}^x Q(n) \sum_{T=x-1}^{\infty} (1-r)^T q(n)^T \\ &= r(1-q(1)) \dots (1-q(x-1)) \sum_{n=1}^x Q(n) \frac{(1-r)^{x-1} q(n)^{x-1}}{1-(1-r)q(n)} \\ &= r(1-r)^{x-1} \frac{(1-q(1)) \dots (1-q(x-1))}{(1-(1-r)q(1)) \dots (1-(1-r)q(x))}. \end{aligned} \quad (\text{A.20})$$

На крају, вероватноћа $f(x)$ се може изразити у зависности од r и од p

$$f(x) = \frac{r}{1-r} \frac{p(1) \dots p(x-1)}{\left(\frac{r}{1-r} + p(1)\right) \dots \left(\frac{r}{1-r} + p(x)\right)}. \quad (\text{A.21})$$

Литература

- [1] H. E. Stanley R. N. Mantegna. *Introduction to Econophysics: Correlations and Complexity in Finance*. Cambridge University Press, 1999. [1](#)
- [2] J. Ross and A. P. Arkin. Complex systems: From chemistry to systems biology. *Proc. Natl. Acad. Sci. USA*, 106:6433, 2009. [1](#)
- [3] D. Sherrington. Physics and complexity. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 368:1175, 2010. [1](#)
- [4] P. Sen and B. K. Chakrabarti. *Sociophysics: an introduction*. Oxford University Press, 2013. [1](#), [7](#)
- [5] J. Kwapien and S. Drożdż. Physical approach to complex systems. *Physics Reports*, 515:115, 2012. [1](#)
- [6] M. San Miguel, J. H. Johnson, J. Kertesz, K. Kaski, A. Díaz-Guilera, R. S. MacKay, V. Loreto, P. Érdi, and D. Helbing. Challenges in complex systems science. *Eur. Phys. J. Special Topics*, 214:245, 2012. [1](#)
- [7] W.-X. Wang, Y.-C. Lai, and C. Grebogi. Data based identification and prediction of nonlinear and complex dynamical systems. *Physics Reports*, 644:1, 2016. [1](#)
- [8] S. N. Dorogovtsev and J. F. F. Mendes. *Evolution of networks: From biological nets to the Internet and WWW*. Oxford University Press, 2003. [2](#), [3](#), [22](#)
- [9] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang. Complex networks: Structure and dynamics. *Physics Reports*, 424:175, 2006. [2](#), [3](#), [4](#), [22](#)

- [10] M. Zanin, D. Papo, P.A. Sousa, E. Menasalvas, A. Nicchi, E. Kubik, and S. Boccaletti. Combining complex networks and data mining: Why and how. *Physics Reports*, 635:1, 2016. [3](#)
- [11] J. Živković, M. Mitrović, and B. Tadić. Correlation patterns in gene expressions along the cell cycle of yeast. In *Complex Networks: Results of the 2009 International Workshop on Complex Networks (CompleNet 2009)*, page 23, 2009. [3](#)
- [12] N. Dianati. Unwinding the hairball graph: Pruning algorithms for weighted complex networks. *Phys. Rev. E*, 93:012304, 2016. [3](#)
- [13] F. Saracco, M. J. Straka, R. Di Clemente, A. Gabrielli, G. Caldarelli, and T. Squartini. Inferring monopartite projections of bipartite networks: an entropy-based approach. *New Journal of Physics*, 19:053022, 2017. [3](#), [29](#), [30](#), [31](#)
- [14] D. J. Watts and S. H. Strogatz. Collective dynamics of ‘small-world’ networks. *Nature*, 393:440, 1998. [3](#), [4](#), [5](#), [26](#)
- [15] A. L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286:509, 1999. [4](#), [5](#), [24](#)
- [16] A. L. Barabási. Scale-free networks: A decade and beyond. *Science*, 325:412, 2009. [4](#)
- [17] K. Klemm and V. M. Eguíluz. Growing scale-free networks with small-world behavior. *Phys. Rev. E*, 65:057102, 2002. [4](#)
- [18] M. Kaiser and C. C. Hilgetag. Modelling the development of cortical systems networks. *Neurocomputing*, 58–60:297, 2004. [4](#)
- [19] L. Li, D. Alderson, W. Willinger, and J. Doyle. A first-principles approach to understanding the internet’s router-level topology. *SIGCOMM Comput. Commun. Rev.*, 34:3, 2004. [4](#)

-
- [20] M. Catanzaro, M. Bogu ná, and R. Pastor-Satorras. Generation of uncorrelated random scale-free networks. *Phys. Rev. E*, 71:027103, 2005. [4](#)
- [21] R. Guimerà, B. Uzzi, J. Spiro, and L. A. N. Amaral. Team assembly mechanisms determine collaboration network structure and team performance. *Science*, 308:697, 2005. [4](#), [7](#)
- [22] V. Nicosia, P. E. Vértés, W. R. Schafer, V. Latora, and E. T. Bullmore. Phase transition in the economically modeled growth of a cellular nervous system. *Proc. Natl. Acad. Sci. USA*, 110:7880, 2013. [4](#)
- [23] C. Orsini, M. M. Dankulov, P. Colomer de Simón, A. Jamakovic, P. Mahadevan, A. Vahdat, K. E. Bassler, Z. Toroczkai, M. Boguñá, G. Caldarelli, S. Fortunato, and D. Krioukov. Quantifying randomness in real networks. *Nature Communications*, 6:8627, 2015. [4](#)
- [24] A. Barrat, M. Barthélemy, and A. Vespignani. *Dynamical processes on complex networks*. Cambridge University Press, 2008. [4](#)
- [25] R. Pastor-Satorras and A. Vespignani. Epidemic spreading in scale-free networks. *Phys. Rev. Lett.*, 86:3200, 2001. [4](#)
- [26] J. Balthrop, S. Forrest, M. E. J. Newman, and M. M. Williamson. Technological networks and the spread of computer viruses. *Science*, 304:527, 2004. [4](#)
- [27] V. Colizza, A. Barrat, M. Barthélemy, and A. Vespignani. The role of the airline transportation network in the prediction and predictability of global epidemics. *Proc. Natl. Acad. Sci. USA*, 103:2015, 2006. [4](#)
- [28] D. Chakrabarti, Y. Wang, C. Wang, J. Leskovec, and C. Faloutsos. Epidemic thresholds in real networks. *ACM Trans. Inf. Syst. Secur.*, 10:1, 2008. [4](#)
- [29] A. V. Goltsev, S. N. Dorogovtsev, J. G. Oliveira, and J. F. F. Mendes. Localization and spreading of diseases in complex networks. *Phys. Rev. Lett.*, 109:128702, 2012. [4](#)

-
- [30] D. Brockmann and D. Helbing. The hidden geometry of complex, network-driven contagion phenomena. *Science*, 342:1337, 2013. [4](#)
- [31] J. Gómez-Gardeñes, M. Campillo, L. M. Floría, and Y. Moreno. Dynamical organization of cooperation in complex topologies. *Phys. Rev. Lett.*, 98:108103, 2007. [4](#)
- [32] A. Pusch, S. Weber, and M. Porto. Impact of topology on the dynamical organization of cooperation in the prisoner’s dilemma game. *Phys. Rev. E*, 77:036120, 2008. [4](#)
- [33] J. Grujić, T. Röhl, D. Semmann, M. Milinski, and A. Traulsen. Consistent strategy updating in spatial and non-spatial behavioral experiments does not promote cooperation in social networks. *PLOS ONE*, 7:1, 2012. [4](#)
- [34] G. Yan, T. Zhou, B. Hu, Z.-Q. Fu, and B.-H. Wang. Efficient routing on complex networks. *Phys. Rev. E*, 73:046108, 2006. [4](#)
- [35] J. Smiljanić and I. Stanković. Efficient routing on small complex networks without buffers. *Phys. A.*, 392:2294, 2013. [4](#)
- [36] T. Nishikawa, A. E. Motter, Y.-C. Lai, and F. C. Hoppensteadt. Heterogeneity in oscillator networks: Are smaller worlds easier to synchronize? *Phys. Rev. Lett.*, 91:014101, 2003. [4](#)
- [37] A. Arenas, A. Díaz-Guilera, J. Kurths, Y. Moreno, and C. Zhou. Synchronization in complex networks. *Physics Reports*, 469:93, 2008. [4](#)
- [38] D. Garlaschelli and M. I. Loffredo. Wealth dynamics on complex networks. *Phys. A*, 338:113, 2004. [4](#)
- [39] T. Ichinomiya. Bouchaud-mézard model on a random network. *Phys. Rev. E*, 86:036111, 2012. [4](#)
- [40] P. Holme and J. Saramäki. Temporal networks. *Physics Reports*, 519:97, 2012. [4](#), [65](#)

-
- [41] S. Boccaletti, G. Bianconi, R. Criado, C.I. del Genio, J. Gómez-Gardeñes, M. Romance, I. Sendiña-Nadal, Z. Wang, and M. Zanin. The structure and dynamics of multilayer networks. *Physics Reports*, 544:1, 2014. [4](#)
- [42] S. M. Rinaldi, J. P. Peerenboom, and T. K. Kelly. Identifying, understanding, and analyzing critical infrastructure interdependencies. *IEEE Control Systems*, 21:11, 2001. [5](#)
- [43] V. Rosato, L. Issacharoff, F. Tiritico, S. Meloni, S. Porcellinis, and R. Setola. Modelling interdependent infrastructures using interacting dynamical models. *International Journal of Critical Infrastructures*, 4:63, 2008. [5](#)
- [44] L. F. Sessions. How offline gatherings affect online communities. *Information, Communication & Society*, 13:375, 2010. [5](#)
- [45] S. Fortunato. Community detection in graphs. *Physics Reports*, 486:75, 2010. [6](#)
- [46] R. S. Burt. The network structure of social capital. *Research in organizational behavior*, 22:345, 2000. [6](#)
- [47] J.-P. Onnela, J. Saramäki, J. Hyvönen, G. Szabó, D. Lazer, K. Kaski, J. Kertész, and A.-L. Barabási. Structure and tie strengths in mobile communication networks. *Proc. Natl. Acad. Sci. USA*, 104:7332, 2007. [6](#)
- [48] G. Kossinets and D. J. Watts. Origins of homophily in an evolving social network. *American Journal of Sociology*, 115:405, 2009. [6](#)
- [49] P. Ball. The physical modelling of human social systems. *ComplexUs*, 1:190, 2003. [6](#), [7](#)
- [50] W.-S. Jung, F. Wang, and H. E. Stanley. Gravity model in the korean highway. *Europhys. Lett.*, 81:48005, 2008. [6](#)
- [51] V. Palchykov, M. Mitrović, H. H. Jo, J. Saramäki, and R. K. Pan. Inferring human mobility using communication patterns. *Scientific Reports*, 4:6174, 2013. [6](#)

-
- [52] L. F. Henderson. The statistics of crowd fluids. *Nature*, 229:381, 1975. [6](#)
- [53] P. Holme and M. E. J. Newman. Nonequilibrium phase transition in the coevolution of networks and opinions. *Phys. Rev. E*, 74:056108, 2006. [7](#)
- [54] C. Castellano, S. Fortunato, and V. Loreto. Statistical physics of social dynamics. *Rev. Mod. Phys.*, 81:591, 2009. [7](#), [22](#)
- [55] Y.-R. Lin, Y. Chi, S. Zhu, H. Sundaram, and B. L. Tseng. 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, page 685, 2008. [7](#)
- [56] S. Aral and D. Walker. Identifying influential and susceptible members of social networks. *Science*, 337:337, 2012. [7](#)
- [57] A. Garas, D. Garcia, M. Skowron, and F. Schweitzer. Emotional persistence in online chatting communities. *Scientific Reports*, 2:402, 2012. [7](#)
- [58] T. Yasseri, R. Sumi, A. Rung, A. Kornai, and J. Kertész. Dynamics of conflicts in wikipedia. *PLoS ONE*, 7:1, 2012. [7](#)
- [59] S. González-Bailón, J. Borge-Holthoefer, and Y. Moreno. Broadcasters and hidden influentials in online protest diffusion. *Am. Behav. Sci.*, 57:943, 2013. [7](#)
- [60] J. Török, G. I niguez, T. Yasseri, M. San Miguel, K. Kaski, and J. Kertész. Opinions, conflicts, and consensus: Modeling social dynamics in a collaborative environment. *Phys. Rev. Lett.*, 110:088701, 2013. [7](#)
- [61] A. Clauset, C. R. Shalizi, and M. E. J. Newman. Power-law distributions in empirical data. *SIAM Review*, 51:661, 2009. [12](#), [13](#), [14](#)
- [62] M. Merkle. *Verovatnoća i statistika*. Akademska misao, 2006. [12](#)
- [63] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes in C. The Art of Scientific Computing*. Cambridge University Press, 1992. [13](#)

-
- [64] K. P. Burnham and D. R. Anderson. *Model Selection and Multimodel Inference*. Springer-Verlag, New York, 2002. [16](#)
- [65] A. M. Petersen, W.-S. Jung, J.-S. Yang, and H. E. Stanley. Quantitative and empirical demonstration of the matthew effect in a study of career longevity. *Proc. Natl. Acad. Sci. USA*, 108:18, 2011. [18](#), [19](#), [60](#)
- [66] N. L. Johnson and S. Kotz. *Urn Models and Their Application: An Approach to Modern Discrete Probability Theory*. John Wiley & Sons Inc, 1977. [21](#)
- [67] E. Drinea, A. Frieze, and M. Mitzenmacher. Balls and bins models with feedback. In *Proceedings of the Thirteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '02, page 308, 2002. [21](#), [22](#)
- [68] F. Chung, S. Handjani, and D. Jungreis. Generalizations of polya's urn problem. *Annals of combinatorics*, 7:141, 2003. [21](#), [22](#)
- [69] R. Pemantle. A survey of random processes with reinforcement. *Probab. Surveys*, 4:1, 2007. [21](#)
- [70] R. I. Oliveira. The onset of dominance in balls-in-bins processes with feedback. *Random Structures & Algorithms*, 34:454, 2009. [21](#)
- [71] J. Smiljanić, A. Chatterjee, T. Kauppinen, and M. Mitrović Dankulov. A theoretical model for the associative nature of conference participation. *PLOS ONE*, 11:1, 2016. [21](#), [33](#)
- [72] G. Kossinets and D. J. Watts. Empirical analysis of an evolving social network. *Science*, 311:88, 2006. [22](#)
- [73] A.L. Barabási, H. Jeong, Z. Néda, E. Ravasz, A. Schubert, and T. Vicsek. Evolution of the social network of scientific collaborations. *Phys. A*, 311:590, 2002. [22](#)
- [74] D. Cvetković and M. Milić. *Teorija grafova i njene primene*. BIGZ, 1971. [22](#)
- [75] M. Newman. The structure and function of complex networks. *SIAM Review*, 45:167, 2003. [22](#)

-
- [76] R. Pastor-Satorras and A. Vespignani. *Evolution and Structure of the Internet: A Statistical Physics Approach*. Cambridge University Press, 2004. [22](#)
- [77] R. Cohen and S. Havlin. *Complex Networks: Structure, Robustness and Function*. Cambridge University Press, 2010. [22](#)
- [78] P. Erdős and A. Rényi. On random graphs. *Publicationes Mathematicae Debrecen*, 6:290, 1959. [24](#)
- [79] P. Erdős and A. Rényi. On the evolution of random graphs. In *Publication of the Mathematical Institute of the Hungarian Academy of Sciences*, page 17, 1960. [24](#)
- [80] M. Faloutsos, P. Faloutsos, and C. Faloutsos. On power-law relationships of the internet topology. *SIGCOMM Comput. Commun. Rev.*, 29:251, 1999. [24](#)
- [81] H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, and A. L. Barabási. The large-scale organization of metabolic networks. *Nature*, 407:651, 2000. [24](#)
- [82] H. Jeong, S. P. Mason, A. L. Barabási, and Z. N. Oltvai. Lethality and centrality in protein networks. *Nature*, 411:41, 2001. [24](#)
- [83] D. Chakrabarti, Y. Zhan, and C. Faloutsos. R-mat: A recursive model for graph mining. In *Proceedings of the 2004 SIAM International Conference on Data Mining*, page 442, 2004. [24](#)
- [84] R. Cohen and S. Havlin. Scale-free networks are ultrasmall. *Phys. Rev. Lett.*, 90:058701, 2003. [25](#)
- [85] L. A. N. Amaral, A. Scala, M. Barthélémy, and H. E. Stanley. Classes of small-world networks. *Proc. Natl. Acad. Sci. USA*, 97:11149, 2000. [25](#)
- [86] R. Guimerà, S. Mossa, A. Turtschi, and L. A. N. Amaral. The worldwide air transportation network: Anomalous centrality, community structure, and cities' global roles. *Proc. Natl. Acad. Sci. USA*, 102:7794, 2005. [25](#)

-
- [87] A. Wagner and D. A. Fell. The small world inside large metabolic networks. *Proceedings of the Royal Society of London B: Biological Sciences*, 268:1803, 2001. [26](#)
- [88] J. M. Montoya and R. V. Solé. Small world patterns in food webs. *Journal of Theoretical Biology*, 214:405, 2002. [26](#)
- [89] O. Shefi, I. Golding, R. Segev, E. Ben-Jacob, and A. Ayali. Morphological characterization of in vitro neuronal networks. *Phys. Rev. E*, 66:021905, 2002. [26](#)
- [90] E. Ravasz and A. L. Barabási. Hierarchical organization in complex networks. *Phys. Rev. E*, 67:026112, 2003. [26](#)
- [91] M. Á. Serrano and M. Bogu ná. Topology of the world trade web. *Phys. Rev. E*, 68:015101, 2003. [26](#)
- [92] A. Barrat, M. Barthélemy, R. Pastor-Satorras, and A. Vespignani. The architecture of complex weighted networks. *Proc Natl Acad Sci USA*, 101:3747, 2004. [27](#)
- [93] M. Barthélemy, A. Barrat, R. Pastor-Satorras, and A. Vespignani. Characterization and modeling of weighted networks. *Phys. A.*, 346:34, 2005. [27](#)
- [94] K.-I. Goh, M. E. Cusick, D. Valle, B. Childs, M. Vidal, and A.-L. Barabási. The human disease network. *Proc Natl Acad Sci USA*, 104:8685, 2007. [28](#)
- [95] M.-S. Shang, L. Lü, Y.-C. Zhang, and T. Zhou. Empirical analysis of web-based user-object bipartite networks. *Europhys. Lett.*, 90:48006, 2010. [28](#)
- [96] M. Mitrović, G. Paltoglou, and B. Tadić. Quantitative analysis of bloggers' collective behavior powered by emotions. *J. Stat. Mech.*, 2011:P02005, 2011. [28](#)

-
- [97] 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. [28](#)
- [98] M. M. Dankulov, R. Melnik, and B. Tadić. The dynamics of meaningful social interactions and the emergence of collective knowledge. *Scientific Reports*, 5:12197, 2015. [28](#)
- [99] F. Saracco, R. Di Clemente, A. Gabrielli, and T. Squartini. Randomizing bipartite networks: the case of the world trade web. *Scientific Reports*, 5:10595, 2015. [28](#), [29](#)
- [100] M. E. J. Newman. Scientific collaboration networks. i. network construction and fundamental results. *Phys. Rev. E*, 64:016131, 2001. [28](#)
- [101] M. E. J. Newman. Scientific collaboration networks. ii. shortest paths, weighted networks, and centrality. *Phys. Rev. E*, 64:016132, 2001. [28](#)
- [102] J. Smiljanić and M. Mitrović Dankulov. Associative nature of event participation dynamics: A network theory approach. *PLOS ONE*, 12:1, 2017. [28](#), [33](#)
- [103] N. Dianati. A maximum entropy approach to separating noise from signal in bimodal affiliation networks. *ArXiv e-prints*, 2016. [29](#), [30](#), [31](#)
- [104] D. Cellai and G. Bianconi. Multiplex networks with heterogeneous activities of the nodes. *Phys. Rev. E*, 93:032302, 2016. [29](#)
- [105] Z. Neal. The backbone of bipartite projections: Inferring relationships from co-authorship, co-sponsorship, co-attendance and other co-behaviors. *Soc. Netw.*, 39:84, 2014. [30](#)
- [106] J. Park and M. E. J. Newman. Statistical mechanics of networks. *Phys. Rev. E*, 70:066117, 2004. [30](#)
- [107] J. Liebig and A. Rao. Fast extraction of the backbone of projected bipartite networks to aid community detection. *Europhys. Lett.*, 113:28003, 2016. [31](#)

-
- [108] Y. Hong. On computing the distribution function for the poisson binomial distribution. *Computational Statistics and Data Analysis*, 59:41, 2013. [31](#)
- [109] C. Cattuto, W. Van den Broeck, A. Barrat, V. Colizza, J.-F. Pinton, and A. Vespignani. Dynamics of person-to-person interactions from distributed rfid sensor networks. *PLOS ONE*, 5:1–9, 2010. [33](#)
- [110] T. Takaguchi, M. Nakamura, N. Sato, K. Yano, and N. Masuda. Predictability of conversation partners. *Phys. Rev. X*, 1:011008, 2011. [33](#)
- [111] M. Génois, C. Vestergaard, C. Cattuto, and A. Barrat. Compensating for population sampling in simulations of epidemic spread on temporal contact networks. *Nature Communications*, 8:8860, 2015. [33](#)
- [112] J. Wu and X.-H. Ding. Author name disambiguation in scientific collaboration and mobility cases. *Scientometrics*, 96:683, 2013. [37](#)
- [113] J. C. Huber. Inventive productivity and the statistics of exceedances. *Scientometrics*, 45:33, 1999. [39](#)
- [114] M. Perc. The matthew effect in empirical data. *J. R. Soc. Interface*, 11(98), 2014. [55](#)
- [115] R. Cohen, K. Erez, D. ben Avraham, and S. Havlin. Resilience of the internet to random breakdowns. *Phys. Rev. Lett.*, 85:4626, 2000. [65](#)
- [116] A. E. Motter and Y.-C. Lai. Cascade-based attacks on complex networks. *Phys. Rev. E*, 66:065102, 2002. [65](#)
- [117] R. Parshani, S. V. Buldyrev, and S. Havlin. Critical effect of dependency groups on the function of networks. *Proc. Natl. Acad. Sci. USA*, 108:1007, 2011. [65](#)
- [118] J. Zhao, D. Li, H. Sanhedrai, R. Cohen, and S. Havlin. Spatio-temporal propagation of cascading overload failures in spatially embedded networks. *Nature Communications*, 7:10094, 2016. [65](#)

- [119] N. Kashtan and U. Alon. Spontaneous evolution of modularity and network motifs. *Proc. Natl. Acad. Sci. USA*, 102:13773–13778, 2005. [65](#)
- [120] J. M. Pacheco, A. Traulsen, and M. A. Nowak. Coevolution of strategy and structure in complex networks with dynamical linking. *Phys. Rev. Lett.*, 97:258103, 2006. [65](#)
- [121] T. Gross and B. Blasius. Adaptive coevolutionary networks: a review. *J. R. Soc. Interface*. [65](#)
- [122] D. S. Bassett, N. F. Wymbs, M. A. Porter, P. J. Mucha, J. M. Carlson, and S. T. Grafton. Dynamic reconfiguration of human brain networks during learning. *Proc. Natl. Acad. Sci. USA*, 108:7641, 2011. [65](#)
- [123] P. Holme and J. Saramäki. *Temporal Networks*. Springer-Verlag, Berlin, Heidelberg, 2013. [65](#)
- [124] M. A. Woodbury. On a probability distribution. *Ann. Math. Statist.*, 20:311, 1949. [85](#)

Списак објављених радова

[1] **J. Smiljanić** and M. M. Dankulov, Associative nature of event participation dynamics: A network theory approach, *PLoS ONE* **12**, e0171565 (2017).

[2] **J. Smiljanić**, A. Chatterjee, T. Kauppinen and M. M. Dankulov, A Theoretical Model for the Associative Nature of Conference Participation, *PLoS ONE* **11**, e0148528 (2016).

[3] **J. Smiljanić**, M. Žeželj, V. Milanović, J. Radovanović, and I. Stanković, MATLAB-based program for optimization of quantum cascade laser active region parameters and calculation of output characteristics in magnetic field, *Comput. Phys. Commun.* **185**, 998 (2014).

[4] I. Stanković, M. Žeželj, **J. Smiljanić**, A. Belić, Modelling of disaster spreading dynamics, in *High-Performance Computing Infrastructure for South East Europe's Research Communities, Springer Book Series on Modeling and Optimization in Science and Technologies* **2**, 31 (2014).

[5] **J. Smiljanić** and I. E. Stanković, Efficient routing in small complex networks without buffers, *Physica A* **392**, 2294 (2013).

Биографија

Јелена Смиљанић је рођена 12.04.1987. у Лазаревцу. 2006. године је завршила Гимназију у Лазаревцу. Исте године уписала је основне студије на Електротехничком факултету Универзитета у Београду. Дипломирала је на смеру Телекомуникације и информационе технологије октобра 2010. године са просечном оценом 9.04. Мастер студије је завршила у септембру 2011. године на истом факултету, смер Системско инжењерство и радио комуникације, са просечном оценом 9.83. Јелена Смиљанић је 2011. године започела докторске студије на смеру Наноелектроника и фотоника на Електротехничком факултету Универзитета у Београду. У оквиру докторских студија положила је све испите предвиђене наставним планом и програмом са просечном оценом 10.

Од децембра 2011. године до новембра 2012. године је као стипендиста Министарства просвете, науке и технолошког развоја Републике Србије ангажована у Лабораторији за примену рачунара у науци Института за физику у Београду на пројекту основних истраживања ОН171017 “Моделовање и нумеричке симулације сложених вишечестичних система”. У новембру 2012. године Јелена Смиљанић је запослена на Институту за физику у Београду као истраживач приправник у Лабораторији за примену рачунара у науци на истом пројекту. Звање истраживач сарадник стекла је у мају 2013. године.

Јелена Смиљанић је аутор четири рада у међународним часописима са СЦИ листе, од којих је један рад објављен у међународном часопису изузетних вредности категорије М21а, два рада су објављена у врхунском међународном часопису категорије М21 и један рад у истакнутом међународном часопису категорије М22. До сада је учествовала на више међународних и националних конференција.

Изјава о ауторству

Име и презиме аутора: **Јелена Смиљанић**

Број индекса: **5012/2011**

Изјављујем

да је докторска дисертација под насловом

Испитивање својстава комплексних мрежа са дискретном динамиком

- резултат сопственог истраживачког рада;
- да дисертација у целини ни у деловима није била предложена за стицање друге дипломе према студијским програмима других високошколских установа;
- да су резултати коректно наведени и
- да нисам кршио/ла ауторска права и користио/ла интелектуалну својину других лица.

Потпис аутора

У Београду, 20.06.2017.

Јелена Смиљанић

Изјава о истоветности штампане и електронске верзије докторског рада

Име и презиме аутора: **Јелена Смиљанић**

Број индекса: **5012/2011**

Студијски програм: **Наноелектроника и фотоника**

Наслов рада: **Испитивање својстава комплексних мрежа са дискретном динамиком**

Ментор: **проф. др Јелена Радовановић**

Изјављујем да је штампана верзија мог докторског рада истоветна електронској верзији коју сам предао/ла ради похрањена у **Дигиталном репозиторијуму Универзитета у Београду**.

Дозвољавам да се објаве моји лични подаци везани за добијање академског назива доктора наука, као што су име и презиме, година и место рођења и датум одбране рада.

Ови лични подаци могу се објавити на мрежним страницама дигиталне библиотеке, у електронском каталогу и у публикацијама Универзитета у Београду.

Потпис аутора

У Београду, 20.06.2017.

Јелена Смиљанић

Изјава о коришћењу

Овлашћујем Универзитетску библиотеку „Светозар Марковић“ да у Дигитални репозиторијум Универзитета у Београду унесе моју докторску дисертацију под насловом:

Моделовање и оптимизација транспортних процеса у савременим наноелектронским уређајима

која је моје ауторско дело.

Дисертацију са свим прилозима предао/ла сам у електронском формату погодном за трајно архивирање.

Моју докторску дисертацију похрањену у Дигиталном репозиторијуму Универзитета у Београду и доступну у отвореном приступу могу да користе сви који поштују одредбе садржане у одабраном типу лиценце Креативне заједнице (Creative Commons) за коју сам се одлучио/ла.

1. Ауторство (CC BY)
2. Ауторство – некомерцијално (CC BY-NC)
3. Ауторство – некомерцијално – без прерада (CC BY-NC-ND)
4. Ауторство – некомерцијално – делити под истим условима (CC BY-NC-SA)
5. Ауторство – без прерада (CC BY-ND)
6. Ауторство – делити под истим условима (CC BY-SA)

(Молимо да заокружите само једну од шест понуђених лиценци.
Кратак опис лиценци је саставни део ове изјаве).

Потпис аутора

У Београду, 20.06.2017.

Јелена Смиљанић

1. **Ауторство.** Дозвољава се умножавање, дистрибуцију и јавно саопштавање дела, и прераде, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце, чак и у комерцијалне сврхе. Ово је најслободнија од свих лиценци.

2. **Ауторство – некомерцијално.** Дозвољава се умножавање, дистрибуцију и јавно саопштавање дела, и прераде, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце. Ова лиценца не дозвољава комерцијалну употребу дела.

3. **Ауторство – некомерцијално – без прерада.** Дозвољава се умножавање, дистрибуцију и јавно саопштавање дела, без промена, преобликовања или употребе дела у свом делу, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце. Ова лиценца не дозвољава комерцијалну употребу дела. У односу на све остале лиценце, овом лиценцом се ограничава највећи обим права коришћења дела.

4. **Ауторство – некомерцијално – делити под истим условима.** Дозвољава се умножавање, дистрибуцију и јавно саопштавање дела, и прераде, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце и ако се прерада дистрибуира под истом или сличном лиценцом. Ова лиценца не дозвољава комерцијалну употребу дела и прерада.

5. **Ауторство – без прерада.** Дозвољава се умножавање, дистрибуцију и јавно саопштавање дела, без промена, преобликовања или употребе дела у свом делу, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце. Ова лиценца дозвољава комерцијалну употребу дела.

6. **Ауторство – делити под истим условима.** Дозвољава се умножавање, дистрибуцију и јавно саопштавање дела, и прераде, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце и ако се прерада дистрибуира под истом или сличном лиценцом. Ова лиценца дозвољава комерцијалну употребу дела и прерада. Слична је софтверским лиценцама, односно лиценцама отвореног кода.

RESEARCH ARTICLE

A Theoretical Model for the Associative Nature of Conference Participation

Jelena Smiljanić^{1,2}, Arnab Chatterjee³, Tomi Kauppinen⁴, Marija Mitrović Dankulov^{1*}

1 Scientific Computing Laboratory, Institute of Physics Belgrade, University of Belgrade, Pregrevica 118, 11080 Belgrade, Serbia, **2** School of Electrical Engineering, University of Belgrade, P.O. Box 35-54, 11120 Belgrade, Serbia, **3** Condensed Matter Physics Division, Saha Institute of Nuclear Physics, 1/AF Bidhannagar, Kolkata 700064, India, **4** Aalto University School of Science, P.O. Box 11000, FI-00076 AALTO, Finland

* mitrovic@ipb.ac.rs



OPEN ACCESS

Citation: Smiljanić J, Chatterjee A, Kauppinen T, Mitrović Dankulov M (2016) A Theoretical Model for the Associative Nature of Conference Participation. PLoS ONE 11(2): e0148528. doi:10.1371/journal.pone.0148528

Editor: Matjaz Perc, University of Maribor, SLOVENIA

Received: November 26, 2015

Accepted: December 8, 2015

Published: February 9, 2016

Copyright: © 2016 Smiljanić et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data files are available from https://figshare.com/articles/Conference_Datasets/2066907.

Funding: This work was supported by ON171017, The Ministry of Education, Science, and Technological Development of the Republic of Serbia (<http://www.mpn.gov.rs/>: J.S. and M.M.D.); and 675121, The European Commission (http://ec.europa.eu/index_en.htm: M.M.D.).

Competing Interests: The authors have declared that no competing interests exist.

Abstract

Participation in conferences is an important part of every scientific career. Conferences provide an opportunity for a fast dissemination of latest results, discussion and exchange of ideas, and broadening of scientists' collaboration network. The decision to participate in a conference depends on several factors like the location, cost, popularity of keynote speakers, and the scientist's association with the community. Here we discuss and formulate the problem of discovering how a scientist's previous participation affects her/his future participations in the same conference series. We develop a stochastic model to examine scientists' participation patterns in conferences and compare our model with data from six conferences across various scientific fields and communities. Our model shows that the probability for a scientist to participate in a given conference series strongly depends on the balance between the number of participations and non-participations during his/her early connections with the community. An active participation in a conference series strengthens the scientist's association with that particular conference community and thus increases the probability of future participations.

Introduction

Social data at a large scale is nowadays available over the internet. Researchers are making the best use of these data to find trends, statistics and patterns, which sometime reveal as robust features, similar to 'laws' in natural science. In recent years, a huge community of researchers [1] including mathematicians, statisticians, computer scientists, theoretical physicists, sociologists, economists, financial analysts, geographers, anthropologists, and biologists of various sub-disciplines have contributed to a larger, developing field, commonly known as 'computational social science' [2]. Empirical data, after a rigorous analysis produces information that is of immense interest for theoreticians. Statistical mechanics, which has been proved to be versatile in modeling phenomena across different areas of physics, and beyond, seems to be the most desired tool even for the above emerging discipline [3, 4].

The abundance of a new data about scientific activities such as publications, collaborations, and citations led to the emergence of a new interdisciplinary field of research about science and how science works [5]. These studies provide insights about the impact of scientists and their publications [6–8], authors' reputation and scientific success [9], patterns of collaboration and their impact on authors' reputation [10, 11], the role of cumulative advantage in career longevity [12, 13] and scientific mobility [14] among many other things. Despite the attention given to publication records and citation patterns, another integral part of modern science, scientific meetings, have so far been largely overlooked. This negligence is particularly interesting, given the pervasive role of the meetings in scientific disciplines. Scientific meetings provide arenas for a fast dissemination of the latest results, exchange and evaluation of ideas as well as a knowledge extension. However, the most important function of scientific meetings is to facilitate social contacts. They provide an opportunity and platform to extend the network of collaborators through the creation of new contacts, and to strengthen existing links by getting acquainted with old friends.

Undoubtedly, conference participation has a very positive impact on scientific career. In addition to the opportunities they provide, attending a scientific meeting can be very costly, both in terms of time and money. Bearing in mind that the number of national and international meetings have drastically increased in the last few decades, it is clear that scientists are now pressed to make a careful selection of the meetings they will attend. Extensive studies [15–17] have shown that conference characteristics, such as the attractiveness and the reachability of the location or the choice of keynote speakers affect the decision of scientists to attend a meeting. The role of the social component in conference choice is so far unexplored, mainly due to lack of quality data. The social component, such as the association with a conference community or conference inclusiveness, are of crucial importance when it comes to whether a conference participation was beneficial or not. This is particularly evident in the case of young scientists, who are new to a community and struggle to overcome the social obstacle of an initial contact [18, 19]. One of the rare studies on conference participation [20] has shown that conferences have a stable core of regularly attending participants, regardless of the conference location and distance. Having in mind that characteristics like the attractiveness of a location and the quality of keynote speakers are fluctuating from one year to another, it is clear that social component of a conference strongly influence the scientists decision to attend the conference and their long-term participation patterns, accordingly.

The association with a conference community and conference inclusiveness, can have a strong influence on scientists persistence in participating at the specific conference. The problem of the order-parameter persistence (first-passage time), is a well studied phenomenon in non-equilibrium statistical dynamics in condensed matter systems [21]. Persistence is defined as the probability that fluctuating variable does not change the sign until time t , and for many non-equilibrium systems this probability decays with time as a power-law [21]. Here we carry out the analysis of persistence of participation patterns of more than 100000 scientists at six national and international conferences of different sizes and from different fields of science. We study the probability of total and successive number of participations, as well as the distribution of time lags between two successive participations. We find that all three measured probabilities have a shape of a truncated power law, regardless of the conference size and degree of specialization. This indicates that the probability for a participant to attend the next meeting is not constant, but rather it grows/decays with a number of participations/non-participations. This observation is directly related to the strength of the association with the conference community. We propose a microscopic stochastic model which includes this influence of balance between the number of participations and non-participations, as well as the role of conference inclusiveness, on the probability to attend the conference next year. Results of our

model show that the studied conferences have a relatively low inclusiveness, i.e. the probability for a scientist to participate in the next meeting after the first attendance. We also show that conference attendance is characterized by *positive feedback*. The growth in the total number of participations results in a stronger attractiveness of the conference community to participants, and vice versa. Longevity of scientific career of publishing in scientific journals is also characterized by a power-law distribution with an exponential cut-off [12]. Using the empirical analysis and stochastic model Petersen et al. [12] have shown that longevity and past success of scientists lead to cumulative advantage in further development of their career. Although the distribution of career longevity and conference persistence have a similar behaviour, there is a significant difference of characteristic exponents, which indicates that a different mechanism underlie these two phenomena.

This paper is structured as follows: first, we perform empirical analysis of participation patterns for six conferences. We then propose and describe the model of conference participation dynamics. Finally, we perform numerical simulations and discuss some properties of the model, and estimate the values of parameters that correspond to empirical data.

Results

Data set

For our empirical analysis we use data for six conference series in different fields of science. We collected and filtered information about abstracts presented at the American Physical Society March Meeting (APSMM), American Physical Society April Meeting (APSAM), Society for Industrial and Applied Mathematics Annual Meetings (SIAM), Neural Information Processing Systems Conference (NIPS), International Conference on Supercomputing (ICS) and Annual International Conference on Research in Computational Molecular Biology (RECOMB). All these scientific meetings are held annually, but they differ in the topic, sizes, degree of specialisation, longevity and degree of localisation (national versus international). When it comes to the meeting size it can vary from a few dozens, like ICS and RECOMB, to several thousands of participants at APSMM. Some of these meetings are on highly focused topic, NIPS, while others are designed to cover the entire scientific fields, like APSMM, APSAM and SIAM. Four of these conferences (SIAM, NIPS, ICS and RECOMB) have an international character with venues all over the world, while APSMM and APSAM are annual conferences of American Physical Society which are always held in North American cities. APSMM, SIAM and APSAM are conferences with a long tradition, while first meetings of NIPS, ICS and RECOMB have been organized during late 80s and early 90s. Detailed information about conferences and data is given in [S1 File](#).

To be able to track participants at the conference over the years, we have labeled them based on name, affiliation and co-authors and performed author name disambiguation (see [Methods](#) for details). We are interested in studying the participation patterns of scientists starting from their first attendance at the conference series. Thus, for conferences for which the data are not available from their beginning (APSMM, APSAM and SIAM), we have filtered out the authors that may have attended the conference before the starting year in our dataset (see [Methods](#) for the details of our filtering procedure).

Empirical results

For all scientists we have the information about the years of their appearance as authors in the book of abstracts of particular a conference series. The information about the list of authors who actually attended the conference is not available for the conferences considered in this paper. Hence, as a proxy for a conference participation in a given year, we use the appearance

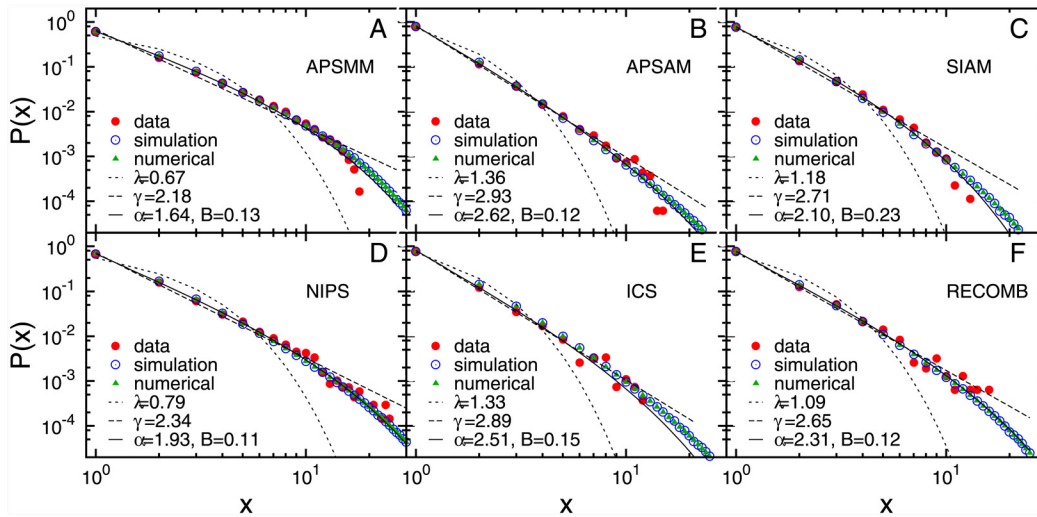


Fig 1. The total number of participations. The probability distribution of the total number of participations obtained from the empirical data (red circles), simulations (blue circles) and numerical iterative algorithm (green triangles). The full line is the best fit to truncated power law, $x^{-\alpha} e^{-Bx}$, while the dashed and dash-dot line denote the best fit to power-law distribution, $x^{-\gamma}$ and exponential distribution, $e^{-\lambda x}$, respectively.

doi:10.1371/journal.pone.0148528.g001

of a scientist as a co-author of at least one abstract in conference proceeding for that year. Not all authors that are mentioned in the book of abstracts have actually attended the conference, but one can argue that as co-authors they have actively contributed to the material presented and thus participate as a contributors in the conference [15].

First we analyse *the total number of author’s participations* (the number of times an author has participated), x , at the given conference series. Fig 1, shows the probability distribution of the total number of participations, $P(x)$, averaged over all participants, for each of the six analysed conferences. The comparison of the quality of fits between exponential, power-law and truncated power-law, Fig 1, shows that all curves are very well represented by power law with exponential cut-off (see Methods), with the value of exponent $\alpha \in (1.6, 2.7)$. The disparity in the total number of participations indicates that most scientists belong to the group of occasional participants, with more than half of all participants attending a particular conference only once. For instance, the percentage of all participants that attend the conference only once is the highest for APSAM and ICS, around 81%, and the lowest for APSMM and NIPS, 63% and 68% respectively. This observation indicates that communities of all these conferences have a relatively low inclusiveness. On the other hand, it is also clear that some of the participants are very regular, attending the conference (almost) every year. These participants form the group of regular attendees whose conference participation is mainly driven by social factors, i.e. their sense of *association with the community*.

In the case of when the probability to attend a conference is constant or random, the expected distribution of total number of attendances is of exponential type. Thus, the power-law nature of the distribution of total participations strongly suggests that the probability of participation at some future conference increases with the number of previous participations. By participating frequently at a particular conference scientists not only expand, but also strengthen, their collaboration network which leads to their further engagement with the community.

We further explore the participation patterns by analysing the number of successive participations (Fig 2) and the time lag between two successive participations (Fig 3). The distributions of these quantities also exhibit the truncated power-law behaviour (see Methods). The observed

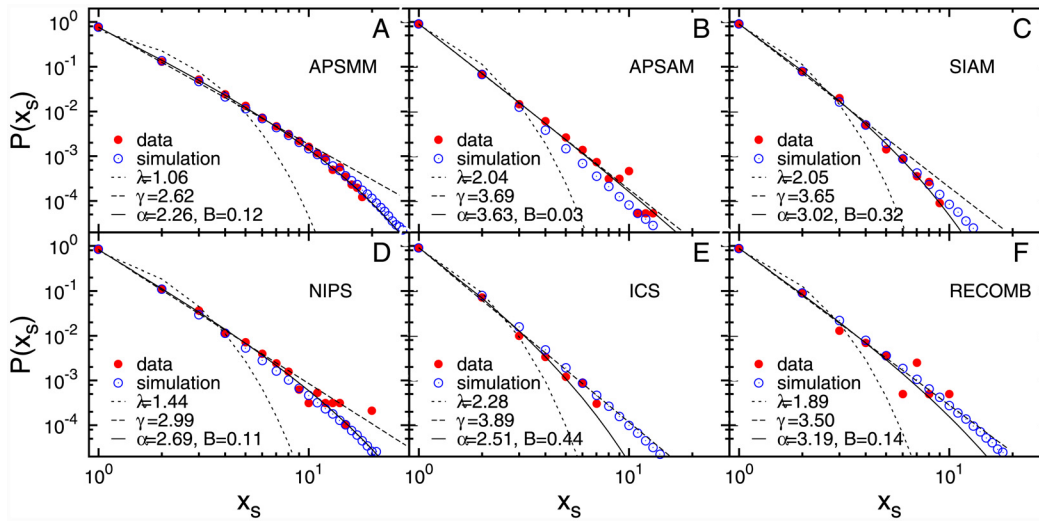


Fig 2. The number of successive participations. The probability distribution of the number of successive participations, x_s , obtained from empirical data (red circles) and numerical simulations of the model (blue circles). The full, dashed and dash-dot line are the best fit to truncated power law, power-law and exponential function respectively.

doi:10.1371/journal.pone.0148528.g002

distributions of the number of successive participations, with exponent $2 \leq \alpha \leq 4$, suggests that even frequent attendees make a pause in their participation, although these breaks are usually short, i.e. long breaks of five and more years occur with a low probability, Fig 3. A long-period of non-participation results in fading of existing collaboration ties with the community while new ones are never formed. Due to this fading, the probability to attend the meeting decreases with total number of non-participations. This indicates that conference participation of most scientists takes place in a limited period of time with a relatively short and small number of breaks.

As it was shown in Ref [12] the distribution of the journal career longevity exhibits a truncated power-law behaviour with cut-off around 10 years. The exponential cut-off in the distribution of all three measures is a consequence of the two combined finite-size effects that influence the asymptotic behaviour, the finite life time of scientist’s association with one community or her/his career in one field of research or in science in general [12], and limitations of used datasets. This effect will be also observed in the distribution of conference participations. The end of a career inevitably results in a termination of participation in conferences and thus also the conference community membership. Also, used datasets have a relatively short time span (less than three decades), due to which they do not include scientists with long careers [12]. Both of these effects affect the value of the exponential cut-off, which is lower in the case of conference participation, between 4 and 9 years, compared to the one observed for the career longevity.

Model

The empirical results from six different series shown in the previous section indicate that the probability for a scientist to attend the next meeting of a conference series depends on the balance of previous participations and non-participations. Petersen et al. [12] show that Matthew (*rich get richer*) effect is responsible for the career longevity in several competitive professions, including science. They argue that it becomes easier to move forward in the career with an increasing past success of an individual, and show, using their stochastic career progressive model, that this mechanism leads to a truncated power-law distribution of the career longevity.

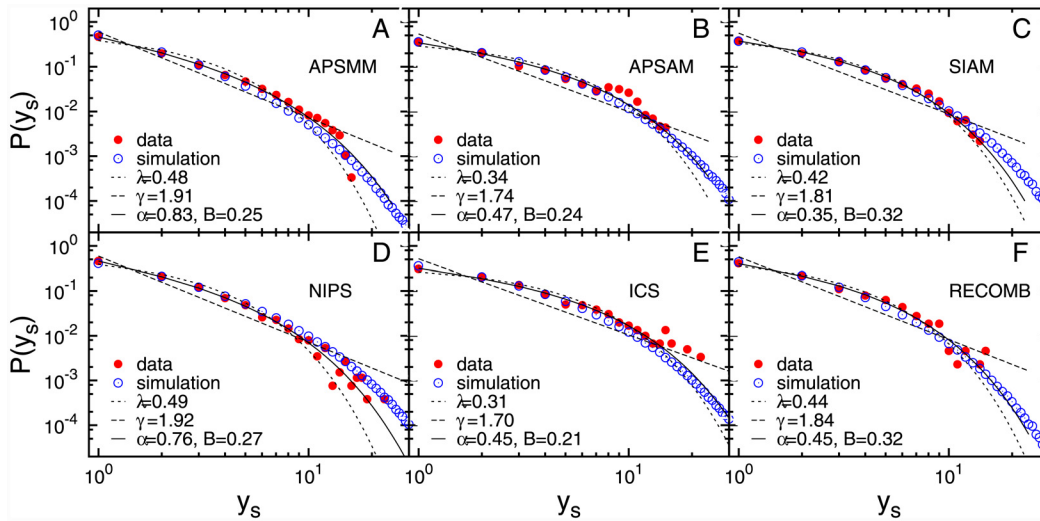


Fig 3. The time lag between the two successive participations. The probability distribution of the time lags between two consecutive conference participations y_s : empirical data (red circles) and numerical simulations data (blue circles). The lines correspond to respective fits as in Figs 1 and 2.

doi:10.1371/journal.pone.0148528.g003

In their model, they assume that the stochastic process governing career progress is similar to Poisson process, where progress is made at any given step with the rate $g(x) \equiv 1 - \exp[-(x/x_c)^\alpha]$, where $1/x_c$ is a hazard rate corresponding to random career ending while the parameter α is the same as power-law exponent in the pdf of career longevity. Using this model for $\alpha < 1$ they were able to obtain truncated power-law distributions for career duration in several professions.

The empirical results of conference participation patterns suggest that the probability for a scientist to participate in a conference is not constant or random, but that it rather grows with the number of participations. This is reflected in the increase of proportion of authors who are going to attend the conference next year with total number of previous conference attendance (see Figure A in S1 File). Higher number of participations of a scientist at the conference results in better connections with the community and thus higher probability that the author will participate in the following conference. But unlike career longevity, where the length of the waiting times between two successive steps in the career does not influence the progress rate, the probability for conference participation is strongly influenced by the number and length of pauses (Figure B in S1 File). The longer the scientists are absent from the community the weaker are their connections and lower are the probabilities to participate in the following events. For this reason and the fact that the pdf obtained from the model proposed in Ref [12] exhibits a truncated power-law only for the exponents $\alpha < 1$ Petersen et al. model [12] cannot be applied for modelling conference participation dynamics.

We propose a new stochastic model for conference attendance dynamics which can explain our empirical findings. Our model is based on a 2-bin generalized Pólya process [22–24] and random termination time of a career. As opposed to the Petersen model where the progress rate depends only on the current position of scientist in his/her career, the 2-bin generalized Pólya incorporates dependence on the balance between participations and non-participations. Let x stands for the total number of participations at the conference, y stands for the number of conferences an author has not participated since she/he appeared at the conference for the first time and t is the number of events held, $t = x + y$. All authors start with $x = 1$ and $y = 0$. According to our model, the probability that a scientist with x total number of participations and y

number of non-participations will appear at the next conference is given by

$$g(x, y) = \frac{x^p}{x^p + (y + y_0)^p} = \frac{z^p}{1 + z^p}, \tag{1}$$

where $z = \frac{x}{y+y_0}$ measures the balance between participations and non-participations, parameter p is the *exponent* of the model, and y_0 determines the initial balance value. The probability that a scientist will not attend the next conference is equal to $1 - g(x, y)$. Depending on the exponent p , the function g can correspond to positive ($p > 1$) or negative feedback ($p < 1$) [22]. When $p = 1$ and $y_0 = 0$, the Eq 1 is equivalent to the equation for a Pólya-Eggenberg problem [25]. As we shall see in the following section, the value of the parameter p for all conferences is larger than one, suggesting that the conference participation dynamics is characterized by the *positive feedback*: scientists who participate in the conference frequently and make less and shorter pauses have a stronger association with the conference community and thus have a higher probability to participate in the following events. The value of the parameter y_0 determines the probability of a scientist to attend the next event after her/his first occurrence at the conference. According to our model this parameter is the same for all scientists attending one conference series, thus it can be interpreted as a measure of the conference community inclusiveness.

Evolution equation. The probability $P(x, t)$ for the author to have x conference participations after t conferences since his/her first participation is equal to the probability to attend the next conference $g(x - 1, t - x)$ times the probability of already attending $x - 1$ conferences at time $t - 1$ plus the probability of skipping the next conference $1 - g(x, t - 1 - x)$ times the probability of already attending x conferences at time $t - 1$:

$$P(x, t) = \frac{(x - 1)^p}{(x - 1)^p + (t - x + y_0)^p} P(x - 1, t - 1) + \frac{(t - 1 - x + y_0)^p}{x^p + (t - 1 - x + y_0)^p} P(x, t - 1). \tag{2}$$

The probability distribution $P(x)$ of the number of total conference attendances for a particular conference series is obtained by summing $P(x, t = T)$ over all possible T :

$$P(x) = \sum_{T=1}^{\infty} P(x, t = T) P(T), \tag{3}$$

where T denotes the duration of a scientist’s membership in the community. In our case, we assume that the duration of a scientist’s membership in a conference community can be terminated at any year after his/her first appearance with probability H , which gives the distribution of time intervals

$$P(T) = H(1 - H)^{T-1}. \tag{4}$$

Numerical simulation results

Since the analytical solution of Eq 3 cannot be obtained, we estimate the model parameters y_0 , H and p using numerical simulations (see [Methods](#)). The best estimates of the model parameters for each of the six conferences are given in Table G in [S1 File](#). As shown in Figs 1, 2 and 3, the model with the properly chosen parameters nicely reproduces the behaviour of participants at six conferences, for all three measured quantities.

For all six conferences the estimated value of parameter p is greater than 1, which suggests that the positive feedback mechanism underlies the conference participation dynamics. This means that the probability for a scientist to attend the next year event grows superlinearly with the balance between the number of participations and pauses (z). The value of the parameter y_0

together with the value of p determines the probability for a scientist to participate in the conference next year after his/her first participation, i.e. the initial inclusiveness of the conference community. Table H in [S1 File](#) shows the estimated value of the initial inclusiveness for all six conferences. The APSMM has the highest probability, around 25%, for newcomers to attend the conference next year, while APSAM has the lowest, 9%. One could assume that the size and diversity of topics of a conference have an essential influence on conference inclusiveness, but according to our results this is not the case. The ordering of the conferences according to size, Table H in [S1 File](#), and their initial inclusiveness do not correlate. APSAM is the second largest conference but has the lowest inclusiveness, while the RECOMB as the smallest conference is ranked as third and has the inclusiveness of 15%. Further, it follows from our results that the diversity of topics covered by the conference does not have a significant effect on the return probability of newcomers. Although the first ranked conference according to inclusiveness, APSMM, covers the widest range of topics among considered conferences, the APSAM and SIAM, which are also considered general conferences, have a lower inclusiveness than NIPS and RECOMB. This suggests that the conference inclusiveness is influenced by some other factors, which are not related to the size, degree of specialisation or localisation (national and international), but rather to social structure and openness of the conference community toward newcomers.

We solve [Eq 3](#) numerically using an iterative method (see SI for more details) and compare it with simulation results. [Fig 1](#) shows an excellent matching between results obtained using the iterative algorithm and numerical simulations for the estimated values of parameters.

Discussion and conclusion

The goal of this paper has been to investigate the conference participation patterns and propose a simple stochastic model of conference participation dynamics. The motivation behind this is to better understand the mechanisms that underlie the repeated participation in the same conference series and explore whether the conference series topic, size, degree of specialisation, longevity and degree of localisation (national and international) influence the participation probability and inclusiveness of the specific community. Our study is based on empirical analysis and modelling of authors participation at six different conference series in the last three decades: APSMM, APSAM, SIAM, NISP, ICS and RECOMB. We note here that it would be important to verify our findings with the data from other conferences.

The set of considered conferences is very heterogeneous. Although they differ in size, topic and topic diversity, national structure of participants and conference longevity, they are characterized with similar participation patterns. The distributions of the total number of participations for all six conferences exhibit the same, truncated power-law, behaviour with values of exponent α between 1.6 and 2.7. A similar behaviour is also observed for the distributions of the number of successive participations and the duration of pauses between them. The observed statistical evidence strongly imply that the dynamics of conference participation is governed by universal forces which are independent of the specific conference features or the scientific field. This and the fact that conferences often have a stable core of attending participants [20] suggests that these have social origins and that social factors, such as the association with a conference community and its inclusiveness, strongly influence the probability for a scientist to attend the future meetings and their participation patterns at the specific conference series, accordingly.

The observed truncated power-law behaviour of the distributions of participations indicates that the probability for a scientist to participate in the next year conference is growing(decreasing) with the balance between the number of participations and pauses. To further explore this

we proposed a stochastic model based on 2-bin generalized Pólya process which incorporates the dependence on the ratio between number of participations and pauses. Our model shows that the positive feedback mechanism underlies the conference participation dynamics. The probability for a scientist to attend a conference grows superlineary with the number of participations, while the frequent pauses have the opposite effect. The scientists who are able to overcome the initial obstacles and create social ties with the conference community by frequent participation at the beginning have a higher probability to attend the conference in the following years. A frequent participation strengthens the scientist's association with a conference community which further increases the probability for future participations. On the other hand, scientists with a small number of initial participations have a low probability to participate in the following conference, thus small number of participations, and eventually stop attending the conference. The initial inclusiveness of the specific conference community has the main influence on early participation patterns. As we showed, this inclusiveness does not depend on the size, degree of specialisation or topic of the conference, but rather on the openness of the community toward newcomers.

Our analysis indicates that social factors, such as the association with the community and the community inclusiveness are the main driving forces of conference participation dynamics. In general the community/group cohesion and the ability to attract and retain newcomers and other members influence the dynamics of their participation in group activities [26]. On the other hand, a member's engagement in group activities strengthens ties to other group/community members, and contributes to the creation of the bonding capital, while the ties of non-attendees dissolve and weaken with time [27]. Conference communities are just one example of these systems, thus we expect to observe the similar group participation patterns in other types of social communities, both online and offline. Further investigations and studies of other social systems will reveal and characterize the connection between a social network structure and group inclusiveness, and participation dynamics in group activities.

Methods

Data filtering Identification of the different authors may involve a few issues. On one hand, an author may use different spelling variants to sign his first and middle name. On the other hand, the author's name may be related to several different authors, thus using only the initials of the last name and first name increases additionally error rates in disambiguating the author names. In our data sets, data from NIPS and RECOMB conferences did not require additional cleaning, while for the SIAM and ICS data, we have used python fuzzy partial string matching of author's first and middle names, which gave a high accuracy. For APSMM and APSAM conferences, where data are highly heterogeneous, we have used a method described in [28] to disambiguate the author names. This method considers pairs of names that match on last name and first name initials. Then it groups the authors based on their affiliation and co-authors. Because the same affiliation could be formatted differently, the two affiliations were considered the same if their fuzzy token set ratio was higher than 50%.

The sources and detailed description of the data are given in Tables A, B and C in [S1 File](#). For NIPS, ICS and RECOMB, we have complete data from their very beginning. Remaining data sets required filtering out the authors with a high probability of attending conference before the starting year in our dataset, Y_0 . Therefore, for APSMM, APSAM and SIAM we have isolated authors with the first recorded year of conference attendance, smaller than $Y_0 + \langle \tau \rangle$, where $\langle \tau \rangle$ is the average waiting time between a consecutive conference attendance for all the authors who took part at the conference during the $[Y_0, Y_f]$ period. This way we excluded between 10% (APSMM and SIAM) and 25% (APSAM) authors from our analysis.

Functional fits We have used the maximum-likelihood fitting method [29] to fit three different functions to the probability distributions of the total number of participations, the number of and the time lags between two successive participations: exponential function $e^{-\lambda x}$, power-law function $x^{-\gamma}$ and truncated power-law $x^{-\alpha} e^{-Bx}$. It follows from the comparison of fits of these three functions to empirical data that the truncated power-law is the best fit for the probability distribution of all three measured quantities, see Figs 1, 2 and 3. In order to compare these three fits we calculate the log likelihood ratio, \mathcal{R} , and π -value (see Ref [29]) which compares the fits to the power-law with exponential cut-off with the pure power-law for the distribution of total number of participations (Table D in S1 File) and the number of successive participations (Table E in S1 File). In the case of nested distributions, the negative value of \mathcal{R} indicates that the larger family of distributions, in this case the truncated power-law, is a superior model. When the value of \mathcal{R} tends to 0, one can use π -value. The small π -value suggests that the smaller family of distributions, in this case power-law, can be ruled-out. Both the log likelihood ratio and the π -value indicate that the truncated power-law is a superior model compared to pure power-law for both distributions. A similar procedure can be applied for the comparison between truncated power-law and exponential fits, but since from the visual inspection it is clear that the distributions do not follow the exponential fits, we have omitted these results. The comparison between exponential and the power-law with exponential cut-off fit, given in Table F in S1 File, indicates that the power-law distribution with exponential cut-off fit is better than exponential fit for the distribution of the time lags. For all six conferences, the power-law with exponential cut-off distribution gives the best fit for all three empirical distributions.

Parameter estimation We simulate the model for $N = 100000$ different authors. Starting from $x = 1$ and $y = 0$ at $t = 1$, an author will appear at the next conference with probability $g(x, y)$ or skip it with the probability $1 - g(x, y)$. The author can terminate his/her membership in the community at each time step with the probability H . In order to estimate the values of parameters p , y_0 and H , we calculate the distribution of total number of attendances x , from the simulations and compare it to the empirical distribution using Kullback-Leibler Distance [30]. We perform the simulations for several different sets of parameter (y_0, H, p) to determine which combination of parameter values makes the model optimally close to the empirical data. For each parameter set the results are averaged across 100 simulations.

Supporting Information

S1 File. Supplementary Information: A theoretical model for the associative nature of conference participation. Proportion of conference participants g with x conference attendances who are going to attend the conference next year (Figure A). Proportion of conference participants ρ with n missed conferences after x -th conference attendance who are going to skip the conference next year, but will take part at some future conference from the observation period (Figure B). Pages on the web from which we downloaded conference data (Table A). Summary of the conference data. Columns 2 and 3 indicate for each conference the year in which data we have collected begin (Y_0) and end (Y_f). The total number of different participants at the conference during that period of time is given in column 4 (Table B). The number of participants at the conference per year (Table C). Log likelihood ratio \mathcal{R} and the π -value compare the fit to the power-law with the fit to the power-law with an exponential cutoff for the probability distribution of number of conferences at which each author appears (Table D). Log likelihood ratio \mathcal{R} and the π -value compare the fit to the power-law with the fit to the power-law with an exponential cutoff for the probability distribution of the number of successive participations at the conference (Table E). Log likelihood ratio \mathcal{R} and the π -value compare the fit to the

exponential with the fit to the power-law with an exponential cutoff for the probability distribution of the time lag between two consecutive conference participations (**Table F**). The optimal parameter values of the model for each conference (**Table G**). Stagnancy rate $1 - g(1, 0)$ at $t = 1$ for each conference and exponent α of power-law with an exponential cutoff distribution fit with the corresponding conference order (**Table H**).
(PDF)

Acknowledgments

Numerical simulations were run on the PARADOX supercomputing facility at the Scientific Computing Laboratory of the Institute of Physics Belgrade.

Author Contributions

Conceived and designed the experiments: JS AC TK MMD. Analyzed the data: JS. Wrote the paper: JS AC TK MMD. Collected the empirical data: JS.

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):721–723. PMID: [19197046](#)
2. Cioffi-Revilla C. Computational social science. *CompStat*. 2010; 2(3):259–271.
3. Castellano C, Fortunato S, Loreto V. Statistical physics of social dynamics. *Rev Mod Phys*. 2009; 81:591–646. doi: [10.1103/RevModPhys.81.591](#)
4. Sen P, Chakrabarti BK. *Sociophysics: An Introduction*. Oxford University Press, Oxford; 2014.
5. Schweitzer F. Scientific networks and success in science. *EPJ Data Science*. 2014; 3:35. doi: [10.1140/epjds/s13688-014-0035-8](#)
6. Radicchi F, Fortunato S, Castellano C. Universality of citation distributions: Toward an objective measure of scientific impact. *Proc Natl Acad Sci USA*. 2008; 105(45):17268–17272. doi: [10.1073/pnas.0806977105](#) PMID: [18978030](#)
7. Radicchi F, Fortunato S, Markines B, Vespignani A. Diffusion of scientific credits and the ranking of scientists. *Phys Rev E*. 2009; 80(5):056103. doi: [10.1103/PhysRevE.80.056103](#)
8. Wang D, Song C, Barabási AL. Quantifying long-term scientific impact. *Science*. 2013; 342(6154):127–132. doi: [10.1126/science.1237825](#) PMID: [24092745](#)
9. Sinatra R, Wang D, Deville P, Song C, Barabasi AL. Scientific impact: the story of your big hit. In: *APS Meeting Abstracts*; 2014. p. 17004.
10. Petersen AM, Riccaboni M, Stanley HE, Pammolli F. Persistence and uncertainty in the academic career. *Proc Natl Acad Sci USA*. 2012; 109(14):5213–5218. doi: [10.1073/pnas.1121429109](#) PMID: [22431620](#)
11. Azoulay P, Graff Zivin JS, Wang J. Superstar Extinction. *QJ of Econ*. 2010; 125(2):549–589. doi: [10.1162/qjec.2010.125.2.549](#)
12. Petersen AM, Jung WS, Yang JS, Stanley HE. Quantitative and empirical demonstration of the Matthew effect in a study of career longevity. *Proc Natl Acad Sci USA*. 2011; 108(1):18–23. doi: [10.1073/pnas.1016733108](#) PMID: [21173276](#)
13. Petersen A, Penner O. Inequality and cumulative advantage in science careers: a case study of high-impact journals. *EPJ Data Science*. 2014; 3(1):24. doi: [10.1140/epjds/s13688-014-0024-y](#)
14. Deville P, Wang D, Sinatra R, Song C, Blondel VD, Barabási AL. Career on the Move: Geography, Stratification, and Scientific Impact. *Sci Rep*. 2014; 4:4770. doi: [10.1038/srep04770](#) PMID: [24759743](#)
15. Borghans L, Romans M, Sauermann J. What makes a good conference? Analysing the preferences of labour economists. *Labour Economics*. 2010; 17(5):868–874.
16. Mair J, Thompson K. The UK association conference attendance decision-making process. *Tourism Management*. 2009; 30(3):400–409.
17. Witt SF, Sykes AM, Dartus M. Forecasting international conference attendance. *Tourism Management*. 1995; 16(8):559–570.
18. Garfield E. Making contacts at conferences. *The Scientist*. 1988; 2:294–295.

19. Garfield E. Making contacts at conferences—a problem for the young scientist. *Essays of an Information Scientist*. 1977; 3:668–672.
20. Van Dijk J, Maier G. ERSAs Conference participation: does location matter? *Papers in Regional Science*. 2006; 85(4):483–504. doi: [10.1111/j.1435-5957.2006.00102.x](https://doi.org/10.1111/j.1435-5957.2006.00102.x)
21. Majumdar SN. Persistence in nonequilibrium systems. *Current Science*. 1999; 77(3):370–375.
22. Drinea E, Frieze A, Mitzenmacher M. Balls and bins models with feedback. In: *Proceedings of the Thirteenth Annual ACM-SIAM Symposium on Discrete Algorithms*. SODA '02; 2002. p. 308–315.
23. Pemantle R. A survey of random processes with reinforcement. *Probab Surveys*. 2007; 4:1–79. doi: [10.1214/07-PS094](https://doi.org/10.1214/07-PS094)
24. Oliveira RI. The onset of dominance in balls-in-bins processes with feedback. *Random Structures & Algorithms*. 2009; 34(4):454–477. doi: [10.1002/rsa.20261](https://doi.org/10.1002/rsa.20261)
25. Johnson NL, Kotz S. *Urn Models and Their Application: An Approach to Modern Discrete Probability Theory*. Wiley; 1977.
26. Friedkin NE. Social cohesion. *Annu Rev Sociol*. 2004; 30:409–425. doi: [10.1146/annurev.soc.30.012703.110625](https://doi.org/10.1146/annurev.soc.30.012703.110625)
27. Sessions LF. How offline gatherings affect online communities. *Information, Communication & Society*. 2010; 13(3):375–395. doi: [10.1080/13691180903468954](https://doi.org/10.1080/13691180903468954)
28. Wu J, Ding XH. Author name disambiguation in scientific collaboration and mobility cases. *Scientometrics*. 2013; 96(3):683–697. doi: [10.1007/s11192-013-0978-8](https://doi.org/10.1007/s11192-013-0978-8)
29. Clauset A, Shalizi CR, Newman MEJ. Power-law distributions in empirical data. *SIAM Review*. 2009; 51(4):661–703. doi: [10.1137/070710111](https://doi.org/10.1137/070710111)
30. Burnham KP, Anderson DR. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. Springer Science & Business Media; 2002.

RESEARCH ARTICLE

Associative nature of event participation dynamics: A network theory approach

Jelena Smiljanić^{1,2*}, Marija Mitrović Dankulov¹

1 Scientific Computing Laboratory, Center for the Study of Complex Systems, Institute of Physics Belgrade, University of Belgrade, Pregrevica 118, 11080 Belgrade, Serbia, **2** School of Electrical Engineering, University of Belgrade, P.O. Box 35-54, 11120 Belgrade, Serbia

* jelenas@ipb.ac.rs



Abstract

The affiliation with various social groups can be a critical factor when it comes to quality of life of each individual, making such groups an essential element of every society. The group dynamics, longevity and effectiveness strongly depend on group's ability to attract new members and keep them engaged in group activities. It was shown that high heterogeneity of scientist's engagement in conference activities of the specific scientific community depends on the balance between the numbers of previous attendances and non-attendances and is directly related to scientist's association with that community. Here we show that the same holds for leisure groups of the Meetup website and further quantify individual members' association with the group. We examine how structure of personal social networks is evolving with the event attendance. Our results show that member's increasing engagement in the group activities is primarily associated with the strengthening of already existing ties and increase in the bonding social capital. We also show that Meetup social networks mostly grow through big events, while small events contribute to the groups cohesiveness.

OPEN ACCESS

Citation: Smiljanić J, Mitrović Dankulov M (2017) Associative nature of event participation dynamics: A network theory approach. PLoS ONE 12(2): e0171565. doi:10.1371/journal.pone.0171565

Editor: Matjaz Perc, University of Maribor, SLOVENIA

Received: November 15, 2016

Accepted: January 22, 2017

Published: February 6, 2017

Copyright: © 2017 Smiljanić, Mitrović Dankulov. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All data files are available at https://figshare.com/articles/Meetup_Datasets/2066904 (DOI: <https://dx.doi.org/10.6084/m9.figshare.2066904.v1>).

Funding: This work was supported by ON171017, The Ministry of Education, Science, and Technological Development of the Republic of Serbia (<http://www.mps.gov.rs/>: J.S. and M.M.D.); and 675121, The European Commission (http://ec.europa.eu/index_en.htm: M.M.D.).

Competing Interests: The authors have declared that no competing interests exist.

Introduction

One of the consequences of the rapid development of the Internet and growing presence of information communication technologies is that a large part of an individual's daily activities, both off and online, is regularly recorded and stored. This newly available data granted us a substantial insight into activities of a large number of individuals during long period of time and led to the development of new methods and tools, which enable better understanding of the dynamics of social groups [1]. The structure and features of social connections both have strong influence and depend on social processes such as cooperation [2, 3], diffusion of innovations [4, 5], and collective knowledge building [6]. Therefore, it is not surprising that the theory of complex networks has proven to be very successful in uncovering mechanisms governing the behavior of individuals and social groups [7, 8].

The human activity patterns, the structure of social networks, and the emergence of collective behavior in different online communities have been extensively studied in the last decade

[6, 9–15]. On the other hand, the dynamics of offline social groups, where the activities take place through offline meetings (events), have drawn relatively little attention, given their importance. Such offline groups, both professional and leisure ones, provide significant benefits and influence everyday lives of individuals, their broader communities, and the society in general: they offer social support to vulnerable individuals [16, 17], can be used for political campaigns and movements [18, 19], or can have an important role in career development [20]. As they have different purpose, they also vary in the structure of participants, dynamics of meetings, and organisation. Some groups, such as cancer support groups or scientific conference communities, are intended for a narrow circle of people while others, leisure groups for instance, bring together people of all professions and ages. In the pre-Internet era these groups have been, by their organisation and means of communication between their members, strictly offline, while today we are witnessing the appearance of a growing number of hybrid groups, which combine both online and offline communication [19]. Although inherently different, all these social groups have two main characteristics in common: they do not have formal organization, although their members follow certain written and non-written rules, and their membership is on a voluntary basis. Bearing this in mind, it is clear that the function, dynamics and longevity of such self-organized communities depend primarily on their ability to attract new and retain old active members. Understanding the reasons and uncovering key factors that influence members to remain active in the social group dynamics are thus important, especially having in mind their relevance for the broader social communities and the society.

The size of social groups and personal social networks, as well as their structure, have been extensively studied. The considerable body of evidence [21–24] suggests that the typical size of natural human communities is approximately 150, that both groups and personal social networks are highly structured, and consist of social layers characterized by different strengths of relationships. The relationships within each layer are characterized by a similar mean frequency of interaction and emotional closeness, both of which decrease rapidly as we move through network layer. These findings have been explained using the Social Brain Hypothesis, which relates the average size of species' personal network with the computational capacity of its brain. Here we confirm that these findings also hold for leisure groups where the frequency of interactions among members is constrained by the event dynamics. We also explore how the number of attended events is related to the size and layered structure of member's personal network.

Previous research on hybrid social groups and interplay between offline and online interactions has shown that offline meetings enhance attendees' engagement with online community and contributes to the creation of a bonding capital [25, 26]. In our previous work [20] it was shown that the participation patterns of scientist in a particular conference series are not random and that they exhibit a universal behaviour independent of conference subject, size or location. Using the empirical analysis and theoretical modeling we have shown that the conference attendance depends on the balance between the numbers of previous attendances and non-attendances and argued that this is driven by scientist's association with the conference community, i.e. with the number and strength of social ties with other members of the conference community. We also argued that similar behaviour can be expected in other social communities when it comes to members' participation patterns in organised group events. Here we provide empirical evidence supporting these claims and further investigate the relationship between the dynamics of participation of individuals in social group activities and the structure of their social networks.

The Meetup portal, whose group dynamics we study here, is an event-based social network. Meetup members use the online communication for the organization of offline gatherings. The online availability of event attendance lists and group membership lists enables us to

examine the event participation dynamics of Meetup groups and its influence on the structure of social networks between group members. The diversity of Meetup groups in terms of the type of activity and size allows us to further examine and confirm the universality of member's participation patterns. We note that previous papers using Meetup source of data have mostly focused on the event recommendation problem [27–31], structural properties of social networks, and relationships between event participants [31, 32] by disregarding evolutionary behaviour of Meetup groups.

In this paper, we examine the event-induced evolution of social networks for four large Meetup groups from different categories. Similarly to the case of conference participation, we study the probability distribution of a total number of meetup attendances and show that it also exhibits a truncated power law for all four groups. This finding suggests that the event participation dynamics of Meetup groups is characterized by a positive feedback mechanism, which is of social origin and is directly related to member's association with social community of the specific Meetup group. Using the theory of complex networks we examine in more detail the correlation between an individual's decisions to participate in an event and her association with other members of that Meetup group. Specifically, we track how member's connectedness with the community changes with the number of attendances by measuring change in the clustering coefficient and relation between the degree and the strength in an evolving weighted social network, where only statistically significant connections are considered. Our results indicate that greater involvement in group activities is more associated with the strengthening of existing than to creation of new ties. This is consistent with previous research on Meetup which has shown that repeated event attendance leads to an increase in bonding and a decrease in bridging social capital [25, 33]. Furthermore, in view of the fact that people interact and networks evolve through events, we examine how particular a event affects the network size and its structure. We investigate effects of event sizes and time ordering on social network organization by studying changes in the network topology, numbers of distinctive links and clustering, caused by the removal of a specific event. We find that large events facilitate new connections, while during the small events already acquainted members strengthen their interpersonal ties. Similar behavior was observed at the level of communities, where small communities are typically closed for new members, while contrary to this, changes in the membership in large communities are looked at favorably [34, 35].

This paper is organized as follows: we first study the distribution of the total number of participations in four Meetup groups from different categories. Next we introduce a filtered weighted social network to characterize significant social connections between members and discuss its structural properties. Specifically, we study how the local topological properties evolve with the growth of the number of participations in order to derive relationships between members' association with the group and their activity patterns. In order to analyze impact of a particular event on the network organization, we remove events using different strategies and study how this influences the social structure.

Results

Event participation patterns of Meetup groups

Meetup is an online social networking platform that enables people with a common interest to start a group with a purpose of arranging offline meetings (events, meetups) all over the world. The groups have various topics and are sorted into 33 different categories, such as careers, hobbies, socializing, health, etc. These groups are of various sizes, have different event dynamics, and hierarchical organisation. They also differ in the type of activity members engage, ranging from socializing events (parties and clubbing) to professional trainings (seminars and

Table 1. Summary of collected data for four selected Meetup groups. N_m is total number of group members, N_e is total number of organised events.

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

doi:10.1371/journal.pone.0171565.t001

lectures). Common to all groups is the way they organize offline events: each member of the group gets an invitation to event to which they reply with yes or no, creating in that way a record of attendance for each event. We use this information to analyze event participation patterns and to study the evolution of the social network.

Here, in particular we analyze four large groups, each belonging to a different category and having more than three thousand organized events (see [Methods](#) and [Table 1](#)). We chose these four groups because of their convenience for statistical analysis, large number of members and organized events, and also for the fact that they are quite different concerning the type of activities and interests their members share. The *geamcIt* (GEAM) group is made of *foodie thrill-seekers* who mostly meet in restaurants and bars in order to try out new exciting foods and drinks, while people in the *VegasHiker* (LVHK) group are hikers who seek excitement through physical activity. The *Pittsburgh-free* (PGHF) is our third group which invites its members to free, or almost free, social events, and the fourth considered group *TechLife Columbus* (TECH), which is about social events and focuses on technology-related community networking, entrepreneurship, environmental sustainability, and professional development.

[Fig 1](#) demonstrates that the probability distributions of total attendance numbers of members in events for all four groups exhibits a truncated power law behavior (see [Fig A](#) and [Table A](#) in [S1 File](#), which show a comparison with the exponential and power law fit), with power law exponent larger than one. Similarly to the conference data [20], the exponential

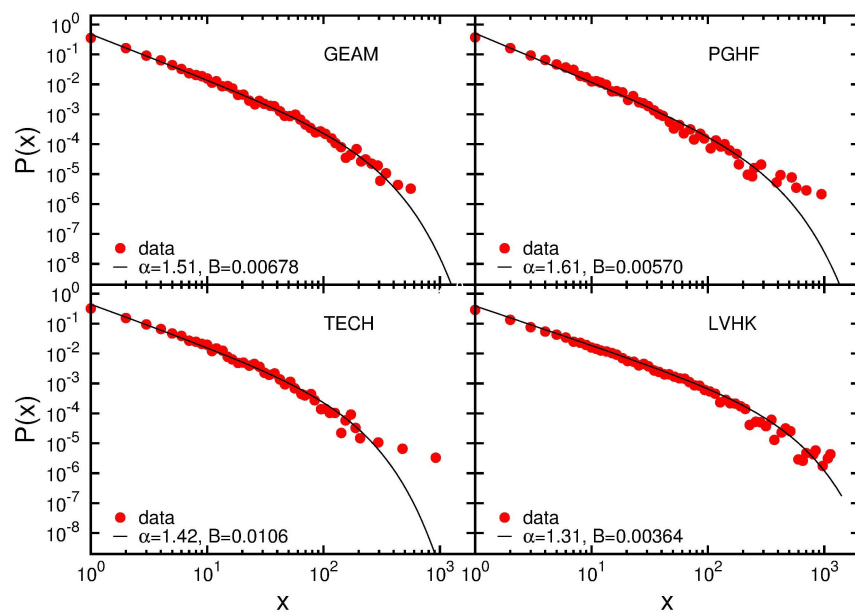


Fig 1. Total number of attended events. Probability distributions $P(x)$ of total number of participations x , for four Meetup groups. Solid line represents best fit to truncated power law distribution, $x^{-\alpha}e^{-Bx}$.

doi:10.1371/journal.pone.0171565.g001

cut-off is a finite size effect. Power law and truncated power law behavior of probability distributions can be observed for the number of and the time lag between two successive participations in group-organized events, Figs B and C in [S1 File](#). In fact, we find that similar participation patterns which differ in values of exponents) can be observed for all Meetup groups, regardless of their size, number of events or category. As in the case of the conference participation dynamics [20], this indicates that the probability to participate in the next event depends exclusively on the balance of numbers of previous participations and non-participations. We argued in [20] that the forces behind conference participation dynamics are of social origin, and it follows from [Fig 1](#) that the same can be argued for the case of the Meetup group participation dynamics. The more participations in group activities member has, the stronger and more numerous are her connections to the other group members, and thus her association with the community. We further explore this assumption by investigating the event-driven evolution of social networks of the four different Meetup groups.

Structure of social event-based network

We construct a social network between group members for each considered group, as a network of co-occurrence on the same event (see [Methods](#) for more details). By definition, these networks are weighted networks with link weights between two members equal to the number of events they participated together. These networks are very dense, as a direct result of the construction method, with broad distribution of link weights (see [Fig D](#) in [S1 File](#)). However, co-occurrence at the same event does not necessarily imply a relationship between two group members. For instance, a member of a group that attends many events, or big events, has a large number of acquaintances, and thus large number of social connections, which are not of equal importance regarding her association with the community. Similarly, two members that attend a large number of events can have relatively large number of co-occurrences, which can be the result of coincidences and not an indicator of their strong relationship. In order to filter out these less important connections we use a filtering technique based on the configuration model of bipartite networks [36, 37] (see [Methods](#)). By applying this technique to weighted networks we reduce their density and put more emphasis on the links that are less likely to be the result of coincidences. In this way we emphasize the links of higher weights without the removal of all links below certain threshold (see [Fig D](#) in [S1 File](#)), a standard procedure for network pruning. We explore the evolution of social networks of significant relationships between Meetup group members by studying how the local characteristics of the nodes (members) change with their growing number of participations in group activities.

Association with the community of a specific Meetup group can be quantitatively expressed through several local and global topological measures of weighted networks. Specifically, here we explore how the number of significant connections (member's degree) and their strength (member's strength), as well as how member's embeddedness in a group non-weighted and weighted clustering coefficient) are changing with the number of attended group events. [Fig 2](#) shows how average strength of a node depends on its degree in filtered networks of four selected Meetup groups. While member's degree equal to the number of member's significant social relationships, the strength measures how strongly she is connected to the rest of the group [38]. In all considered Meetup groups members with small and medium number of acquaintances ($q \leq 50$) have similar values of strengths and degrees, i.e., their association with the community is quantified by the number of people they know and not through the strength of their connections (see [Fig 2](#)). Having in mind that the average size of an event in these four groups is less than 20, we can conclude that majority of members with a degree less than 50 are the ones that attended only a few group meetups. A previous study [30] has found that the

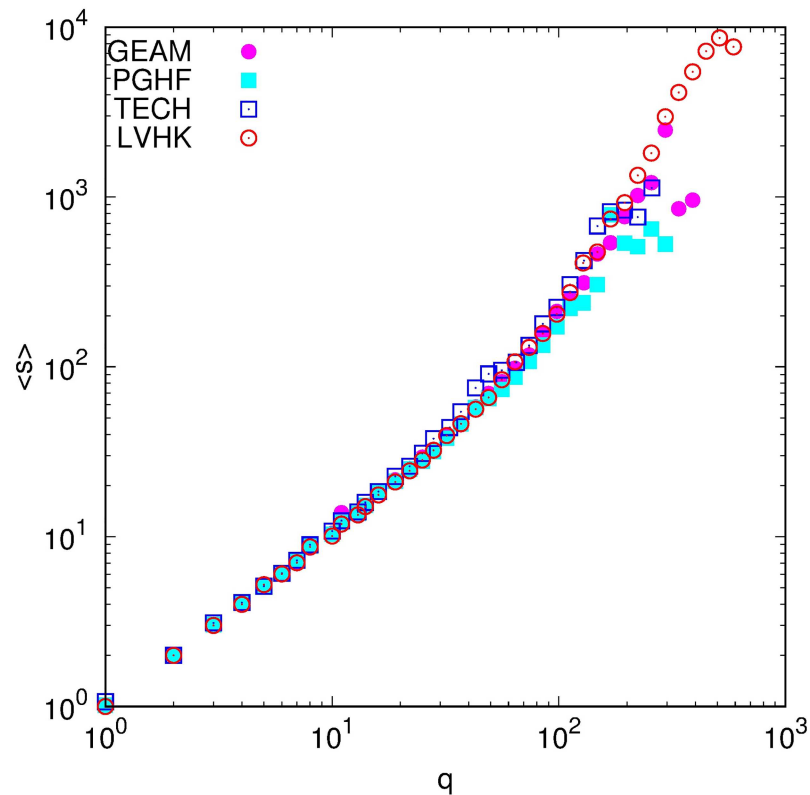


Fig 2. Node strength dependence on node degree. Dependence of average member's strength (s) on her degree q in social network of significant links for considered groups.

doi:10.1371/journal.pone.0171565.g002

probability for a member to attend a group event strongly depends on whether her friends will also attend. The non-linear relationship between the degree and the average strength for $q > 50$ shows that event participation of already engaged members (ones who already attended few meetings) is more linked to the strength of social relations than to their number. This means that at the beginning of their engagement in group activities, when the association is relatively small, the participation is conditioned by a number of members a person knows, while later, when the association becomes stronger, the intensity of relations with already known members becomes more important.

This finding is further supported if we consider the change of the average degree and strength with the number of participations. Fig 3 shows how the average member's degree and strength evolve with the number of participations in group's events. At the beginning, the degree and strength have the same value and grow at the same rate, but after only few participations the strength becomes larger than the degree, and starts to grow much faster for members of all four Meetup communities. After 100 attended events the average strength of a member is up to ten times larger than her degree (see Fig E in S1 File). This indicates that the event participation dynamics is mostly governed by the need of a member to maintain and strengthen her relationships with already known members of a community. As a matter of fact, our analysis of member's embeddedness in these social networks shows that members maintain strong relations with single members of the community, but also with small sub-groups of members. A comparison with randomized data (Figs E and F in S1 File) reveals that both the degree and strength grow slower with the number of events, and that their ratio is

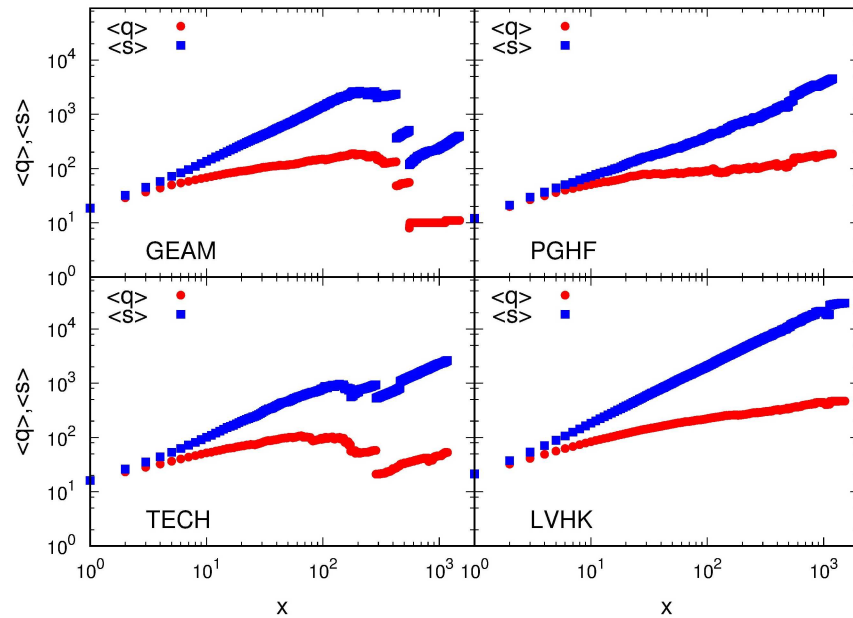


Fig 3. Event driven evolution of member's degree and strength. Dependence of member's average degree $\langle q \rangle$ and strength $\langle s \rangle$ on number of attended group events by member x for four considered Meetup groups.

doi:10.1371/journal.pone.0171565.g003

higher than in the original data. Relatively high value of the average clustering coefficient $\langle c_i \rangle$, shown in Fig 4 indicates that there is a high probability (more than 10% on average) that friends of a member also form significant relationships. The slow decay of $\langle c_i \rangle$ with the number of participations and the fact that it remains relatively large (above 0.2) even for participants with a thousand of attended meetups, Fig 4, show that personal networks of members have tendency to remain clustered, i.e., have relatively high number of closed triplets compared to random networks.

We now further examine the structure of these triplets and its change with the number of participations by calculating the averaged weighted clustering coefficient. The weighted clustering coefficient c_i^W measures the local cohesiveness of personal networks by taking into account the intensity of interactions between local triplets [38]. This measure does not just take into account a number of closed triplets of a node i but also their total relative weight with respect to the total strength of the nodes (see Methods). We also examine how the value of weighted clustering coefficient, averaged over all participants that have attended x events, designated as $\langle c_i^W(x) \rangle$, with the number of attended events. As it is shown in Fig 4, a member's network of personal contacts shows high level of cohesiveness, on the average. Like its non-weighted counterpart, the value of $\langle c_i^W \rangle$ only slightly decreases during member's early involvement in group activities, while later it remains constant and independent of the number of participations. A comparison of the values of weighted and non-weighted clustering coefficients reveals the role of strong relationships in local networks, i.e., whether they form triplets or bridges between different cohesive groups [38]. At the beginning of member's involvement in a group, these two coefficients have similar values, Fig 4, which indicates that the cohesiveness of a subgroup of personal contacts is not that important for the early participation dynamics. As a number of attended events grows, as well as a number and strength of personal contacts, the weighted clustering coefficient becomes larger than its non-weighted counterpart, indicating member's strongest ties with other members who are also friends. The fact that in later

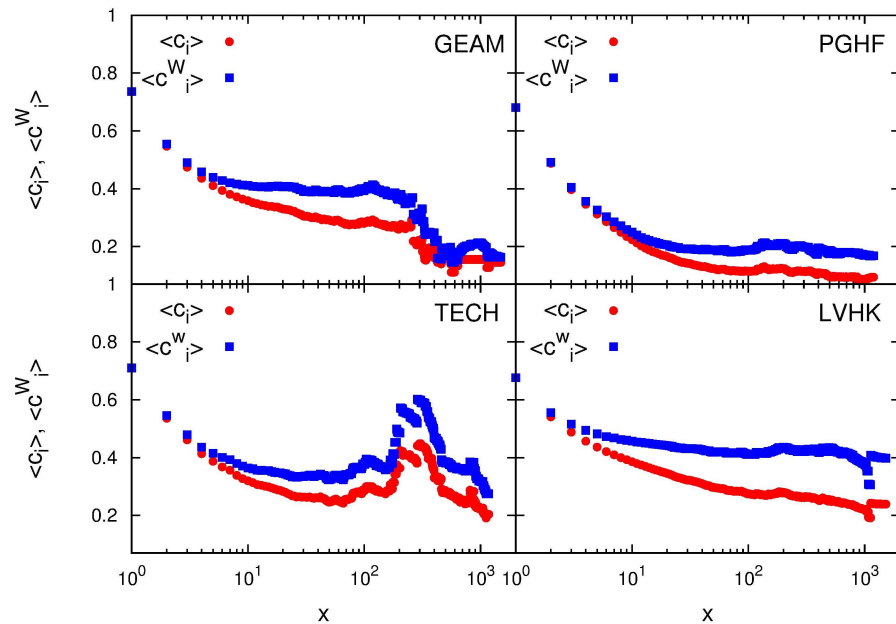


Fig 4. Local cohesiveness of social networks of significant links. Evolution of local cohesiveness of members personal networks, measured by averaged non-weighted $\langle c_i \rangle$ and weighted clustering coefficients $\langle c_i^w \rangle$, with the number of events attended by the member x .

doi:10.1371/journal.pone.0171565.g004

engagement the weighted clustering coefficient is larger than its non-weighted counterpart indicates that the clustering has an important role in the network organization of Meetup groups and thus in the group participation dynamics [38]. Low and very similar values of the clustering and weighted clustering coefficients in networks obtained for randomized data (Fig G in S1 File) further confirm our conclusion about the importance of clustering in the event participation dynamics. The observed discontinuity and decrease of values of the degree, strength and both clustering coefficients, Figs 3 and 4, for groups GEAM and TECH are consequences of a small number of members who attended more than 300 events.

Event importance in group participation dynamics

In our previous work [20], we have shown that the conference participation dynamics is independent of the conference topic, type and size. The same holds true for the Meetup participation dynamics, i.e., the member’s participation patterns in the Meetup group activities do not depend on the group size, category, location or type of activity. However, the size of group events and their time order may influence the structure of network and thus group dynamics. We explore how topological properties of networks, specifically the number of acquaintances and network cohesion, change after the removal of events according to a certain order (see Methods for details).

Firstly, we study how the removal of events according to a certain order influences the number of overall acquaintances in the network. For this purpose we define a measure η (see Methods), which we use to quantify the percentage of the remaining significant acquaintances after the removal of an event. Fig 5 shows the change of measure η after the removal of a fraction r of events according to a chosen strategy. We see that most of new significant connections are usually made during the largest events. The importance of large events for the creation of new acquaintances is especially striking for the groups GEAM, PGHF, and TECH, where

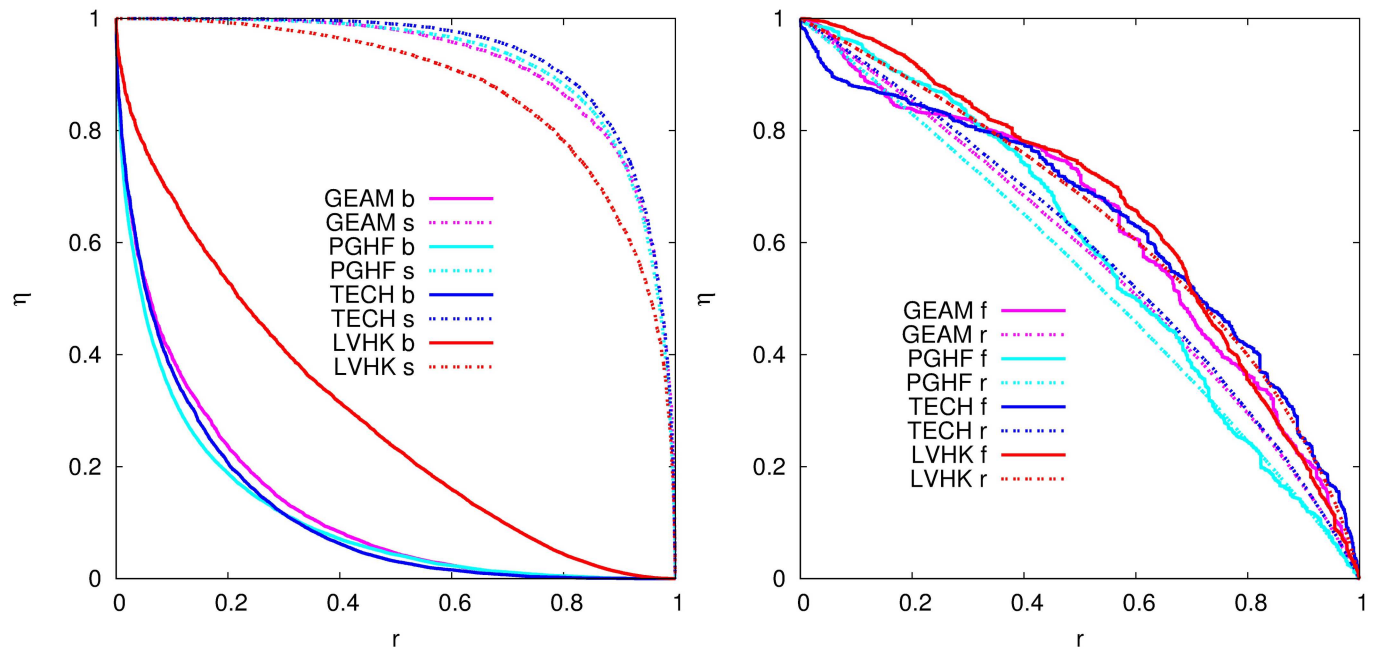


Fig 5. Importance of event size for number of distinctive links in the networks. Change of η with removal of events according to their size (left) and temporal and random order (right). Abbreviations indicate order in which we remove events: **b**—from the largest to the smallest, **s**—from the smallest to the largest, **f**—from the first to the last and **r**—random.

doi:10.1371/journal.pone.0171565.g005

about 80% of acquaintances only met at top 20% of events. For LVHK the decrease is slower, probably due to a difference in the event size fluctuations (see Fig E in S1 File), but still more than 50% of acquaintances disappear if we remove top 40% of events, which is still much higher percentage of contacts compared to random removal of events (see Fig 5 (right)). Similar results are observed when we remove events in the opposite order, Fig 5 (left). Only 20% of acquaintances are being destroyed after the removal of 80% of events, for all four groups. This indicates that new and weak connections are usually formed during large events, while these acquaintances are further strengthened during small meetups. On the other hand, the removal of events according to their temporal order, Fig 5, has very similar effect as random removal, i.e., the value of parameter η decreases gradually as we remove events.

Similar conclusions can be drawn based on the change of average weighted clustering coefficient $\langle C^W \rangle$ (now averaged over all nodes in the network) with the removal of events, Fig 6. Removal of events according to decreasing order of their sizes, does not result in the significant change of $\langle C^W \rangle$. The same value of weighted clustering coefficient, observed even after the removal of 80% of events, shows that small events are not attended by a pair of but rather by a group of old friends. On the other hand, the removal of events in the opposite order results in gradual decrease of $\langle C^W \rangle$. A certain fraction of triads in networks are made by at least one link of low weight. These links are most likely to vanish after the removal of the largest events, which results in the gradual decrease of $\langle C^W \rangle$. Removal of events according to their temporal order results in the change of $\langle C^W \rangle$ similar to one obtained for random removal of events, confirming further that the time ordering of events does not influence the structure of studied networks.

Discussion and conclusion

In this article we explore the event participation dynamics and underlying social mechanism of the Meetup groups. The motivation behind this was to further explore the event driven

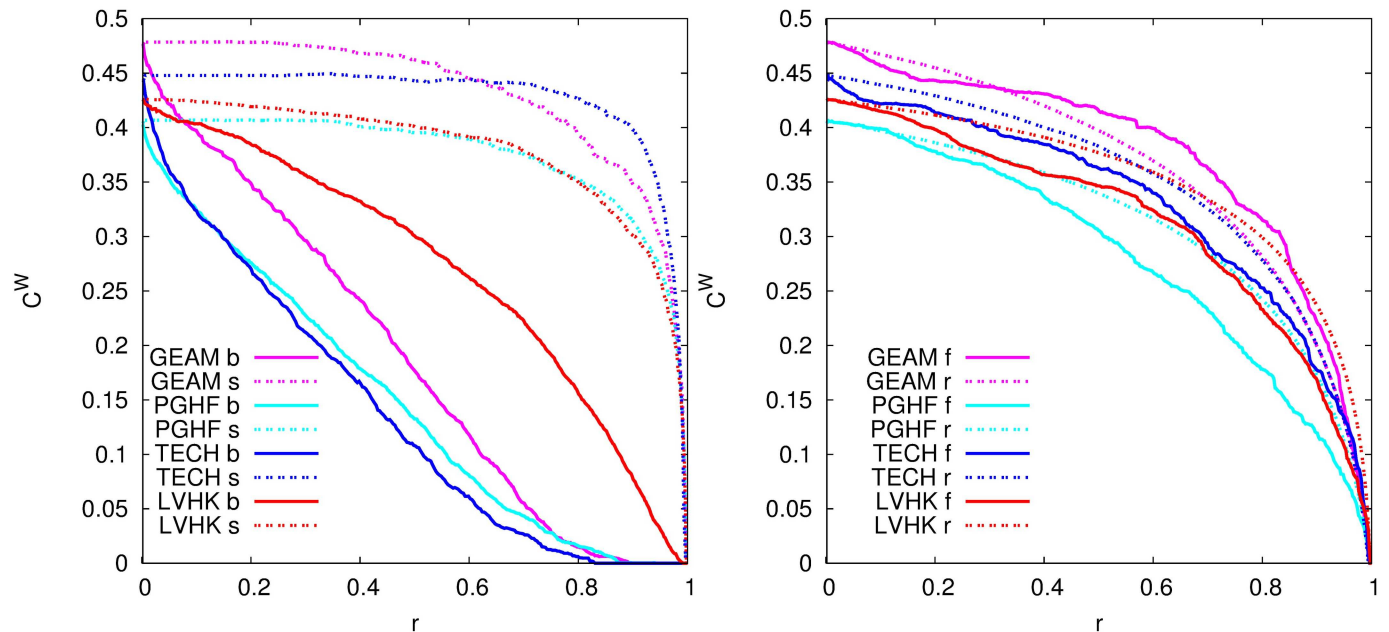


Fig 6. Importance of event size for the network cohesiveness. Change of local network cohesiveness with removal of events according to their size (left) and temporal and random order (right). Abbreviations indicate order in which we remove events: **b**—from the largest to the smallest, **s**—from the smallest to the largest and **r**—random.

doi:10.1371/journal.pone.0171565.g006

dynamics, work we have started by exploring participation patterns of scientists at scientific conferences [20], and to better examine the social origins behind the repeated attendance at group events, which was not feasible with the conference data. The results in this manuscript are based on empirical analysis of participation patterns and topological characteristics of networks for four different Meetup groups made up of people who have different motives and readiness to participate in group activities: GEAM, PGHF, TECH, LVHK.

Although these four groups differ in category and type of activity, we have shown that they are all characterized with similar participation patterns: the probability distributions of total number of participations, number of successive participations and time lag between two successive participations follow a power law and truncated power law behavior, with the value of power law exponents between 1 and 3. The resemblance of these patterns to those observed for conference participations [20] indicates that these two, seemingly different, social system dynamics are governed by similar mechanism. This means that the probability for a member to participate in future events depends non-linearly on the balance between the numbers of previous participations and non-participations. As in the case of conferences [20], this behavior is independent of the group category, size, or location, meaning that members association with the community of a Meetup group strongly influence their event participation patterns, and thus the frequency and longevity of their engagement in the group activities.

The Member's association with the community is primarily manifested through her interconnectedness with other members of a specific Meetup group, i.e., in the structure of her personal social network. We have examined topological properties of filtered weighted social networks constructed from the members event co-occurrence. Through network filtering we have emphasized the importance of significant links, the ones which are not the result of coincidence but rather an indicator of social relations. The analysis of local topological properties of these networks has revealed that the strength of connectedness with the community, for the members with small number of participations, is predominantly the consequence of the width

of their social circles. Average strength and degree of members with $q \lesssim 50$, which on average corresponds to only a few participations, are equal, while the strength of members who know more than 50 people and have participated in more than a few events, is several times higher than their degree. This means that after a few participations strengthening of existing ties becomes more important than meeting new people. These arguments are further extended with our observation of the evolution of average strength and degree with the growth of number of participations. Both, average degree and strength, grow, but the growth rate of strength is higher than one of the degree, for all four Meetup groups. All four groups are characterised with very high cohesiveness of their social communities. The evolutions of clustering coefficients, non and weighted one, and their ratio, show that bonding with the community becomes more important as the members' engagement in the group activity progresses. As in the case of conference participations, frequent attendees of group activities tend to form a core whose stability grows with the number of participations [20, 39]. The need of frequent attendees to maintain and increase their bonding with the rest of the community influences their probability to attend future meetings and thus governs the event participation dynamics of the Meetup groups.

The observed structure of personal social networks of the Meetup members is in accordance with previous research on this topic [21–24]. The average size of personal social networks for the most frequent attendees of the Meetup groups GEAM, PGHF, and TECH, is 150 or lower, while the size of the LVHK personal network is less than 500 different connections, i.e., of the same order. This is consistent with the predictions of the Social Brain Hypothesis for the typical human group size. The faster growth of the strength, compared to the one observed for degree, and the constant, non-trivial, value of the clustering coefficients are indicators of the layered structure of social networks. The comparable values of strength and degree, as well as weighted and non-weighted clustering coefficients, observed for small numbers of attendances, indicate that at the beginning all social connections are of the equal importance. As members' engagement with the community grows, she begins to interact with a certain members of the group more often, which results in the non-linear growth of her strength. The higher value of weighted clustering coefficient, compared to its non-weighted counterpart, indicates that member's personal network consists of layers, subgroups of members, characterized with similar strength of mutual relations.

While the group category, type of activity and size do not significantly affect the participation dynamics in the group activities and structure of networks, the size of separate events does have an influence on the evolution of social networks. Large events represent an opportunity for members to make new acquaintances, i.e., to establish new connections. On the other hand, small meetings are typically the gatherings of members with preexisting connections, and their main purpose is to facilitate the stronger bonding among group members. We find that the time order of events is irrelevant for group dynamics.

The universality of the event participation patterns, shown in this and previous work [20], and its socially driven nature give us a better insight not only about the dynamics of studied social communities but also about others which are organised on very similar principles: communities that bring together people with the similar interests and where the participation is voluntary. Having in mind that these type of groups constitute a large part of human life, including all life aspects, understanding their functioning and dynamics is of great importance. Our results not only contribute to the corpus of increasing knowledge, but also indicate the key factor which influences the group longevity and successful functioning: the association of group members with the community. This and recent success stories [40] suggest that complex network theory can be an extremely useful tool in creating successful communities. Future studies will be conducted towards further confirmation of universality of event participation

patterns and better understanding of how social association and contacts can be used for creating conditions for successful functioning of learning and health support groups.

Materials and methods

Data

There are more than 240000 groups in 181 countries classified into 33 categories active in the Meetup community [41]. For each of selected four groups, we have used the Meetup public API to access the data and collect the list of events organized by the group and the information on the members who confirmed their participation (RSVP) in the given event since the group's beginnings. Each member has a unique id which enables us to follow her activity in the group events during the time. The collected data have been fully anonymised and we did not collect any personal information about the group members. We have complied with terms of use of Meetup website. More details about the group sizes and the number of events is given in [Table 1](#).

Network construction and filtering

Network construction. We start with a bipartite member event network, which we represent with participation matrix B . Let N_m denotes total number of members in the group and N_e is total number of events organized by the group. If the member i participated in the event l element of matrix B_{il} takes a value 1, otherwise $B_{il} = 0$. In the bipartite network created in this way, members' degree is equal to total number of events member participated in, while events' degree is defined as total number of members that have attended that event. The social network, which is the result of members interactions during the Meetup events and is represented by weighted matrix W , is created from the weighted projection of bipartite network to members partition [42, 43]. In the obtained weighted network nodes correspond to individual members while the value of the element of weighted matrix W_{ij} corresponds to number of common events two members have attended together.

Network filtering. The observed weighted network is dense network where some of the non-zero edges can be the result of coincidence. For instance, these edges can be found between members who attended large number of events or events with many participants, and therefore they do not necessarily indicate social connections between members. The pruning of these type of networks and separation of significant edges from non-significant ones is not a trivial task [36, 37, 44]. For this reasons we start from bipartite network and use method that determines the significance of W_{ij} link based on configuration model of random bipartite networks [36, 37, 45, 46]. In this model of random networks the event size and the number of events a member attended are fixed, while all other correlations are destroyed (see SI for further explanations). Based on this model, for each link in bipartite network, B_{il} , we determine the probability p_{il} that user i has attended event l . The assumption of uncorrelated network enables us to also estimate the probability that two members, i and j , have attended the same event, which is equal to $p_{il}p_{jl}$. Probability that two members have attended the same w events is then given by Poisson binomial distribution

$$P_{ij}(w) = \sum_{M_w} \prod_{l \in M_w} p_{il}p_{jl} \prod_{l \notin M_w} (1 - p_{il}p_{jl}) \quad (1)$$

where M_w is the subset of w events that can be chosen from given M events [36, 37, 47]. We define p -value as probability that two members i and j has co-occurred on at least w_{ij} events,

i.e., that the link weight between these two members is w_{ij} or higher

$$p\text{-value}(w_{ij}) = \sum_{w \geq w_{ij}} P_{ij}(w). \tag{2}$$

The relationship between users i and j will be considered statistically significant if $p\text{-value}(w_{ij}) \leq p_{trs}$. In our case, threshold $p_{trs} = 0.05$. All links with $p\text{-value}(w_{ij}) > p_{trs}$ are consequence of chance and are considered as non-significant and thus removed from the network. This way we obtain weighted social network of significant relations between members of the Meetup group W_{ij}^S . The details on how we estimate p_{ij} and $P_{ij}(w)$ for each link are given in SI.

Topological measures. All topological measures considered in this work are calculated for weighted social network of significant relations W_{ij}^S . We consider the following topological measures of the nodes:

- The node degree $q_i = \sum_j \mathcal{H}(W_{ij}^S)$, where \mathcal{H} is Heaviside function ($\mathcal{H}(x) = 1$ if $x > 0$ otherwise $\mathcal{H}(x) = 0$);
- The node strength $s_i = \sum_j W_{ij}^S$ [7];
- Non-weighted clustering coefficient of the node $c_i = \frac{1}{q_i(q_i-1)} \sum_{j,m} \mathcal{H}(W_{ij}^S) \mathcal{H}(W_{im}^S) \mathcal{H}(W_{jm}^S)$ [7].
- Weighted clustering coefficient of the node $c_i^W = \frac{1}{s_i(q_i-1)} \sum_{j,m} \frac{W_{ij}^S + W_{im}^S}{2} \mathcal{H}(W_{ij}^S) \mathcal{H}(W_{im}^S) \mathcal{H}(W_{jm}^S)$ [38].

Weighted clustering coefficient of the network $\langle C^W \rangle$ and its non-weighted counterpart $\langle C \rangle$ are values averaged over all nodes in the network.

The event relevance

In order to explore the relevance of event size and time ordering for the evolution of social network topology we analyze how removal of events, according to specific ordering, influences the number of acquaintance and network cohesion. Specifically, we observe change of measure η , which represents the fraction of the remaining acquaintances, and weighted clustering coefficient $\langle C^W \rangle$ after the removal of a fraction r of events. The removal of event results in change of link weights between group members. For instance, if two members, i and j , have participated in event l , the removal of this event will result in the decrease of the link weight W_{ij}^S by one. Further removal of events in which these two members have co-occurred will eventually lead to termination of their social connection, i.e., $W_{ij}^S = 0$. If $W^S(r)$ is the matrix of link weights after the removal of a fraction r of events and W^S is the original matrix of significant relations, then the value of parameter η after the removal of r events is calculated as

$$\eta(r) = \frac{\sum_{ij} \mathcal{H}(W_{ij}^S(r))}{\sum_{ij} \mathcal{H}(W_{ij}^S)}, \tag{3}$$

The value of weighted clustering coefficient $\langle C^W \rangle$ after the removal of a fraction r of events is calculated using the same formula as for the $\langle C^W \rangle$ just using the value of $W^S(r)$ instead of W^S .

We remove events according to several different strategies:

- We sort events according to their size. Then, we remove sorted events in descending and ascending order.
- We remove events according to their time-order, from the first to the last.

- We remove events in random order. We perform this procedure for each list of events 100 times.

Supporting information

S1 File. Supplementary information: Associative nature of event participation dynamics: a network theory approach. The probability distribution $P(x)$ of total numbers of participations in group events x , obtained from the empirical data for the four selected Meetup groups (blue circles). We also show truncated power law fit $x^{-\alpha}e^{-Bx}$ (solid lines), power law fit $x^{-\gamma}$ (dotted-dashed lines), and exponential fit $e^{-\lambda x}$ (dotted lines). **Fig A** Log likelihood ratio \mathcal{R} and the π -value compare fits to the power law and fits to the truncated power law for the probability distribution of total numbers of participations in group events. **Table A** The probability distribution of successive numbers of participations in group events x_s , for the four selected Meetup groups. The probability distribution follows power law behavior $P(x_s) \sim x_s^{-\gamma}$. **Fig B** The probability distribution of time lags between two successive participations in group events y_s , for the four selected Meetup groups. The probability distribution follows truncated power law behavior $P(y_s) \sim y_s^{-\alpha}e^{-By_s}$. **Fig C** The probability distribution of link weights in a weighted network before and after filtering, for the four selected Meetup groups. **Fig D** The dependence of a degree strength ratio on the number of participations, averaged over all members for the four considered Meetup groups. Red circles correspond to results obtained from empirical data, while blue squares correspond to randomized data. **Fig E** The dependence of group members' average degree $\langle q \rangle$ and strength $\langle s \rangle$ on numbers of participations for a real weighted network and a randomized network. **Fig F** The dependence of group members' average non-weighted $\langle c_i \rangle$ and weighted clustering coefficient $\langle c_i^W \rangle$ on numbers of participations for a real weighted network and a randomized network. **Fig G** The probability distribution of relative size fluctuations $\frac{\langle e \rangle - e}{\langle e \rangle}$, for the four considered Meetup groups, where e is the event size and $\langle e \rangle$ is the average event size. **Fig H.**
(PDF)

Acknowledgments

Numerical simulations were run on the PARADOX supercomputing facility at the Scientific Computing Laboratory of the Institute of Physics Belgrade.

Author contributions

Conceptualization: JS MMD.

Data curation: JS.

Formal analysis: JS MMD.

Investigation: JS.

Methodology: JS MMD.

Visualization: JS MMD.

Writing – original draft: JS MMD.

References

1. Castellano C, Fortunato S, Loreto V. Statistical physics of social dynamics. *Rev Mod Phys.* 2009; 81:591–646. doi: [10.1103/RevModPhys.81.591](https://doi.org/10.1103/RevModPhys.81.591)

2. Nowak MA. Five Rules for the Evolution of Cooperation. *Science*. 2006; 314(5805):1560–1563. doi: [10.1126/science.1133755](https://doi.org/10.1126/science.1133755) PMID: [17158317](https://pubmed.ncbi.nlm.nih.gov/17158317/)
3. Fowler JH, Christakis NA. Cooperative behavior cascades in human social networks. *Proc Natl Acad Sci USA*. 2010; 107(12):5334–5338. doi: [10.1073/pnas.0913149107](https://doi.org/10.1073/pnas.0913149107) PMID: [20212120](https://pubmed.ncbi.nlm.nih.gov/20212120/)
4. Granovetter MS. The Strength of Weak Ties. *Am J Sociol*. 1973; 78(6):1360–1380. doi: [10.1086/225469](https://doi.org/10.1086/225469)
5. Pastor-Satorras R, Castellano C, Van Mieghem P, Vespignani A. Epidemic processes in complex networks. *Rev Mod Phys*. 2015; 87:925–979. doi: [10.1103/RevModPhys.87.925](https://doi.org/10.1103/RevModPhys.87.925)
6. Mitrović Dankulov M, Melnik R, Tadić B. The dynamics of meaningful social interactions and the emergence of collective knowledge. *Sci Rep*. 2015; 5:12197. doi: [10.1038/srep12197](https://doi.org/10.1038/srep12197)
7. Boccaletti S, Latora V, Moreno Y, Chavez M, Hwang DU. Complex networks: Structure and dynamics. *Phys Rep*. 2006; 424(4–5):175–308. doi: [10.1016/j.physrep.2005.10.009](https://doi.org/10.1016/j.physrep.2005.10.009)
8. Holme P, Saramäki J. Temporal networks. *Phys Rep*. 2012; 519(3):97–125. doi: [10.1016/j.physrep.2012.03.001](https://doi.org/10.1016/j.physrep.2012.03.001)
9. Aral S, Walker D. Identifying Influential and Susceptible Members of Social Networks. *Science*. 2012; 337(6092):337–341. doi: [10.1126/science.1215842](https://doi.org/10.1126/science.1215842) PMID: [22722253](https://pubmed.ncbi.nlm.nih.gov/22722253/)
10. González-Bailón S, Borge-Holthoefer J, Moreno Y. Broadcasters and Hidden Influentials in Online Protest Diffusion. *Am Behav Sci*. 2013; 57(7):943–965. doi: [10.1177/0002764213479371](https://doi.org/10.1177/0002764213479371)
11. 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.
12. Mitrović M, Paltoglou G, Tadić B. Quantitative analysis of bloggers' collective behavior powered by emotions. *J Stat Mech*. 2011; 2011(02):P02005.
13. Garas A, Garcia D, Skowron M, Schweitzer F. Emotional persistence in online chatting communities. *Sci Rep*. 2012; 2:402. doi: [10.1038/srep00402](https://doi.org/10.1038/srep00402) PMID: [22577512](https://pubmed.ncbi.nlm.nih.gov/22577512/)
14. Török J, Iñiguez G, Yasserli T, San Miguel M, Kaski K, Kertész J. Opinions, Conflicts, and Consensus: Modeling Social Dynamics in a Collaborative Environment. *Phys Rev Lett*. 2013; 110:088701. doi: [10.1103/PhysRevLett.110.088701](https://doi.org/10.1103/PhysRevLett.110.088701) PMID: [23473207](https://pubmed.ncbi.nlm.nih.gov/23473207/)
15. Yasserli T, Sumi R, Rung A, Kornai A, Kertész J. Dynamics of Conflicts in Wikipedia. *PLoS ONE*. 2012; 7(6):1–12. doi: [10.1371/journal.pone.0038869](https://doi.org/10.1371/journal.pone.0038869)
16. 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 Educ Couns*. 2001; 45:195–198. doi: [10.1016/S0738-3991\(01\)00121-5](https://doi.org/10.1016/S0738-3991(01)00121-5) PMID: [11722855](https://pubmed.ncbi.nlm.nih.gov/11722855/)
17. Davison KP, Pennebaker JW, Dickerson SS. Who talks? The social psychology of illness support groups. *Am Psychol*. 2000; 55:205–217. doi: [10.1037/0003-066X.55.2.205](https://doi.org/10.1037/0003-066X.55.2.205) PMID: [10717968](https://pubmed.ncbi.nlm.nih.gov/10717968/)
18. Tam Cho WK, Gimpel JG, Shaw DR. The Tea Party Movement and the Geography of Collective Action. *Q J Polit Sci*. 2012; 7:105–133. doi: [10.1561/100.00011051](https://doi.org/10.1561/100.00011051)
19. Weinberg BD, Williams CB. The 2004 US Presidential campaign: Impact of hybrid offline and online 'meetup' communities. *J Direct Data Digit Mark Pract*. 2006; 8(1):46–57. doi: [10.1057/palgrave.ddmp.4340552](https://doi.org/10.1057/palgrave.ddmp.4340552)
20. Smiljanić J, Chatterjee A, Kauppinen T, Mitrović Dankulov M. A Theoretical Model for the Associative Nature of Conference Participation. *PLoS ONE*. 2016; 11(2):1–12.
21. Dunbar RIM. Coevolution of neocortical size, group size and language in humans. *Behavioral and Brain Sciences*. 1993; 16(4):681–694. doi: [10.1017/S0140525X00032325](https://doi.org/10.1017/S0140525X00032325)
22. Dunbar RIM. Mind the Gap: Or Why Humans Aren't Just Great Apes. *Proceedings of the British Academy*. 2008; 154:403–423.
23. Hill RA, Dunbar RIM. Social network size in humans. *Human Nature*. 2003; 14(1):53–72. doi: [10.1007/s12110-003-1016-y](https://doi.org/10.1007/s12110-003-1016-y) PMID: [26189988](https://pubmed.ncbi.nlm.nih.gov/26189988/)
24. Dunbar RIM. Constraints on the evolution of social institutions and their implications for information flow. *Journal of Institutional Economics*. 2011; 7(3):345–371. doi: [10.1017/S1744137410000366](https://doi.org/10.1017/S1744137410000366)
25. Sessions LF. How offline gatherings affect online communities. *Information, Communication & Society*. 2010; 13(3):375–395. doi: [10.1080/13691180903468954](https://doi.org/10.1080/13691180903468954)
26. Hristova D, Quattrone G, Mashhadi A, Capra L. The Life of the Party: Impact of Social Mapping in OpenStreetMap. In: *Proceedings of the Seventh International AAAI Conference on Web and Social Media*. ICWSM'13; 2013. p. 234–243.

27. 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.
28. 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.
29. Pham TAN, Li X, Cong G, Zhang Z. A general graph-based model for recommendation in event-based social networks. In: 2015 IEEE 31st International Conference on Data Engineering; 2015. p. 567–578.
30. Macedo AQ, Marinho LB, Santos RLT. Context-Aware Event Recommendation in Event-based Social Networks. In: Proceedings of the 9th ACM Conference on Recommender Systems. RecSys'15; 2015. p. 123–130.
31. 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.
32. Jiang JY, Li CT. Analyzing Social Event Participants for a Single Organizer. In: International AAAI Conference on Web and Social Media; 2016. p. 599–602.
33. McCully W, Lampe C, Sarkar C, Velasquez A, Sreevivasan A. Online and Offline Interactions in Online Communities. In: Proceedings of the 7th International Symposium on Wikis and Open Collaboration. WikiSym'11; 2011. p. 39–48.
34. Palla G, Barabási AL, Vicsek T. Quantifying social group evolution. *Nature*. 2007; 446(7136):664–667. doi: [10.1038/nature05670](https://doi.org/10.1038/nature05670) PMID: [17410175](https://pubmed.ncbi.nlm.nih.gov/17410175/)
35. Backstrom L, Huttenlocher D, Kleinberg J, Lan X. Group Formation in Large Social Networks: Membership, Growth, and Evolution. In: Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. KDD'06; 2006. p. 44–54.
36. Dianati N. A maximum entropy approach to separating noise from signal in bimodal affiliation networks. *ArXiv e-prints*. 2016;.
37. Saracco F, Di Clemente R, Gabrielli A, Squartini T. Grandcanonical projection of bipartite networks. *ArXiv e-prints*. 2016;.
38. Barrat A, Barthélemy M, Pastor-Satorras R, Vespignani A. The architecture of complex weighted networks. *Proc Natl Acad Sci USA*. 2004; 101(11):3747–3752. doi: [10.1073/pnas.0400087101](https://doi.org/10.1073/pnas.0400087101) PMID: [15007165](https://pubmed.ncbi.nlm.nih.gov/15007165/)
39. Van Dijk J, Maier G. ERSa Conference participation: does location matter? *Pap Reg Sci*. 2006; 85(4):483–504. doi: [10.1111/j.1435-5957.2006.00102.x](https://doi.org/10.1111/j.1435-5957.2006.00102.x)
40. Cosgrave P. Engineering Serendipity: The Story of Web Summit's Growth; 2014. Available from: <https://goo.gl/H3aWMI>.
41. Meetup Datasets;. Available from: <https://www.meetup.com/>.
42. Mitrović M, Tadić B. Bloggers behavior and emergent communities in Blog space. *Eur Phys J B*. 2010; 73(2):293–301. doi: [10.1140/epjb/e2009-00431-9](https://doi.org/10.1140/epjb/e2009-00431-9)
43. Mitrović M, Paltoglou G, Tadić B. Networks and emotion-driven user communities at popular blogs. *Eur Phys J B*. 2010; 77(4):597–609. doi: [10.1140/epjb/e2010-00279-x](https://doi.org/10.1140/epjb/e2010-00279-x)
44. Dianati N. Unwinding the hairball graph: Pruning algorithms for weighted complex networks. *Phys Rev E*. 2016; 93:012304. doi: [10.1103/PhysRevE.93.012304](https://doi.org/10.1103/PhysRevE.93.012304) PMID: [26871089](https://pubmed.ncbi.nlm.nih.gov/26871089/)
45. Saracco F, Di Clemente R, Gabrielli A, Squartini T. Randomizing bipartite networks: the case of the World Trade Web. *Sci Rep*. 2015; 5:10595. doi: [10.1038/srep10595](https://doi.org/10.1038/srep10595) PMID: [26029820](https://pubmed.ncbi.nlm.nih.gov/26029820/)
46. Cellai D, Bianconi G. Multiplex networks with heterogeneous activities of the nodes. *Phys Rev E*. 2016; 93:032302. doi: [10.1103/PhysRevE.93.032302](https://doi.org/10.1103/PhysRevE.93.032302) PMID: [27078361](https://pubmed.ncbi.nlm.nih.gov/27078361/)
47. Liebig J, Rao A. Fast extraction of the backbone of projected bipartite networks to aid community detection. *Europhys Lett*. 2016; 113(2):28003. doi: [10.1209/0295-5075/113/28003](https://doi.org/10.1209/0295-5075/113/28003)

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 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.

PACS: 05.65.+b, 89.75.Fb, 89.75.Hc

Keywords: complex networks, social groups, self-organization

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 fields 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 used 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

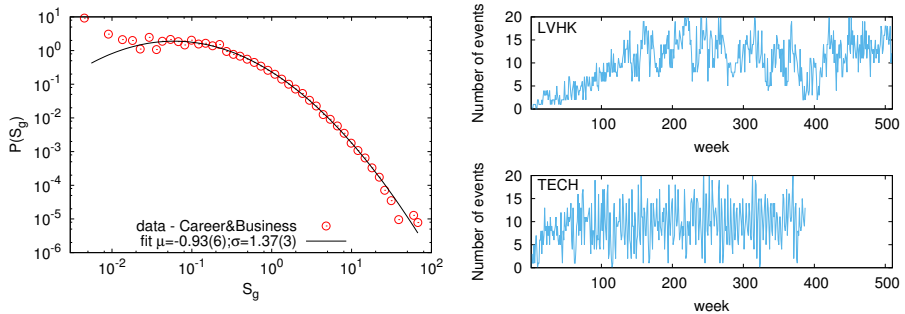


Figure 1. (left) Probability density distribution of group sizes S_g belonging to Meetup category *Career&Business*, which exhibits log-normal behaviour $f(S_g) = \frac{1}{S_g \sigma_g \sqrt{2\pi}} \exp\left(-\frac{(S_g - \mu_g)^2}{2\sigma_g^2}\right)$. Here S_g is group size rescaled with average group 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 identifier 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.

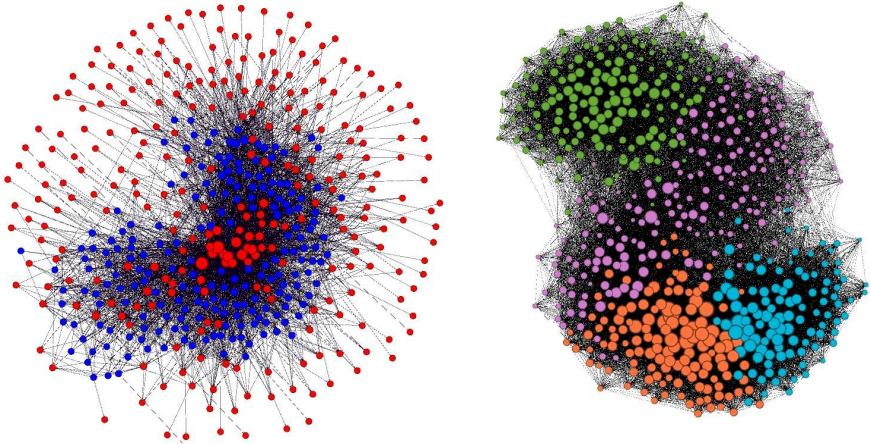


Figure 2. (left) A part of Meetup bipartite network of GEAM community: light gray circles represent nodes from Members partition while nodes from Events partition are represented with white circles. (right) Weighted network of significant social relations between the 10% most active members of LVHK group.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

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

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,

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

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 \mathcal{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

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

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

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

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

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

and

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

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

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

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

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

where Ew 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 $w_{i_1^M i_2^M}$ events

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

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.

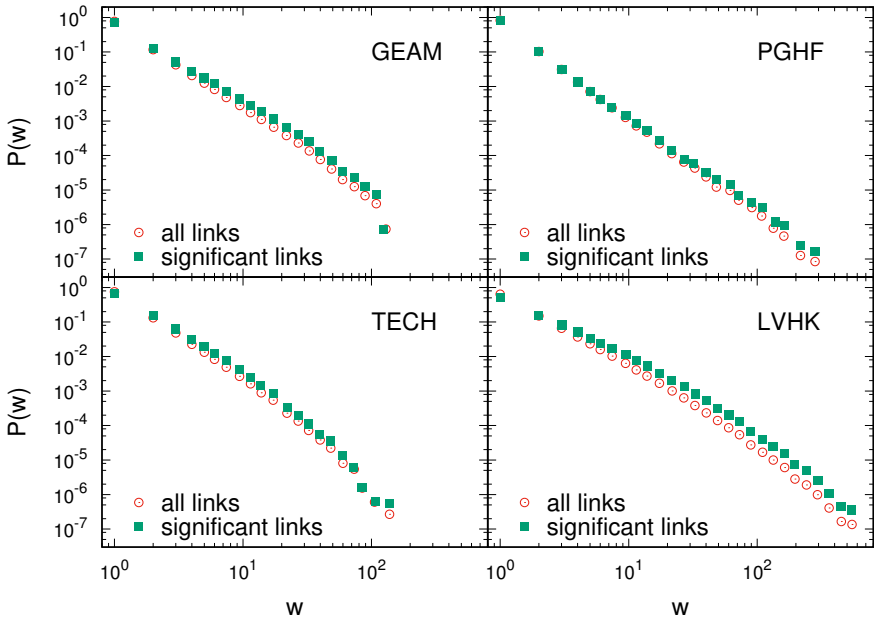


Figure 3. Probability distribution of link weights in a weighted network before and after filtering, for four considered Meetup groups. [Figure right reprinted with permissions from [27]]

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.

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

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.

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

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

$$\langle q_{nn}^\kappa \rangle = \frac{\sum_{j^{\kappa'}} q_{j^{\kappa'}}^{\kappa'}}{q^\kappa}. \quad (12)$$

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 $\langle q_{nn}^k \rangle$ 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 $\langle q_{nn}^k \rangle(q)$ 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

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

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

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

and weighted clustering coefficient

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

While the non-weighted clustering coefficient measures the number of closed triples formed by node and its neighbours, the $c_{i_1^M}^W$ 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]:

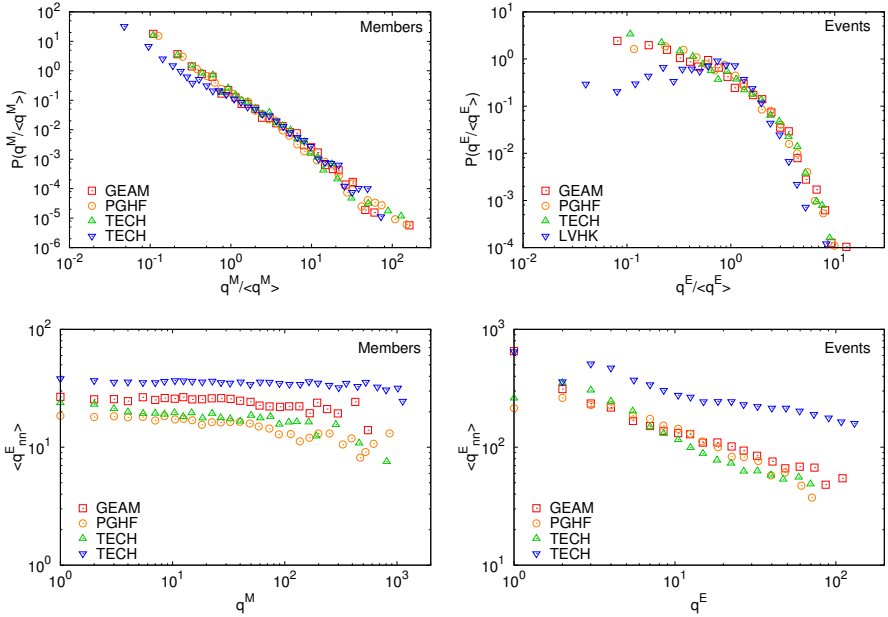


Figure 4. Probability density distribution of rescaled degree for nodes in Members (top left) and Events (top right) partition. The degree is rescaled by the value of an average degree in a given partition. The dependence of average nearest neighbor degree on degree in Members (bottom left) and Events (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

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

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

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

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.

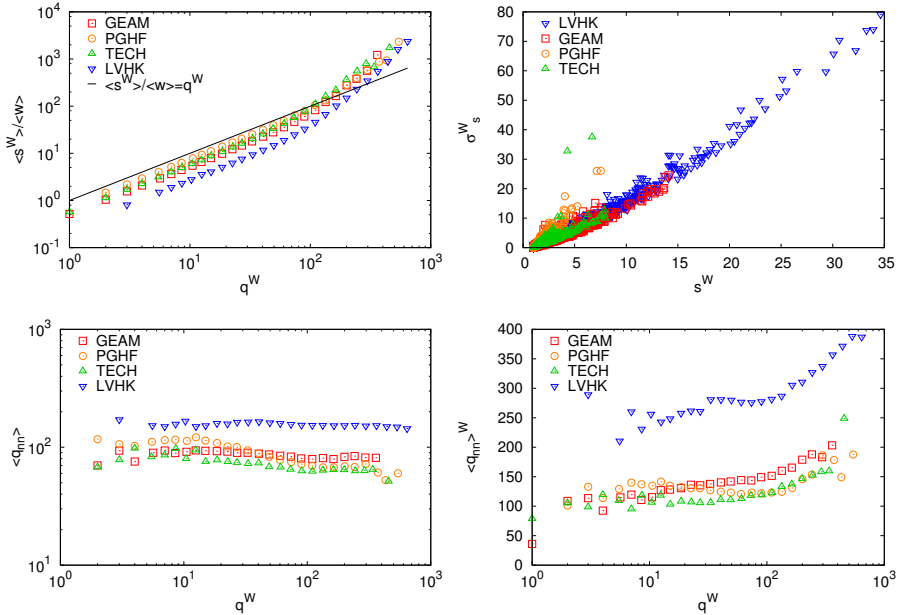


Figure 5. (top left) Dependence of normalized node strength on its degree. (top right) Scatter plot of nodes strength versus standard deviation of link weights adjacent to that node. Average nearest neighbor degree (bottom left) and its weighted counterpart (bottom right) in weighted network.

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 $\langle w_{i^M} \rangle = \frac{1}{q_{i^M}^M} \sum_{i_2^M} W_{i_1^M i_2^M}$ 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

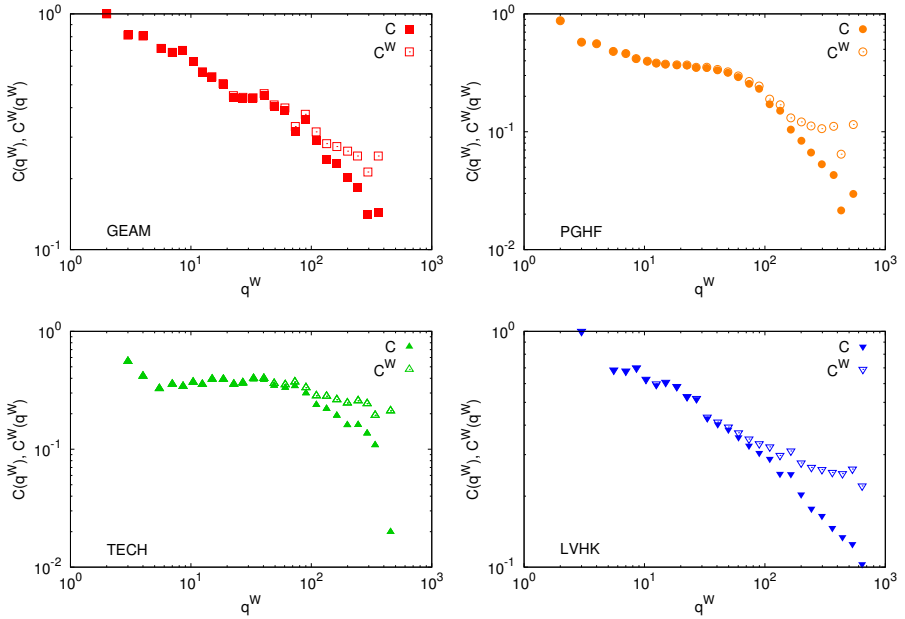


Figure 6. Dependence of non-weighted and weighted clustering coefficient on node degree for four studied Meetup social groups.

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.

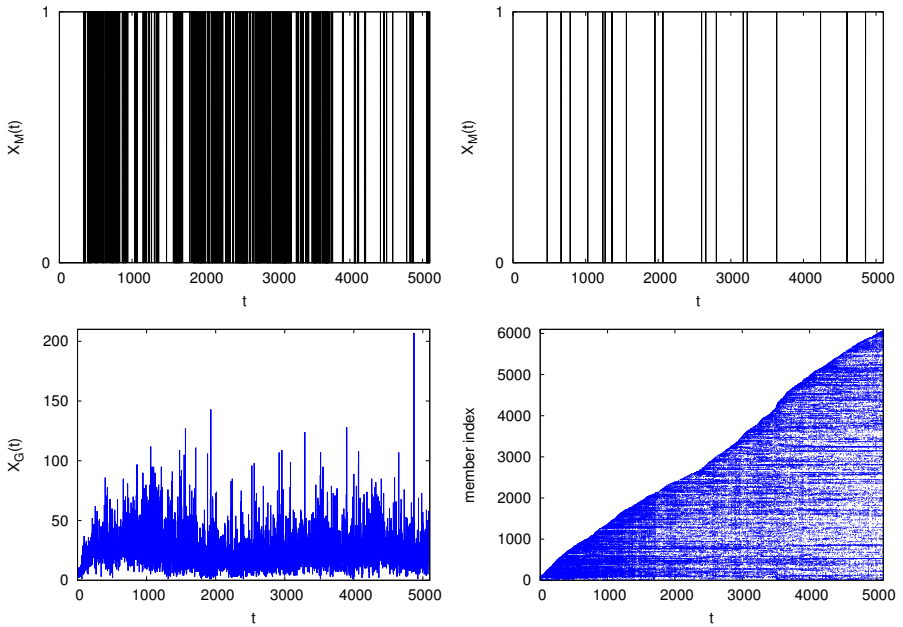


Figure 7. Time series of activity of one of the very active members (top left) and low active members (top right) of LVHK group. (bottom left) Time series of total activity in LVHK group. (bottom right) LVHK members activity pattern. The x axes in Fig.7 corresponds to event-time while y axes shows member's 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 $\{\Delta T_1^{i^M}, \dots, \Delta T_{q^M-1}^{i^M}\}$. 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).

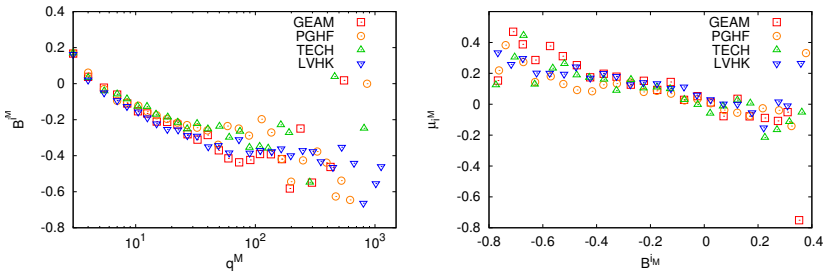


Figure 8. (left) Dependence of member’s burstiness on the number of attended events. (right) Dependence of memory coefficient on burstiness.

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

$$B_{i^M} = \frac{\langle \Delta T^{i^M} \rangle - \sigma_{\Delta T^{i^M}}}{\langle \Delta T^{i^M} \rangle + \sigma_{\Delta T^{i^M}}}, \quad (18)$$

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

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

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

where $\langle \Delta T^{i^M} \rangle_1$ ($\langle \Delta T^{i^M} \rangle_2$) and σ_1 (σ_2) are sample mean and sample standard deviation of $\Delta T_k^{i^M}$ ($\Delta T_{k+1}^{i^M}$) respectively.

We quantify the temporal patterns of individual member's activities by calculating the burstiness parameter B_{i^M} and memory coefficient μ_{i^M} [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_{i^M} 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-

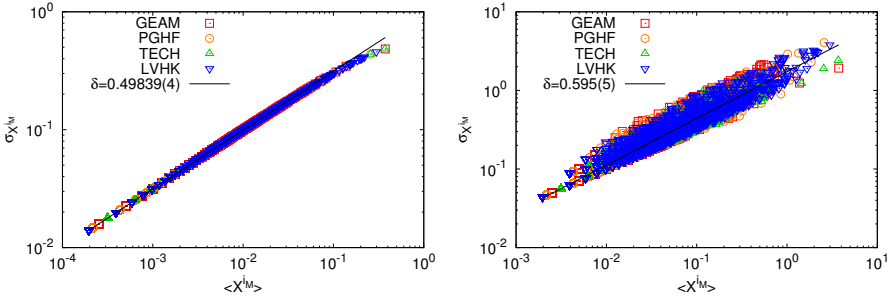


Figure 9. The relationship between fluctuations $\sigma_{X_{iM}}$ and average $\langle X_{iM} \rangle$ activity of member time series at the scale of one (left) and ten events (right).

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

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

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.

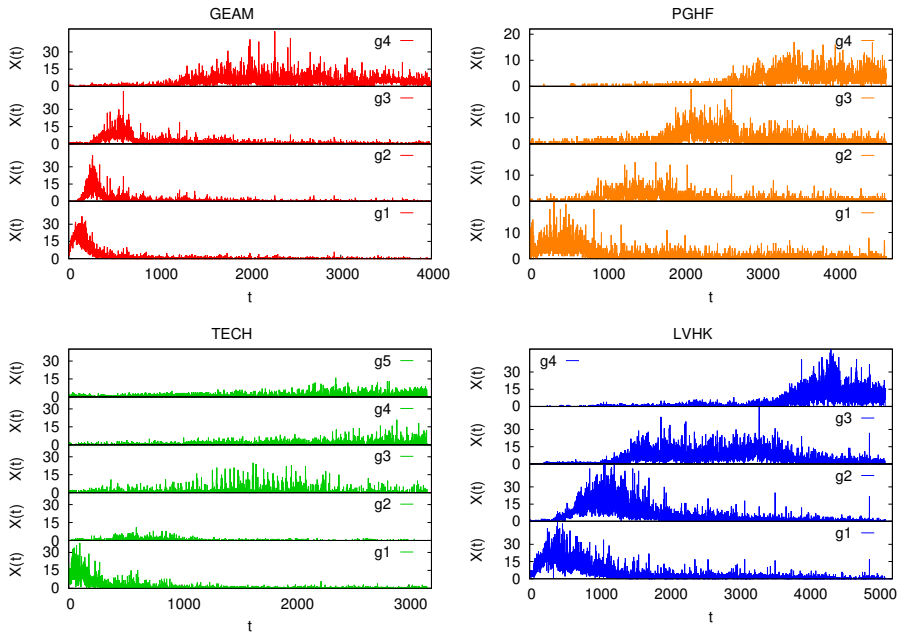


Figure 10. Total activity of subgroups found in weighted networks of the 10% most active members using LOUVAIN method.

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

This work was supported by the Ministry of Education, Science, and Technological Development of the Republic of Serbia under project ON171017. Numerical simulations were performed on the PARADOX supercomputing facility at the Scientific Computing Laboratory of the Institute of Physics Belgrade.

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.



MATLAB-based program for optimization of quantum cascade laser active region parameters and calculation of output characteristics in magnetic field[☆]



J. Smiljanić^{a,b,*}, M. Žeželj^{a,b}, V. Milanović^b, J. Radovanović^b, I. Stanković^a

^a Scientific Computing Laboratory, Institute of Physics Belgrade, University of Belgrade, Pregrevica 118, 11080 Belgrade, Serbia

^b School of Electrical Engineering, University of Belgrade, Bulevar Kralja Aleksandra 73, 11120 Belgrade, Serbia

ARTICLE INFO

Article history:

Received 21 March 2013

Received in revised form

21 October 2013

Accepted 23 October 2013

Available online 31 October 2013

Keywords:

Quantum cascade laser

Scattering

LO-phonons

Interface roughness

MATLAB program

ABSTRACT

A strong magnetic field applied along the growth direction of a quantum cascade laser (QCL) active region gives rise to a spectrum of discrete energy states, the Landau levels. By combining quantum engineering of a QCL with a static magnetic field, we can selectively inhibit/enhance non-radiative electron relaxation process between the relevant Landau levels of a triple quantum well and realize a tunable surface emitting device. An efficient numerical algorithm implementation is presented of optimization of GaAs/AlGaAs QCL region parameters and calculation of output properties in the magnetic field. Both theoretical analysis and MATLAB implementation are given for LO-phonon and interface roughness scattering mechanisms on the operation of QCL. At elevated temperatures, electrons in the relevant laser states absorb/emit more LO-phonons which results in reduction of the optical gain. The decrease in the optical gain is moderated by the occurrence of interface roughness scattering, which remains unchanged with increasing temperature. Using the calculated scattering rates as input data, rate equations can be solved and population inversion and the optical gain obtained. Incorporation of the interface roughness scattering mechanism into the model did not create new resonant peaks of the optical gain. However, it resulted in shifting the existing peaks positions and overall reduction of the optical gain.

Program summary

Program title: QCL

Catalogue identifier: AERL_v1_0

Program summary URL: http://cpc.cs.qub.ac.uk/summaries/AERL_v1_0.html

Program obtainable from: CPC Program Library, Queen's University, Belfast, N. Ireland

Licensing provisions: Standard CPC licence, <http://cpc.cs.qub.ac.uk/licence/licence.html>

No. of lines in distributed program, including test data, etc.: 37763

No. of bytes in distributed program, including test data, etc.: 2757956

Distribution format: tar.gz

Programming language: MATLAB.

Computer: Any capable of running MATLAB version R2010a or higher.

Operating system: Any platform supporting MATLAB version R2010a or higher.

RAM: Minimum required is 1 GB. Memory usage increases for less intense magnetic fields.

Classification: 15.

[☆] This paper and its associated computer program are available via the Computer Physics Communication homepage on ScienceDirect (<http://www.sciencedirect.com/science/journal/00104655>).

* Corresponding author at: Scientific Computing Laboratory, Institute of Physics Belgrade, University of Belgrade, Pregrevica 118, 11080 Belgrade, Serbia. Tel.: +381 113713152.

E-mail address: jelena.smiljanic@ipb.ac.rs (J. Smiljanić).

Nature of problem:

The nature of the problem is to provide an efficient numerical algorithm implementation for optimization of GaAs/AlGaAs QCL active region parameters and calculation of output properties in the magnetic field.

Solution method:

The optimization of the QCL laser performance at selected wavelength is performed at entire free-parameters space using simulated annealing algorithm. The scattering rates are calculated in the presence and without magnetic field and used as coefficients in rate equations. The standard MATLAB procedures were used to solve iteratively this system of equations and obtain distribution of electron densities over electronic states.

Restrictions:

The machine must provide the necessary main memory which decreases roughly quadratically with the increase of the magnetic field intensity.

Running time:

Optimization time on Intel 3 GHz processor is about 2×10^4 s. The calculation time of laser output properties for values set automatically in GUI is 5×10^4 s.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

Molecular beam epitaxy (MBE) technique [1], allowed the realization of complex heterostructures, with the quantum well as a basic structural building block. The inherent design variability of these structures gave rise to a new direction of research that relies on the so-called band-structure engineering. In superlattices comprising many coupled quantum wells, energy subbands are formed in analogy with what happens when atoms are brought together to form a solid [2]. The first intersubband heterostructure laser was created by Jerome Faist et al. [3]. In recent years the scientific community has witnessed a rapid progress in the development of semiconductor quantum cascade lasers (QCLs) [4–8]. The optimization of the structure and prediction of performance of the QCL is an important contribution of theory to the development [9–14].

Quantum cascade lasers have become light sources of choice for a broad variety of applications including high-sensitivity gas sensing, infrared imaging, military countermeasures, security monitoring, non-invasive medical diagnostics, optical communications [3,4,15–19]. In the mid- and far-infrared spectral range, these powerful light sources are particularly appreciated for the wide scope of operating wavelengths which can be achieved by using the same heterostructure material combination. Some of the features that make this kind of device so unique are: unipolarity, the population inversion achieved through careful quantum engineering of the lifetimes of the states (typically in the ps range) and the electron recycling due to the cascading scheme. Their performance under the influence of high magnetic field has been intensively studied as well, due to the fact that magnetic field may serve as an efficient tool to determine the nature and magnitude of the scattering mechanisms in the active region [5–8,20–24]. Early experiments have demonstrated QCL emission enhancement and a magneto-phonon-resonance effects [5–7] which have led to a deeper insight into the physics of intersubband lasers. Furthermore, in the THz frequency range where the performances of QCLs deteriorate rapidly with temperature, due to detrimental thermal activation of non-radiative losses, the highest operating temperature of 225 K is reported for the structure assisted by external magnetic field for additional carrier confinement [21]. Recently, a portable QCL based infrared magnetospectrometer covering the spectral range from 5 to 120 μm has been presented, for applications in cyclotron-resonance spectroscopy measurements where high magnetic fields (up to 60 T) are required [16]. Hence, a detailed understanding of various scattering mechanisms under the influence of strong magnetic field is an important factor for improving QCL performances and applicability.

The typical design of the QCL active region entails a three-level system. The carriers are injected into the upper laser state, from which they can relax into the lower state by means of photon-assisted transition or by scattering, mainly by interactions with LO-phonons. In order to achieve the population inversion the lifetimes of electrons in the quantum states must satisfy the relation $\tau_{32} > \tau_{21}$: this condition is achieved by means of two key points. The lifetime τ_{21} is reduced by making the energy E_{21} resonant with the optical phonon energy, which is the most efficient scattering mechanism. The lifetime τ_{32} is increased by employing a transition with a reduced spatial overlap of the wavefunctions. Further extension of otherwise short carrier lifetime (of the order of 1 ps) is achieved using an intense magnetic field parallel to the growth direction of semiconductor layers [6–8,20]. The magnetic field breaks the two-dimensional (2D) in-plane continuous energy subbands into discrete Landau levels. The lifetime of electrons in the excited laser state is strongly modulated by the applied magnetic field. The dependence of electron lifetime on magnetic field intensity results in oscillations in the laser emission intensity. The inelastic scattering by LO-phonons and elastic scattering by interface roughness were identified as main mechanisms behind this effect by Leuliet et al. [8]. Given that the scattering processes between the two states depend on their energy spacing, certain relaxation mechanisms can be enhanced or inhibited by varying the magnetic field strength, although they may be influenced by the operating temperature as well. LO-phonon scattering is well explained in previous theoretical and experimental works [6–8,20,25]. Increasing magnetic field reduces the number of relevant Landau levels and changes the energy differences between individual levels, thus affecting the lifetime of carriers in higher states. On the other hand, the influence of interface roughness scattering has only recently been studied in detail [26]. The strength of the interface roughness scattering in a particular sample is determined by the morphology of the interfaces. In contrast to LO-phonon, interface roughness (IR) scattering does not depend on the temperature. As a result, efficiency of interface roughness scattering mechanism remains constant with increasing temperature, while the efficiency of LO-phonon scattering is reduced due to their higher absorption [27].

The purpose of the present paper is to give a simple and efficient program for the numerical solution of the rate equations based model and analysis of the optical gain in the active region of the QCL in magnetic field. An easy-to-use MATLAB® user interface is provided. In order to find an optimal design of the structure for operation at specific wavelength and temperature, optimization procedure is provided as well. The procedures for calculation of

LO-phonon and IR scattering rates are also included both with and without magnetic field. The electron distribution over the states of the system is found by solving the full set of rate equations which describe the transitions between levels, and subsequently used to determine the optical gain. These results can be easily calculated by a moderate PC configuration using the MATLAB[®] programs provided, due to efficient numerical implementation. This is important, since the number of levels involved is inversely proportional to the applied magnetic field.

In Section 2, we present a theoretical description of QCL active region and introduce models for LO-phonon and IR scattering rate with and without the presence of an external magnetic field. The rate equations which describe the population change of each Landau level are presented as well. The stationary solution of these equations allows for evaluating the population inversion and the resulting optical gain. In Section 3 we present a description of the MATLAB[®] program, an explanation about inputs and how to use the graphical user interface. In Section 4, we demonstrate the program features by calculating the optimal QCL active region parameters at different wavelengths. The calculations are also presented for the scattering rates and the total relaxation rate from the upper laser state, for a wide range of magnetic fields (3–60 T) and two temperatures $T = 77$ and 300 K. By using the calculated scattering rates as input data, rate equations are solved and population inversion and the optical gain are obtained. For both the population inversion and the gain, interface roughness scattering is shown to have a significant influence in terms of reducing the predicted magnitude, especially at low temperatures. Finally, in Section 5 we present a brief summary of our work.

2. Theoretical considerations

The active region of the QCL structure under consideration comprises three coupled quantum wells (QWs) biased by an external electric field \vec{E} . In the absence of the magnetic field this system has three energy states, i.e., subbands ($n = 1, 2, 3$), and the laser transition occurs between subbands $n = 3$ and $n = 2$. This active region is surrounded by suitable emitter/collector regions in the form of superlattices, designed as Bragg reflectors, which inject electrons into state $n = 3$ on one side, and allow for rapid extraction of carriers from the lowest subband $n = 1$, on the other side. The energy difference between E_2 and E_1 should match the LO-phonon energy in order to ensure fast depopulation via LO-phonon scattering and maintain a short lifetime for the lower laser level. In addition, we introduce in our calculations the interface roughness scattering as an additional non-radiative relaxation mechanism.

In the absence of an external magnetic field, the electronic subbands have a free particle-like energy dispersion in the direction parallel to the QW planes $E_n + \hbar^2 k_{\parallel}^2 / 2m_{\parallel n}(E_n)$, where $m_{\parallel n}(E_n)$ is the energy-dependent in-plane wave vector. However, when this structure is subjected to a strong magnetic field B in the z -direction, continuous subbands transform into series of individual (strictly discrete) states, the total energies of which are [7] $E_{n,l} \approx E_n + (l + 1/2)\hbar\omega_{c_n}$ where $l = 0, 1, 2, \dots$ is the Landau index, $E_n \equiv E_n(k_{\parallel} = 0)$, the term $(l + 1/2)\hbar\omega_{c_n}$ originates from the in-plane kinetic energy part of the subband, and $\omega_{c_n} = eB/m_{\parallel n}$ is the corresponding cyclotron frequency. The values of B which give rise to resonant LO-phonon emission are found by solving the equation $E_{3,0} - E_{n,l} = \hbar\omega_{LO}$ where $n = 1, 2$, while $\hbar\omega_{LO}$ is the LO-phonon energy.

According to [8,20], to account for the variations of the well widths, a Gaussian probability density is introduced:

$$\Pi(L_i) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(L_i - L_{i0})^2 / 2\sigma^2} \quad (1)$$

for the i th well width L_i , $i = 1, 2, 3$. In order to keep the results as analytical as possible, we assume that around a mean value L_{i0} the

energy difference varies linearly with L_i , i.e.

$$E_{n_i, l_i}(L_i) - E_{n_f, l_f}(L_i) \approx E_{n_i, l_i}(L_{i0}) - E_{n_f, l_f}(L_{i0}) - \gamma(L_i - L_{i0}) \quad (2)$$

with the factor γ taken the same for all Landau levels, according to [8,20].

By introducing Eqs. (1) and (2) into the Fermi golden rule, we obtain the following function:

$$\begin{aligned} J^s &= \int_{-\infty}^{\infty} \Pi(L_i) \delta[E_{n_i, l_i} - E_{n_f, l_f} - \Delta E_s] dL_i \\ &= \frac{1}{\delta\sqrt{2\pi}} e^{-(E_{n_i, l_i} - E_{n_f, l_f} - \Delta E_s)^2 / 2\delta^2}, \end{aligned} \quad (3)$$

where $\delta = \sigma\gamma$ is the width of the Gaussian distribution of energy difference $E_{n_i} - E_{n_f} + (l_i\omega_{c_{n_i}} - l_f\omega_{c_{n_f}})\hbar + \hbar(\omega_{c_{n_i}} - \omega_{c_{n_f}})/2 - \Delta E_s$, and s denotes the scattering mechanism (electron–LO-phonon scattering (LO) or interface roughness (IR) scattering). In our notation $\Delta E_{LO} = \hbar\omega_{LO}$ and $\Delta E_{IR} = 0$. The terms $E_{n,l}$ represent the total energies of Landau levels and a more detailed explanation of their calculation will be provided in the continuation of this section (see Eq. (19)).

The material parameters for GaAs used in the calculation are $m_{\text{GaAs}} = 0.067m_0$ and for AlAs $m_{\text{AlAs}} = 0.15m_0$ (m_0 is the free electron mass), effective mass in $\text{Al}_x\text{Ga}_{1-x}\text{As}$ is $m = xm_{\text{AlAs}} + (1-x)m_{\text{GaAs}}$. The dielectric permittivities used in the numerical calculations are $\epsilon_{\infty} = 10.67$, $\epsilon_s = 12.51$ and $\delta = 6$ meV.

2.1. Electron–LO-phonon scattering in magnetic field

The electron–LO-phonon scattering rates for phonon emission between the initial state E_{n_i, l_i} and the final state E_{n_f, l_f} may be found from

$$\begin{aligned} &\frac{1}{\tau_{(n_i, l_i) \rightarrow (n_f, l_f)}^{LO}} \\ &= \frac{2\pi}{\hbar} \sum_{\vec{q}} |\langle n_f, l_f, k_{x_f}, n_q + 1 | \hat{H}_{e-ph}(\vec{q}) | n_i, l_i, k_{x_i}, n_q \rangle|^2 J^{LO}. \end{aligned} \quad (4)$$

In this expression, electron–LO-phonon Hamiltonian $\hat{H}_{e-ph}(\vec{q})$ is the sum of the interaction Hamiltonians with each phonon mode defined by its 3D wave vector \vec{q} , see [20], and k_{x_i} and k_{x_f} are the initial and the final state wave vector components, respectively. From the previous equation one obtains the following analytical expression for the scattering rate:

$$\begin{aligned} \frac{1}{\tau_{(n_i, l_i) \rightarrow (n_f, l_f)}^{LO, e}} &= \frac{e^2\omega_{LO}}{4\epsilon_0} \left(\frac{1}{\epsilon_{\infty}} - \frac{1}{\epsilon_s} \right) \\ &\times \frac{1}{\delta\sqrt{2\pi}} e^{-(E_{n_i, l_i} - E_{n_f, l_f} - \hbar\omega_{LO})^2 / 2\delta^2} (n_q + 1) \\ &\times \int_0^{\infty} |F(q_{\parallel})|^2 G(q_{\parallel}) dq_{\parallel}, \end{aligned} \quad (5)$$

where ϵ_{∞} and ϵ_s are the static and the high-frequency relative dielectric constant, respectively, ϵ_0 is the vacuum dielectric permittivity and $n_q = [\exp(\hbar\omega_{LO}/k_B T) - 1]^{-1}$ is the mean number of LO-phonons. Furthermore, q_{\parallel} is the in-plane component of the phonon wave vector $\vec{q} = (q_z, q_{\parallel})$ and $F(q_{\parallel})$ is the lateral overlap integral

$$|F(q_{\parallel}, l_i, l_f)|^2 = e^{-(q_{\parallel}^2/2\beta^2)} \frac{l_i!}{l_f!} \left(\frac{q_{\parallel}^2}{2\beta} \right)^{l_f - l_i} \left[L_{l_i}^{l_f - l_i} \left(\frac{q_{\parallel}^2}{2\beta} \right) \right]^2, \quad (6)$$

where $\beta = \sqrt{eB/\hbar}$ is the magnetic length and $L_m^k(x)$ represents the associate Laguerre polynomial [28]. Finally, $G(q_{\parallel})$ stands for the

form factor given by

$$G(q_{\parallel}) = \int \int \eta_i^*(z) \eta_f(z) \eta_i(z') \eta_f^*(z') e^{-q_{\parallel}|z-z'|} dz dz', \quad (7)$$

where η_i and η_f denote the z -dependent parts of the electronic wave functions. The electron-LO-phonon scattering rate for phonon absorption [29] is

$$\frac{1}{\tau_{(n_f, l_f) \rightarrow (n_i, l_i)}^{LO, [a]}} = \frac{1}{\tau_{(n_i, l_i) \rightarrow (n_f, l_f)}^{LO, [e]}} \frac{1}{e^{\hbar\omega_{LO}/kT}}. \quad (8)$$

Phonon absorption is significant at room temperature ($T = 300$ K) and it almost vanishes at low temperatures ($T = 77$ K).

2.2. Electron-LO-phonon scattering without the magnetic field

When the magnetic field is not present, the electron-LO-phonon scattering rate may be evaluated from

$$\frac{1}{\tau_{if}^{LO}(B=0)} = \frac{2\pi}{\hbar} \sum_{\vec{q}} \left| \langle n_f, \vec{k}_f, n_q \right. \\ \left. + 1 | \hat{H}_{e-ph}(\vec{q}) | n_i, \vec{k}_i, n_q \rangle \right|^2 J^{LO}. \quad (9)$$

The Hamiltonian \hat{H}_{e-ph} is in this case the sum of interaction Hamiltonians with each phonon mode defined by its 3D wave vector \vec{q} , see [30], and \vec{k}_i and \vec{k}_f are the initial and the final state in-plane wave vectors, respectively. From the previous equation one obtains the following expression for the scattering rate in the absence of magnetic field:

$$\frac{1}{\tau_{if}^{LO}(B=0)} = \frac{e^2 \omega_{LO} m}{4\pi \hbar^2 \epsilon_0} \left(\frac{1}{\epsilon_{\infty}} - \frac{1}{\epsilon_s} \right) (n_q + 1) \int_0^{\pi} \frac{G(q_{\parallel})}{q_{\parallel}} d\theta, \quad (10)$$

where $G(q_{\parallel})$ and q_{\parallel} are the form factor and the in-plane component of phonon wave vector, respectively. The form factor is already defined in Eq. (7) and $q_{\parallel}^2 = 2m(E_i - E_f - \hbar\omega_{LO})/\hbar^2$.

2.3. IR scattering in magnetic field

We use the model for interface roughness scattering proposed by Leuliet et al., cf. [8]. This model assumes the in-plane terrace-like surface defects, as explained in [31]. In order to evaluate the interface roughness scattering rate, we introduce spatial distribution of roughness which follows the Gaussian correlation function [8,32–34,29]:

$$\langle \Delta(\vec{r}) \Delta(\vec{r}') \rangle = \Delta^2 e^{-|\vec{r}-\vec{r}'|^2/\Lambda^2}, \quad (11)$$

with Δ being the mean height of the roughness and Λ the correlation length. We also introduce the corresponding perturbation Hamiltonian [8],

$$\hat{H}_{IR} = U_0 \delta(z - z_i) \Delta(x, y), \quad (12)$$

where U_0 is the barrier height at interface position z_i .

The electron-interface roughness scattering rate can be calculated from the following expression:

$$\left\langle \frac{1}{\tau_{(n_i, l_i) \rightarrow (n_f, l_f)}^{IR}}(z_i) \right\rangle \\ = \frac{2\pi}{\hbar} \left\langle \sum_{k_{x_i}, k_{x_f}} \left| \langle n_f, l_f, k_{x_f} | \hat{H}_{IR} | n_i, l_i, k_{x_i} \rangle \right|^2 \right\rangle J^{IR}. \quad (13)$$

In the above expression, the averaging is performed over space (as follows from Eq. (11)), and over the initial state wave vector

component k_{x_i} :

$$\frac{1}{\tau_{n_i, l_i \rightarrow n_f, l_f}^{IR}} = \frac{\sqrt{2}}{\hbar \delta} e^{-\frac{(E_{n_i, l_i} - E_{n_f, l_f})^2}{2\delta^2}} \times |F_{if}|^2 \Delta^2 \Lambda \alpha \\ \times \int_{-\infty}^{+\infty} e^{-\frac{\Delta k_x^2}{2}} \left[\frac{2 + \Lambda^2 \beta^2}{2\beta^2} \right] \zeta(\Delta k_x) d(\Delta k_x). \quad (14)$$

Here, $\Delta k_x = k_{x_i} - k_{x_f}$, $F_{if} = U_0 \eta_i^*(z_i) \eta_f(z_i)$ and $\alpha = \beta^2 / (\pi l_f! l_i! 2^{l_f+l_i})$. The form factor $\zeta(\Delta k_x)$ is given by:

$$\zeta(\Delta k_x) = \int \int \exp \left(-\beta^2 [t^2 + (t - \omega)^2] - \frac{\omega^2}{\Lambda^2} \right) \\ \times H_{l_i} \left[\beta t - \frac{\Delta k_x}{2\beta} \right] H_{l_i} \left[\beta(t - \omega) - \frac{\Delta k_x}{2\beta} \right] \\ \times H_{l_f} \left[\beta t + \frac{\Delta k_x}{2\beta} \right] H_{l_f} \left[\beta(t - \omega) + \frac{\Delta k_x}{2\beta} \right] dt d\omega \quad (15)$$

and H_j is the Hermite polynomial of order j .

2.4. IR scattering without magnetic field

The electron-interface roughness scattering rate can be calculated from the following expression:

$$\left\langle \frac{1}{\tau_{n_i \rightarrow n_f}^{IR}}(z_i) \right\rangle = \frac{2\pi}{\hbar} \left\langle \sum_{\vec{k}_i, \vec{k}_f} \left| \langle n_f, \vec{k}_f | \hat{H}_{IR} | n_i, \vec{k}_i \rangle \right|^2 \right\rangle J^{IR}. \quad (16)$$

The Hamiltonian \hat{H}_{IR} is the same as in the case with magnetic field, i.e., Eq. (12). The averaging is performed over space, and over amplitude of the initial k_i and the final k_f state wavevector. The following expression for IR scattering rate without magnetic field is obtained:

$$\left\langle \frac{1}{\tau_{if}^{IR}(B=0)} \right\rangle = \frac{\sqrt{2\pi}}{\hbar \delta} e^{-\frac{(\Delta E_{if})^2}{2\delta^2}} \frac{1}{4} |F_{if}|^2 (\Delta \Lambda)^2 \\ \times \int_0^{+\infty} e^{-\frac{\hbar^4}{4m^2} x^2 - \left[\frac{\Lambda^2}{4} - \frac{\hbar^2}{m} \frac{\Delta E_{if}}{2\delta^2} \right] x} dx. \quad (17)$$

2.5. Rate equations and optical gain

Using previous expressions obtained for the scattering rates in the presence/absence of magnetic field, we can write the expression for the total scattering rate of the system as

$$\frac{1}{\tau} = \frac{1}{\tau^{LO}} + \sum_{z_i} \left\langle \frac{1}{\tau^{IR}}(z_i) \right\rangle. \quad (18)$$

Finally, if one wants to compare the effects of electron-LO-phonon scattering and electron-interface roughness scattering, two things can be noted: (1) due to the nature of electron-interface roughness interactions, scattering rates for transition from lower to upper and from upper to lower energy level are equal. (2) the LO-phonon scattering has maximum influence when the energy difference between two states is close to phonon energy, $\Delta E_{LO} = \hbar\omega_{LO}$; on the other hand, the effects of interface roughness are maximal when the energy difference approaches zero. Therefore, the two mechanisms of scatterings are complementary.

When the magnetic field B is applied, continuous subbands E_n transform into discrete Landau levels. The expression for energies

$E_{n,l}$ is given by Ekenberg as [36]

$$E_{n,l} = E_n + (l + 1/2) \frac{\hbar e B}{m_{\parallel n}} + [(8l^2 + 8l + 5)\langle\alpha_0\rangle + (l^2 + l + 1)\langle\beta_0\rangle] \frac{e^2 B^2}{2\hbar^2}. \quad (19)$$

In Eq. (19), $m_{\parallel n}$ represents the parallel effective mass of the n th subband in the absence of the magnetic field [36], while α_0 and β_0 are the nonparabolicity parameters. The optical gain corresponds to transitions $(3, l) \rightarrow (2, l)$ and is given by [25,35]

$$g_{3 \rightarrow 2} = \frac{2e^2 \pi^2 d_{3 \rightarrow 2}^2}{\bar{n} \epsilon_0 \lambda} \sum_l \delta(E_{3,l} - E_{2,l} - \hbar\omega) (N_{3,l} - N_{2,l}), \quad (20)$$

where \bar{n} is the material refractive index, λ and ω denote the wavelength and the frequency of the emitted light, respectively, while $N_{3,l} - N_{2,l}$ represents the degree of population inversion. The Dirac function in the above equation is replaced in numerical calculations by a Lorentzian with the linewidth parameter $\Gamma = 4.25$ meV [37], $\delta(E_{3,l} - E_{2,l} - \hbar\omega) \rightarrow \frac{\Gamma}{\pi [E_{3,l} - E_{2,l} - \hbar\omega]^2 + \Gamma^2}$. The energy difference between states is dependent on the Landau index l , $E_{3,l} - E_{2,l} \approx E_3 - E_2 + (l + 1/2)\hbar e B \left(\frac{1}{m_{\parallel 3}} - \frac{1}{m_{\parallel 2}} \right)$. Therefore, the Dirac function (i.e. the Lorentzian) cannot be put in front of summation in Eq. (20). The transition matrix element is calculated as $d_{3 \rightarrow 2} = \int \eta_3^*(z) z \eta_2(z) dz$, where the wavefunctions η_n are found by solving the Schrödinger equation in the form [36]

$$\frac{d^2}{dz^2} \alpha_0 \frac{d^2 \eta_n}{dz^2} - \frac{\hbar^2}{2} \frac{d}{dz} \frac{1}{m} \frac{d \eta_n}{dz} + U(z) \eta_n = E_n \eta_n. \quad (21)$$

Here, m represents the effective mass at the conduction-band minimum.

To calculate the optical gain we need to find the inverse population which is the solution of a nonlinear system of rate equations:

$$N_i \sum_{j \neq i} \frac{\bar{f}_j}{\tau_{i \rightarrow j}} - \bar{f}_i \sum_{j \neq i} \frac{N_j}{\tau_{j \rightarrow i}} + \frac{J_i}{e} = 0, \quad (22)$$

where indices $i, j = 1, 2, \dots$ denote the electronic states sorted by energy and

$$\bar{f}_i = 1 - \frac{\pi \hbar}{e B} N_i \quad (23)$$

is the probability that the state i is not occupied according to the Fermi–Dirac distribution. The electrons arrive in the active region by a constant current, and they are injected only into a limited number of Landau levels of the excited laser state, i.e. levels $(3, 0), \dots, (3, l_{3, \max})$. The injection current can be represented as a sum of all currents J_i which inject electrons into levels $(3, l_3)$, and in a similar manner, the extraction current can be expressed as a sum of all currents J_i which extract the electrons from levels $(1, l_1)$. The energy values of maximal Landau levels for each subband described by $l_{1, \max}$, $l_{2, \max}$ and $l_{3, \max}$, are taken in this work to be roughly $E_{3,0} + 5k_B T$ and it is reasonable to assume that the levels above are almost empty, cf. [25].

In the absence of an external magnetic field, the electronic subbands have free particle-like energy dispersion in the direction parallel to the QW planes, which in the parabolic approximation reads as $E_{\parallel} = \hbar^2 k_{\parallel}^2 / 2m^*$, where m^* is the effective mass and k_{\parallel} is the in-plane wave vector. The non-radiative lifetime for the state $|3, k_{\parallel}\rangle$ is limited by the electron–LO-phonon scattering into the two lower subbands of the active region, and the optical gain may be described by the following expression:

$$g_{3,2} = \frac{e^2 \omega}{2\bar{n} \epsilon_0 c} \int_0^{+\infty} F_{3,2} |d_{3,2}|^2 \delta(E_3 - E_2 - \hbar\omega) d(k_{\parallel}^2), \quad (24)$$

where ϵ_0 is the vacuum dielectric permittivity, c is the speed of light in vacuum, $\hbar\omega$ is the photon energy, $F_{3,2}$ stands for the difference of Fermi–Dirac functions for the initial and the final state, while $d_{i,f} = \langle \eta_i | z | \eta_f \rangle$ is the transition matrix element, and η_i and η_f denote the z -dependent parts of the wave functions.

2.6. Optimization procedure of the QCL active region

In order to optimize the laser performance at selected wavelength, one must consider the entire free-parameters space and this type of search is best performed by some established method for global optimization, such as simulated annealing algorithm [38] employed in this work. This algorithm belongs to the class of stochastic global optimization methods and uses the Metropolis function for the acceptance probability. The annealing algorithm in each step randomly generates new configurations and calculates a fitness function value. Any downhill step is automatically accepted while an uphill step may be accepted according to the Metropolis criteria. The algorithm starts initially with annealing control parameter set to a high value, and as the annealing proceeds, the value of annealing control parameter declines. In this way, the system is expected to wander initially towards a broad region of the active parameter space containing good solutions and then the search towards minimum is narrowed down. One of the most important phases in the implementation of any simulated annealing algorithm is the selection of a formal fitness function, which should be defined to encompass the goals of optimization. Here, the objective is to optimize the optical gain at selected wavelength, hence the fitness function is taken in the following form [11]

$$F = - \frac{g(B=0)}{\left[\left(\frac{E_3 - E_2}{\hbar\omega} - 1 \right)^2 + \Theta^2 \right] \left[\left(\frac{E_2 - E_1}{\hbar\omega_{LO}} - 1 \right)^2 + \Theta^2 \right]}, \quad (25)$$

where the term in the denominator favors achieving specified emission wavelength (i.e. photon energy $\hbar\omega$) and the LO-phonon resonance. In addition, Θ is a nonzero constant, which ensures that F is strongly driven towards resonance in the course of optimization, while remaining finite at the exact resonance, and $g(B=0)$ is the optical gain in the absence of the magnetic field, given by Eq. (24). In numerical calculation, the optical gain can be expressed via the gain coefficient $g^* = (1 - \tau_{21}/\tau_{32}) \tau_3 d_{3,2}^2$, where τ_{21} and τ_{32} are the scattering times and τ_3 is the upper laser level lifetime.

3. Program usage

In this section we provide instructions on how to operate the QCL program with a user-friendly graphical user interface (GUI) which was developed using MATLAB[®]. The QCL GUI enables the end user to easily manipulate the active region parameters, set and run simulations. The main m -file for running the program is *QCL.m*. The GUI has been designed so as to meet the following functions:

- The user can define her/his own input parameters, and to save them if desired. The user is able to choose only the input values within the allowed parameter range, see Table 1.
- The user is able to perform a search for the optimal QCL parameters for a given wavelength and temperature and to save the obtained values.
- The user is able to specify the range of magnetic field and visualize/save laser output properties in the presence of an external magnetic field.

Fig. 1 shows the graphical user interface. The window is basically divided into four panels. These panels are described in detail below.

Table 1
Summary of the input parameters and allowed range of values.

Parameter	Description	Minimum value	Maximum value
λ	Wavelength (μm)	5	15
T	Temperature (K)	1	350
Δ	IR scattering parameter (\AA)	0	2
Λ	IR scattering parameter (\AA)	0	200
X	Coefficient of $\text{Al}_x\text{Ga}_{1-x}\text{As}$ alloy composite	0.01	0.4
E	Electric field (kV/m)	3000	5000
$\text{well}(1)$	1st well width (\AA)	8^a	60
$\text{barrier}(1)$	1st barrier width (\AA)	8^a	60
$\text{well}(2)$	2nd well width (\AA)	8^a	60
$\text{barrier}(2)$	2nd barrier width (\AA)	8^a	60
$\text{well}(3)$	3rd well width (\AA)	8^a	60
B_{min}	Lower endpoint of range of magnetic field (T)	3	60
B_{max}	Upper endpoint of range of magnetic field (T)	3	60
dB	Step size for calculations in magnetic field (T)	/	/

^a If the system has two wells, one of the well or barrier widths is set to zero value, and other layer widths are in the range 8–60 \AA .

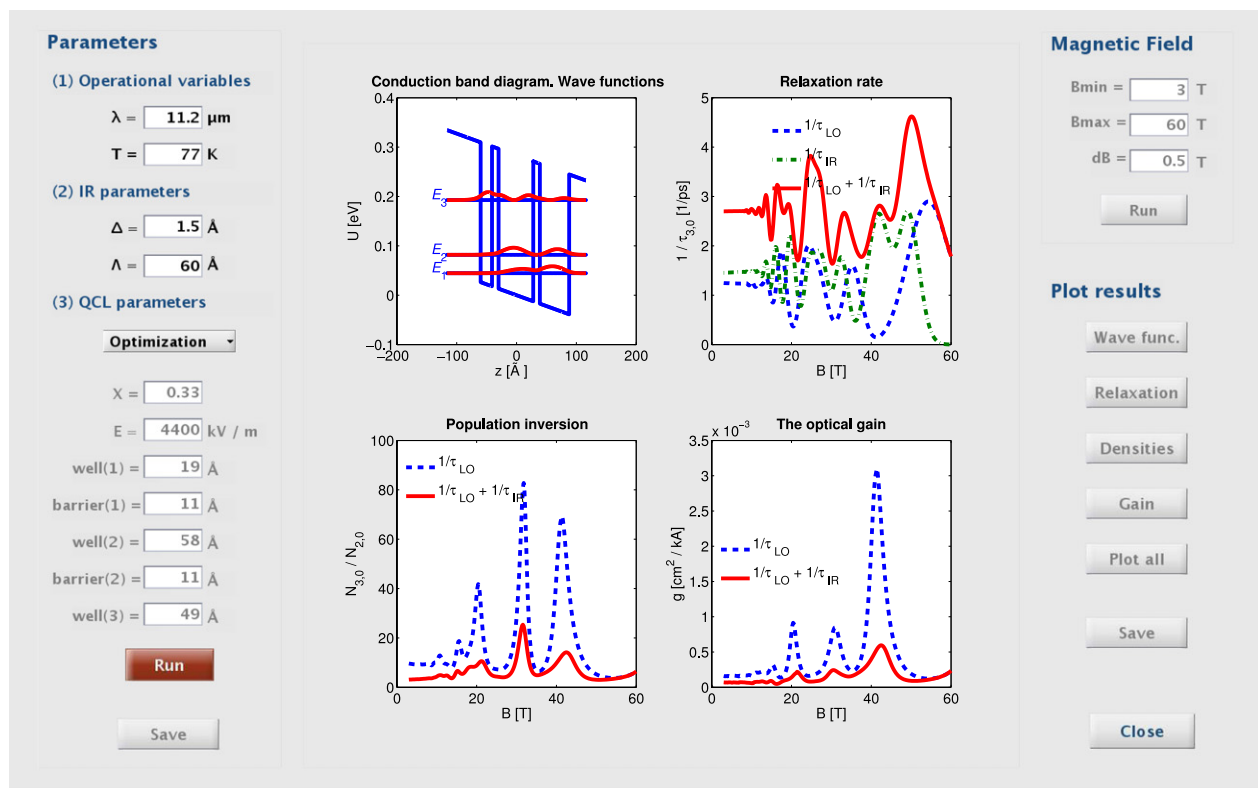


Fig. 1. QCL main window.

On the *Parameters* panel, there are three sub-panels. In the first two sub-panels, the input parameters are defined only by the user. In case of input entrance outside the allowed range (see Table 1), the parameter is automatically set to the closest allowed value or to NaN value (for example in case of text input). It is useful to note that if one of the IR parameters is chosen to be zero, the interface roughness scattering influence on the operation of QCL will not be considered. In the third *QCL parameters* sub-panel, the user has two options: (i) s/he can define the parameters manually, or (ii) s/he can optimize the laser performance at selected wavelength and temperature in the absence of magnetic field. If the optimization is selected, the ‘Run’ button will be enabled and after a click on this button alloy composition coefficient, electric field and layer widths lying in the appropriate range will be computed automatically using the simulated annealing algorithm. When the optimization is completed, calculated values will be entered in their boxes and the user will not be able to modify

these parameters. If s/he wants to modify or delete them, s/he has to select ‘Enter manually’ from the popup menu. The user can save the parameters. The ‘Save’ button is used for all input parameters of the *Parameters* panel. The input parameters are summarized in Table 1.

Once the QCL parameters are specified by using one of the above-mentioned ways, the user can calculate the scattering rates, the total relaxation rate, population inversion and the optical gain for a wide range of magnetic fields. On the *Magnetic field* panel, after defining the range of magnetic field, the user must click on the ‘Run’ button to analyze QCL dynamics when it is subjected to a magnetic field. After calculating QCL properties in magnetic field, the user can visualize results by clicking the appropriate button in the *Plot results* panel. Graphics appear in the central panel. The user can plot the active region of QCL, the total electron relaxation rate, the ratio of the total areal densities or the optical gain. There are options to plot a single graphic or to plot all the graphics together.

Table 2
Functions used for the optimization of QCL active region parameters.

Program	Description
QCLparam_opt.m	Finding optimal QCL parameters for given wavelength using annealing algorithm <i>calls</i> : fitness_fun.m, anneal.m
fitness_fun.m	Fitness function, function to be optimized (Eq. (25)) <i>inputs</i> : QCL parameters for optimization <i>calls</i> : QCLgain_opt.m
QCLgain_opt.m	Finding optical gain i.e. gain coefficient in the absence of the magnetic field (Eq. (24)) <i>inputs</i> : QCL parameters for optimization, wave functions, energy eigenvalues <i>calls</i> : WLO_opt.m, WIR_opt.m
WLO_opt.m	Finding LO-phonon scattering rate at zero magnetic field (Eq. (10)) <i>inputs</i> : wave functions, energy eigenvalues, temperature
WIR_opt.m	Finding IR scattering rate at zero magnetic field (Eq. (17)) <i>inputs</i> : energy eigenvalues, IR scattering parameters, width of the Gaussian distribution
anneal.m	Simulated annealing optimization method <i>inputs</i> : fitness function, initial values of parameters to be optimized

These graphics can be saved in the form of a picture file (.tif) by using the 'Save' button.

The 'Close' button on the right side of the GUI is a general button that closes the program and clears any temporary information.

4. Running the program and examples

In the following examples, applications of QCL program are presented. We illustrate how the program can be used to obtain (i) a gain-maximized structure and (ii) laser output properties in magnetic field.

4.1. Optimization of the QCL active region parameters

The first test run is concerned with the optimization of QCL active region parameters, described in the Section 2, by means of the simulated annealing algorithm. Here, we demonstrate the optimization of QCL active region parameters at two wavelengths $\lambda = 10 \mu\text{m}$ and $\lambda = 11 \mu\text{m}$.

An overview of the principal functions used for optimization is given in Table 2. The principal script is *QCLparam_opt.m*. The parameters m_{GaAs} and m_{AlAs} are defined in *param1.m*. The dielectric permittivities used in numerical calculations ϵ_{∞} , ϵ_s and the width δ are defined in *phyconst.m*. Nonparabolicity parameters α_1 and β_1 are defined in *param2.m*.

Starting from the predefined initial values of active region parameters, we minimize the fitness function given by Eq. (25) using the simulated annealing algorithm [38]. The optimization time on Intel 3 GHz processor is about 2×10^4 s. The QCL configurations with corresponding wave functions squared, obtained using the optimization algorithm, are given in Fig. 2(a) and (b). The structural parameters for the QCL active region for $\lambda = 10 \mu\text{m}$ are 25, 30, 59, 12 and 59 Å (for the well and the barrier widths, respectively, going from left to right) and the applied electric field in the z direction is $E = 3581$ kV/m. The obtained Al mole fraction is $X = 0.4$, so the structure may be realized by GaAs/Al_{0.4}Ga_{0.6}As. Another optimization is performed for $\lambda = 11 \mu\text{m}$, and the optimization procedure has delivered a structure with the following layer widths: 26, 23, 58, 12, 51 Å. The applied electric field is $E = 4600$ kV/m and Al mole fraction $X = 0.37$.

Table 3
Functions used for calculation of the output characteristics of QCL in the magnetic field.

Program	Description
main1.m	Finding subband positions at zero magnetic field and corresponding wave functions <i>calls</i> : phyconst.m, param1.m, shooting.m
main2.m	Finding scattering rates, the total relaxation rates, population inversion and the optical gain in the presence of an intense external magnetic field <i>calls</i> : phyconst.m, param1.m, param2.m, dipol.m, GQP.m, Ekenberg.m, WLO.m, WIR.m, iterrateequ.m, gainQCL.m
shooting.m	Shooting method for the Schrödinger equation <i>inputs</i> : energy eigenvalue
dipol.m	Calculation of dipole (transition matrix) element <i>inputs</i> : wave functions, energy eigenvalues, parallel masses
GQP.m	Calculation of form factor <i>inputs</i> : wave functions, in-plane component of the phonon wave vector
Ekenberg.m	Energies given by Ekenberg (Eq. (19)) <i>inputs</i> : energy eigenvalue, parallel mass, magnetic field
WLO.m	Finding LO-phonon scattering rate (Eq. (5)) <i>inputs</i> : energy eigenvalues, form factor, in-plane component of the phonon wave vector, magnetic field, temperature, width of the Gaussian distribution
WIR.m	Finding IR scattering rate (Eq. (14)) <i>inputs</i> : energy eigenvalues, magnetic field, IR scattering parameters, width of the Gaussian distribution
iterrateequ.m	Calculation of inverse population which is obtained from the solution of a nonlinear system of rate equations (Eq. (22)) <i>inputs</i> : scattering rate, distribution of current, magnetic field
gainQCL.m	Calculation of the optical gain (Eq. (20)) <i>inputs</i> : energy eigenvalues, inverse population, dipole element

4.2. Output characteristics of QCL in magnetic field

This section presents calculations of the output characteristics of QCL in the presence of an external magnetic field. The aim of this example is to demonstrate how the optical gain, resulting from the combined action of LO-phonon and IR scattering mechanisms, is influenced by varying the magnetic field strength. The overview of the scripts and functions is given in Table 3.

Each run has a set of unique parameters associated with it, including wavelength, temperature, IR scattering parameters, the alloy composition coefficient, electric field and layer widths. In the first calculation, we use parameters previously obtained by optimization at $\lambda = 10 \mu\text{m}$. In Fig. 3(a) and (b) the relaxation rates for the low temperatures ($T = 77$ K) and the room temperature ($T = 300$ K) are shown, respectively. The first bottleneck of simulation is calculation of the form factor, cf. Eq. (7). This takes roughly 3×10^3 s at 3 GHz processor. After that, relaxation rates are calculated for every energy level within subbands. Since the number of levels is inversely proportional to the applied magnetic field and three subbands are involved, the calculation time decreases as B^{-2} with magnetic field. Calculation time at $B = 10$ T on Intel 3 GHz processor is about 600 s. Oscillations of the relaxation rate with B are very pronounced, and very prominent peaks are found at values of the magnetic field which satisfy the resonance conditions for LO-phonon emission. If the relaxation rates due to interface roughness and LO-phonon scattering are compared, one can see that the local relaxation rate maxima are of the same order of magnitude and not correlated with respect to the applied magnetic field. This is due to the fact that interface roughness scattering has the largest influence when the energy difference between states is diminishing. In contrast, for LO-phonon scattering, when the arrangement

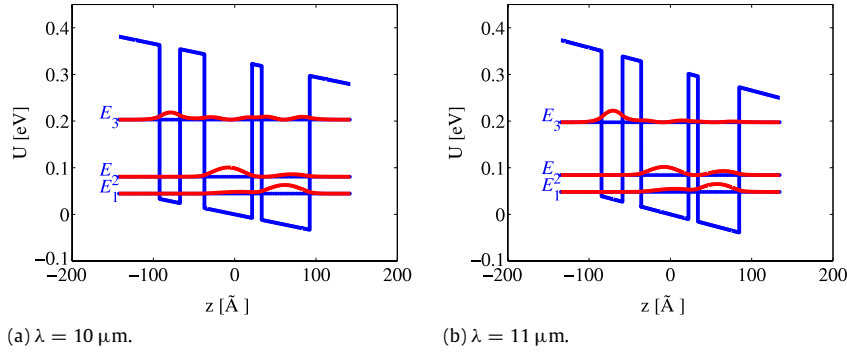


Fig. 2. (a) The active region of QCL, optimized for emission at $\lambda = 10 \mu\text{m}$. (b) The active region of QCL, optimized for emission at $\lambda = 11 \mu\text{m}$.

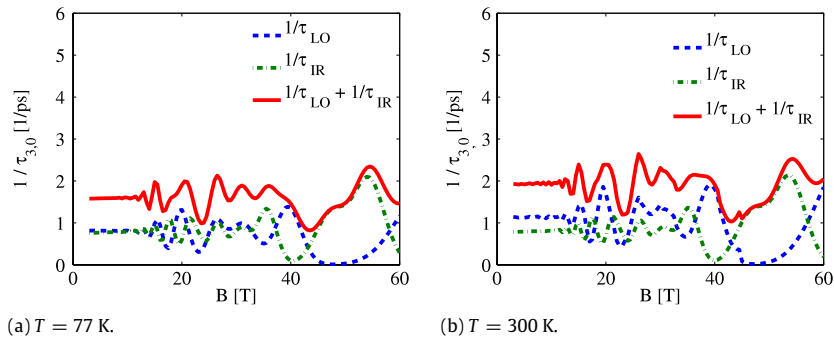


Fig. 3. The total electron relaxation rate due to the electron LO-phonon scattering and interface roughness scattering for transitions from the ground laser level of the third subband into the two sets of Landau levels of the lower subbands, for the magnetic fields in the range of $B = 3\text{--}60 \text{ T}$ and at temperature (a) $T = 77 \text{ K}$ and (b) $T = 300 \text{ K}$. The QCL active region parameters used are obtained by optimization at $\lambda = 10 \mu\text{m}$. The IR scattering parameters used in calculation are $\Delta = 60 \text{ \AA}$ and $\Lambda = 1.5 \text{ \AA}$.

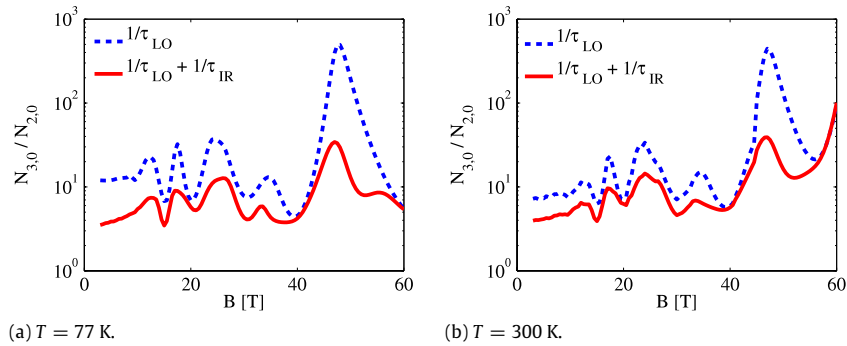


Fig. 4. The ratio of the total electron areal densities due to the electron LO-phonon and interface roughness scattering, in the ground laser levels of the third and the second subband, for the magnetic fields in the range of $B = 3\text{--}60 \text{ T}$ and at temperature (a) $T = 77 \text{ K}$ and (b) $T = 300 \text{ K}$. The QCL active region parameters used are obtained by optimization at $\lambda = 10 \mu\text{m}$. The IR scattering parameters used in calculation are $\Delta = 60 \text{ \AA}$ and $\Lambda = 1.5 \text{ \AA}$.

of laser levels is such that there is a level situated at $\hbar\omega_{LO}$ below the state $(3, 0)$, this type of scattering is enhanced. One can also see that the peaks at magnetic fields $B < 20 \text{ T}$ are a result of combined action of two scattering mechanisms. Assuming a constant current injection, the modulation of lifetimes of all the states in the system results in either suppression or an enhancement of population inversion between states $(3, 0)$ and $(2, 0)$, see Fig. 4(a) and (b).

In Fig. 5, the optical gain is shown at three wavelengths $\lambda = 9, 10, \text{ and } 11 \mu\text{m}$. The QCL active region parameters are obtained by optimization at $\lambda = 10 \mu\text{m}$. One can observe significant changes in the magnitude of the optical gain, but at the same time, the gain as a function of magnetic field looks qualitatively similar for different (close by) wavelengths. The difference in magnitude of the gain is about one order of magnitude between resonant wavelength $\lambda = 10 \mu\text{m}$ and other two wavelengths $\lambda = 9, 11 \mu\text{m}$. Finally, we should note that the QCL operating in the mid-IR spectral range was chosen to demonstrate our simulation code since experimental data were readily available [6].

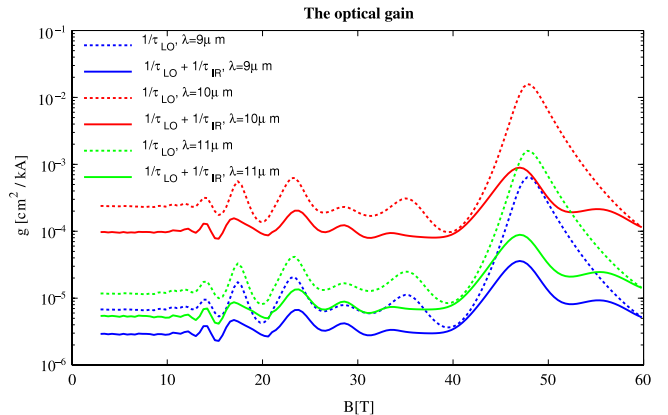


Fig. 5. The optical gain (per unit injection current) as a function of the applied magnetic field in the range $B = 3\text{--}60 \text{ T}$, at temperature $T = 77 \text{ K}$.

5. Conclusion

MATLAB[®]-based program is introduced, which solves the rate equations-based model and analyzes the optical gain in the active region of a QCL in magnetic field perpendicular to the layers of the structure. The scattering on both LO-phonons and interface roughness is included in modeling. Magnetic field alters the number of relevant in-plane electronic levels and the corresponding relaxation rates between them, by positioning some states on or off resonance with the upper laser level. Scattering effects are also important for relaxation of electrons from the second level to the first (ground) level. While interface roughness relaxation has maximal effects when the energy difference between levels is negligible, the LO-phonon scattering is enhanced if energy difference is close to resonant phonon energy. It is evident from the obtained numerical results that the inclusion of interface roughness scattering does not introduce additional (strong) peaks of inverse population and optical gain with varying magnetic field, it mainly affects their magnitude. The work aims to distribute an open source program which can be easily understood, used for efficient modeling and understanding of QCL operation and/or straightforwardly modified for the THz spectral range.

Acknowledgments

The authors acknowledge support by the Ministry of Education and Science of the Republic of Serbia under Projects No. ON171017, III45018, III45010 and NATO SfP Grant, ref. no. ISEG.EAP.SFPP 984068. Numerical simulations were developed and tested on the AEGIS e-Infrastructure, supported in part by FP7 projects EGI-InSPIRE, PRACE-1IP, PRACE-2IP, and HPSEE.

References

- [1] A.Y. Cho (Ed.), *Molecular Beam Epitaxy*, American Institute of Physics AIP Press, New York, 1994.
- [2] L. Esaki, R. Tsu, *IBM J. Res. Dev.* 14 (1970) 61.
- [3] J. Faist, F. Capasso, D.L. Sivco, C. Sirtori, A.L. Hutchinson, A.Y. Cho, *Science* 264 (1994) 553.
- [4] C. Gmachl, F. Capasso, D.L. Sivco, A.Y. Cho, *Rep. Progr. Phys.* 64 (2001) 1533.
- [5] C. Becker, C. Sirtori, O. Drachenko, V. Rylkov, D. Smirnov, J. Leotin, *Appl. Phys. Lett.* 81 (2002) 2941.
- [6] D. Smirnov, O. Drachenko, J. Leotin, H. Page, C. Becker, C. Sirtori, V. Apalkov, T. Chakraborty, *Phys. Rev. B* 66 (2002) 125317.
- [7] D. Smirnov, C. Becker, O. Drachenko, V.V. Rylkov, H. Page, J. Leotin, C. Sirtori, *Phys. Rev. B* 66 (2002) 121305.
- [8] A. Leuliet, A. Vasanelli, A. Wade, G. Fedorov, D. Smirnov, G. Bastard, C. Sirtori, *Phys. Rev. B* 73 (2006) 085311.
- [9] A. Bismuto, R. Terazzi, B. Hinkov, M. Beck, J. Faist, *Appl. Phys. Lett.* 101 (2012) 021103.
- [10] M.T. Arafin, N. Islam, S. Roy, S. Islam, *Opt. Quantum Electron.* 44 (2012) 701.
- [11] A. Daničić, J. Radovanović, V. Milanović, D. Indjin, Z. Ikonić, *J. Phys. D: Appl. Phys.* 43 (2010) 045101.
- [12] A. Mirčetić, D. Indjin, Z. Ikonić, P. Harrison, V. Milanović, R.W. Kelsall, *J. Appl. Phys.* 97 (2005) 084506.
- [13] D. Indjin, S. Tomić, Z. Ikonić, P. Harrison, R.W. Kelsall, V. Milanović, S. Kočinac, *Appl. Phys. Lett.* 81 (2002) 2163.
- [14] S. Tomić, M. Tadić, V. Milanović, Z. Ikonić, *J. Appl. Phys.* 87 (2000) 7965.
- [15] B.S. Williams, *Nat. Photonics* 1 (2007) 517.
- [16] O. Drachenko, S. Winnerl, H. Schneider, M. Helm, J. Wosnitza, J. Leotin, *Rev. Sci. Instrum.* 82 (2011) 033108.
- [17] A. Hugi, R. Maulini, J. Faist, *Semicond. Sci. Technol.* 25 (2010) 083001.
- [18] R.F. Curl, F. Capasso, C. Gmachl, A.A. Kosterev, B. McManus, R. Lewicki, M. Pusharsky, G. Wysocki, F.K. Tittel, *Chem. Phys. Lett.* 487 (2010) 1.
- [19] M. Tonouchi, *Nat. Photonics* 1 (2007) 97.
- [20] C. Becker, A. Vasanelli, C. Sirtori, G. Bastard, *Phys. Rev. B* 69 (2004) 115328.
- [21] A. Wade, G. Fedorov, D. Smirnov, S. Kumar, B.S. Williams, Q. Hu, J.L. Reno, *Nat. Photonics* 3 (2009) 41.
- [22] G. Scalari, C. Walther, L. Sirigu, M.L. Sadowski, H. Beere, D. Ritchie, *Phys. Rev. B* 76 (2007) 115305.
- [23] F.R. Jasnot, et al., *Appl. Phys. Lett.* 100 (2012) 102103.
- [24] G. Scalari, et al., *Appl. Phys. Lett.* 97 (2010) 081110.
- [25] J. Radovanović, V. Milanović, Z. Ikonić, D. Indjin, P. Harrison, *J. Appl. Phys.* 97 (2005) 103109.
- [26] M. Žeželj, V. Milanović, J. Radovanović, I. Stanković, *J. Phys. D: Appl. Phys.* 44 (2011) 325105.
- [27] H.B. Teng, J.P. Sun, G.I. Haddad, M.A. Stroschio, S. Yu, K.W. Kim, *J. Appl. Phys.* 84 (2008) 2155.
- [28] A.P. Prudnikov, Yu.A. Brychkov, O.I. Marichev, *Integrals and Series: More Special Functions*, Vol. 3, Gordon and Breach, New York, 1990.
- [29] A. Valavanis, Z. Ikonić, R.W. Kelsall, *Phys. Rev. B* 77 (2008) 075312.
- [30] T. Unuma, T. Takahashi, N. Takeshi, M. Yoshita, T. Noda, H. Sakaki, M. Baba, H. Akiyama, *Appl. Phys. Lett.* 78 (2001) 3448.
- [31] C. Weisbuch, R. Dingle, A.C. Gossard, W. Wiegmann, *Solid State Commun.* 38 (1981) 709.
- [32] T. Ando, A.B. Fowler, F. Stern, *Rev. Modern Phys.* 54 (1982) 437.
- [33] T. Unuma, M. Yoshita, T. Noda, H. Sakaki, H. Akiyama, *J. Appl. Phys.* 93 (2003) 1586.
- [34] N. Califano, et al., *Phys. Rev. B* 75 (2007) 045338.
- [35] S. Živanović, V. Milanović, Z. Ikonić, *Phys. Rev. B* 52 (1995) 8305.
- [36] U. Ekenberg, *Phys. Rev. B* 40 (1989) 7714.
- [37] J. Wang, J.-P. Leburton, Z. Moussa, F.H. Julien, A. Sa'ar, *J. Appl. Phys.* 80 (1996) 1970.
- [38] S. Kirkpatrick, C.D. Gelatt, M.P. Vecchi, *Science* 220 (1983) 4598.



Efficient routing on small complex networks without buffers



J. Smiljanić*, I. Stanković

Scientific Computing Laboratory, Institute of Physics Belgrade, University of Belgrade, Pregrevica 118, 11080 Belgrade, Serbia

ARTICLE INFO

Article history:

Received 10 August 2012

Received in revised form 29 November 2012

Available online 17 January 2013

Keywords:

Scale-free networks

Routing

Network capacity

ABSTRACT

In this paper, we are exploring strategies for the reduction of the congestion in the complex networks. The nodes without buffers are considered, so, if the congestion occurs, the information packets will be dropped. The focus is on the efficient routing. The routing strategies are compared using two generic models, i.e., Barabási–Albert scale-free network and scale-free network on lattice, and the academic router networks of the Netherlands and France. We propose a dynamic deflection routing algorithm which automatically extends path of the packet before it arrives at congested node. The simulation results indicate that the dynamic routing strategy can further reduce the number of dropped packets in a combination with the efficient path routing proposed by Yan et al. [5].

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

Complex networks are important for the functioning of the modern society. To ensure a free, uncongested traffic flow on the complex networks is of great interest. Intuitively, the traffic congestion could be largely reduced or completely avoided with a very large average degree of connectivity and/or node capacity for information packet delivery. The capacity of nodes to deliver information cannot be infinite. Also, upgrading the infrastructure is often not economically feasible [1,2]. The performance of the communication systems can be improved by implementing the more appropriate routing protocols without changing the underlying network structure [3–17], which is more realizable in the practice. Such work presents two problems. The first is finding out the optimal strategies for the traffic routing on a defined network structure. The second problem is finding a procedure to draw general conclusions about performance of routing strategies due to the variation of the real network topologies. The increasing speed of the network interfaces raises an important question concerning the size of buffers, complexity and cost. A considerable research effort is currently under way in an attempt to resolve compromise between buffer latency and complexity on one side and capacity on the other [18–25]. In the previous studies, the node buffer size in the traffic-flow model is set as infinite [3–16]. Our intention is to produce a relatively simple methodology for evaluating routing strategies in networks with limited buffering capability or without optical buffers. This should be important for the optical networks with either large volumes of the information traffic or without buffering capacity.

A number of network models are introduced in the past two decades [26,27]. A particular class of models is dedicated to networks embedded in the space [28]. Here we are interested in evaluating the models for representation of the spatially constrained networks, both in terms of distance between nodes and the extent of the network. We are interested in information flow optimization in small networks. Small networks should represent bulk of telecommunication networks or other dedicated information networks, e.g., regional optical backbones and academic networks. As examples of the real-world networks, we analyze the national research and educational networks (NRENs) of the Netherlands [29], France [30], Norway [31] and Spain [32] and compare them with Barabási–Albert scale-free network and scale-free network on lattice. The system size dependence of topological characteristics of the scale-free network on lattice is also analyzed.

* Corresponding author. Tel.: +381 113713000.

E-mail address: jelena.smiljanic@ipb.ac.rs (J. Smiljanić).

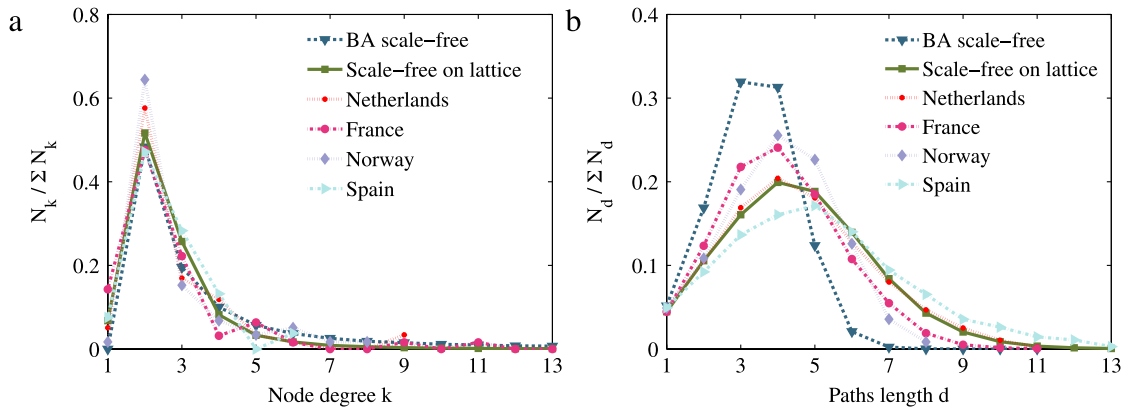


Fig. 1. Topological characteristics: (a) degree distribution; (b) length of path probability density function for NRENs of the Netherlands ($N = 59$), France ($N = 63$), Norway ($N = 58$) and Spain ($N = 53$) and generic network models: BA scale-free network ($N = 64, m = 2$) and scale-free network on lattice ($N = 64, \lambda = 2$).

In this work, three routing strategies are implemented and evaluated: the shortest path routing, efficient path routing and dynamic deflection routing. The shortest path routing is widely used routing strategy in praxis (by “shortest” we mean the path with the smallest number of links) [33]. However, in the shortest path routing strategy load distribution is not homogeneous. The majority of the shortest paths pass through the nodes that are highly connected, while other nodes carry much less traffic [34]. Yan et al. [5] presented an approach to redistribute traffic load in highly connected nodes to other nodes using link weight. An improvement is achieved through a targeted traffic redistribution from the most congested nodes. As result the congestion is reduced at the expense of a slight increase of the total path length and traffic. We compare this routing strategy with a dynamic routing strategy. The dynamic routing strategy improves the control of the congestion in the heavily loaded nodes by dynamically returning packet one step back. In this way, the dynamic strategy uses the redundant capacity of the links in network to temporarily store information, until congested node capacity is free. Further, we test a possibility of combining dynamic and static routing strategies. In order to check how different routing strategies behave in the larger networks, we have evaluated information loss dependence on packet generation rate and network size in case of the scale-free network on lattices.

The paper is organized as follows: in Section 2 we introduce the generic scale-free model and the scale-free model on lattice and compare their network characteristics with the national research and educational networks (NRENs) of the Netherlands, France, Norway and Spain. Furthermore, the characteristics of the scale-free model on lattice for different system sizes are considered. In Section 3, the information flow model and a measure of system performance are introduced. The static and dynamic routing strategies are described in Section 4 and their performance is analyzed in Section 5.

2. Network models

In this work, we compare topological network characteristics of Barabási–Albert scale-free model [26], the scale-free model on lattice [28] and the national research and educational networks (NRENs) of the Netherlands, France, Norway and Spain. Barabási and Albert observed an existence of a high degree of self-organization characterizing the large-scale properties of complex networks [26]. They have introduced a model of the scale-free networks with two key elements: probability that a new node connects to the existing nodes is not uniform and there is a higher probability that it will be linked to a node that already has a large number of connections. Thus, Barabási–Albert (BA) scale-free network model is formed in a series of steps in which new nodes are incorporated into the network. Algorithm is starting with a small number (N_0) of nodes, and at every time step new nodes with m connections are added. To incorporate preferential attachment, the model assumes that the probability of the new connection with the node i depends on its connectivity k_i and equals $P(k_i) = k_i / \sum k_j$. After a few algorithm steps, distribution of number of links per node takes scale-free form $P(k) \sim k^{-\lambda}$. In this work, for $N_0 = 3$ and $m = 2$, $n = 61$ algorithm steps are preformed. Obtained network consists of $N = 64$ nodes with $\lambda = 2$. The obtained network degree distribution corresponds well to the NRENs; cf. Fig. 1(a). The number of nodes in different NRENs is in the case of the Netherlands $N = 59$, France $N = 63$, Norway $N = 58$ and Spain $N = 53$.

However, the real-life networks are embedded into the geographical space and constrained by the cost of the links between the nodes. In the scale-free model on lattice (cf. [28]) the algorithm starts with a set of nodes that are identified with the set of lattice vertices in an $M \times M$ square. The lattice distance between two nodes is defined as the minimal number of “lattice steps” separating them in the regular lattice. In this model, network nodes are randomly assigned with the number of links (k) according to scale-free distribution $P(k) = Ak^{-\lambda}$, $m \leq k < K$ and connected to its closest neighbors. Therefore, exponent λ is a model parameter. We set $\lambda = 2$, as obtained from the connectivity distribution of NRENs; cf. Fig. 1(a). The choice of model parameter λ is also in accordance with the distribution of the number of links per node obtained with BA model. Normalization constant is $A \approx (\lambda - 1)m^{\lambda-1}$.

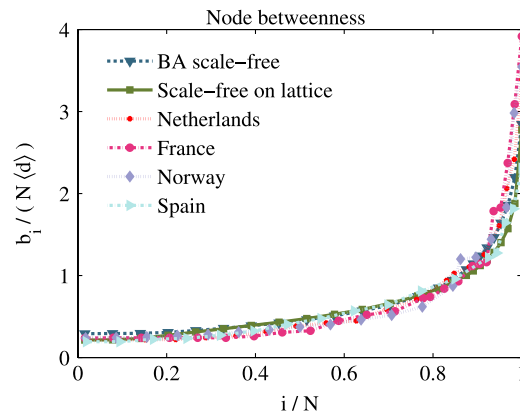


Fig. 2. Distribution of normalized node betweenness for NRENs of the Netherlands ($N = 59$, $\langle d \rangle = 4.54$), France ($N = 63$, $\langle d \rangle = 4.08$), Norway ($N = 58$, $\langle d \rangle = 4.06$) and Spain ($N = 53$, $\langle d \rangle = 5.08$) and generic network models: BA scale-free network ($N = 64$, $m = 2$, $\langle d \rangle = 3.36$) and scale-free network on lattice ($N = 64$, $\lambda = 2$, $\langle d \rangle = 4.59$).

At the first glance the most efficient mean to transfer information through the network is along the shortest paths. The distribution of the shortest path lengths is given in Fig. 1(b). The network diameter D , can be defined as the maximal length of the shortest path between any two nodes in the network, i.e., $D = \max\{d_{ij}\}$, where length of shortest path from the node i to the node j is d_{ij} . The small network diameter means that packets transmitted through the network, travel from one node to another quickly along the shortest path. As result the possibility of loss due to the congestion of the transmitting nodes is reduced. From Fig. 1(b), one can observe that the path length distribution of the BA scale-free network does not match NRENs well. This is not surprising, since in the BA scale-free model the Euclidean distance between nodes is irrelevant. For the version of scale-free network on lattice, the model has desirable properties in terms of path lengths. Network diameter in different NRENs is in the case of the Netherlands $D = 11$, France $D = 11$, Norway $D = 8$ and Spain $D = 13$. Scale-free network (BA) generation algorithm generates networks with considerably smaller network diameters $D = 6.84$. Average network diameter obtained for scale-free on lattice model is $D = 10.81$, and compares well with diameters of NRENs.

The quality of the communication of the two non-adjacent nodes, i.e., node j and node k , depends on the nodes belonging to the paths connecting the nodes j and k . Consequently, a measure of the relevance of a given node for overall network performance can be obtained by counting the number of geodesics going through it, and defining the so-called node betweenness. More precisely, the betweenness b_i of the node i , sometimes referred to also as load, is defined as [34,35]:

$$b_i = \sum_{j,k \in N, j \neq k} n_{jk}(i) \quad (1)$$

where $n_{jk}(i)$ is the number of the shortest paths connecting j and k and passing through the node i . Let $\langle d \rangle$ denote the average path length of the given network measured according to the shortest path routing rule. The normalized betweenness distribution for NRENs of the Netherlands, France, Norway and Spain and BA scale-free network and scale-free network on lattice is shown in Fig. 2. We observe that both models reproduce betweenness characteristics of real networks well.

In Fig. 3 we examine the system-size dependent ($N = 64, 256, 1024$ and 2500) behavior of the path length and node betweenness distribution for the scale-free on lattice model. The average path length initially increases fast with the system size (i.e., for $N < 1024$). In the larger systems the path length distribution changes little with the system size; cf. $N = 1024$ and 2500 in Fig. 3(a). The reason for this is existence of one or several nodes with high degree i.e. close to N ; cf. Ref. [36]. The average betweenness is roughly proportional to the system size N and the average path length $\langle d \rangle$ as $\langle b_i \rangle \sim N \langle d \rangle$. Also, the maximal value of node betweenness increases roughly with the square of system size N ; cf. Ref. [34].

3. Information flow model

In the information flow model all nodes are treated as both hosts and routers. Each node has a predefined maximum packet routing capacity C and communication channels have an infinite capacity to transmit the packets. If packets arrive to the node whose routing capacity has been already reached (i.e., congested node), they will be dropped. The dynamics of the model is as follows. At each time step t , an information packet is created at random node with the probability p . Therefore p is the control parameter: small values of p correspond to the uncongested (free) flow of packets and high values of p correspond to the high flow rate of packets. When a new packet is created, a destination node, different from the origin, is chosen randomly in the network. In this paper, we analyze the case that each node is able to send one packet at each time step. The travel time T of a packet is defined as the time spent by the packet between its source and destination. Here we do not take into account the time delay of the information transfer at each node or link, so that all data are delivered in a unit time, regardless of the distance between any two nodes. Thus, during the following time steps $t + 1, t + 2, \dots, t + T$, the

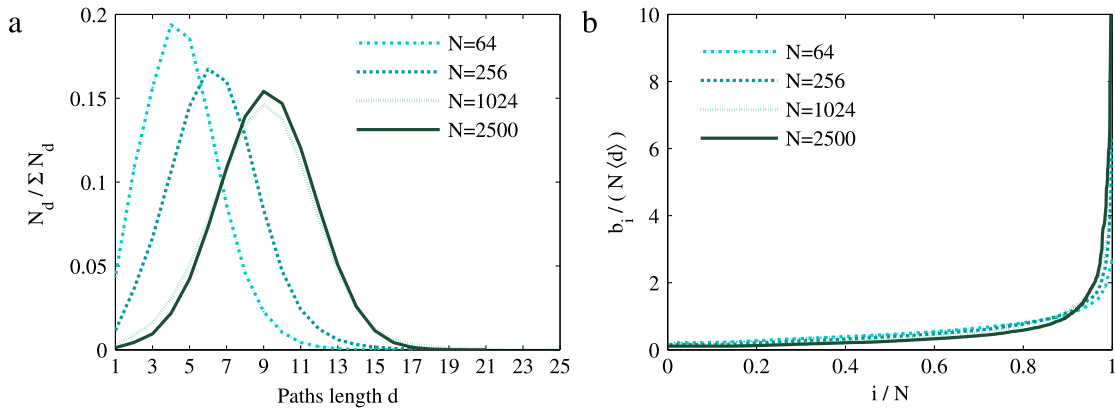


Fig. 3. (a) Length of path probability density function for scale-free network on lattice for various system sizes N . (b) Distribution of normalized node betweenness for scale-free network on lattice for various system sizes N , where $\langle d \rangle$ is the average path length ($N = 64, 256, 1024, 2500$ and $\langle d \rangle = 4.59, 6.38, 8.98, 9.14$ respectively).

packet travels towards its destination and the time T is related to the path length. Once the packet reaches the destination node, it is delivered and disappears from the network.

When the amount of packets is small, the network is able to deliver all packets that are generated. Conversely, when p is large enough the number of generated packets is larger than the number of packets that the network can manage to solve and the nodes enter in a state of congestion. The characteristic that measures the system performance is the packet drop probability η ,

$$\eta = \frac{R_d}{R} \tag{2}$$

defined as a ratio of the total number of deleted packets R_d and the total number of generated packets R . A high drop probability indicates that a large percentage of packets cannot reach their destinations. Then, the quality of service is poorer.

4. Routing strategies

4.1. Static routing

Packets can be delivered according to different routing strategies. When static weighted routing strategy is used, packets choose the routing path with the minimum sum weight of links. For any pair of source and destination node, there may be several paths with the same weight between them. We randomly choose one of these paths and put it into the fixed routing table which is followed by all packets. It has to be assumed that each node has the same capability of delivering packets, that is, at each time step all the nodes can deliver at most C packets one step towards their destinations according to the fixed routing table. Here we compare two static routing strategies: (i) the shortest path routing, i.e., the links in the network have the same weight, $w_{st} = 1$, where w_{st} is the weight of the link going from s to t . Routing communication along the shortest paths is of course beneficial for speed, but if there is a limit to the node load and network traffic is heavy, congestion is a threat to the nodes with the largest betweenness. Obviously, bypassing high-degree nodes, packet will have more chance to reach its destination. (ii) In the second, efficient path routing strategy, the weight of the link between the nodes s and t is defined as [5]:

$$w_{st} = \left(\frac{k_s + k_t}{\min_{i \neq j} (k_i + k_j)} \right)^\beta, \tag{3}$$

where k_i denotes the degree of node i and β is an adjustable parameter. The efficient path between nodes i and j is corresponding to the route that makes the sum weight of links minimum. As for any pair of source and destination, there may be several efficient paths between them. We randomly choose one of them and put it into the fixed routing table which is followed by all the information packets.

In Fig. 4(a), we show the betweenness deviation as a function of β on different scale-free networks. The optimal routing strategy for generic scale-free network and scale-free network on lattice is corresponding to $\beta_{opt} = 1 \pm 0.1$, where betweenness deviation is the smallest. This is also the optimal value for NREN of the Netherlands. The optimal value for French NREN is higher, i.e., $\beta_{opt} = 1.2 \pm 0.1$, due to larger betweenness deviation. Obviously $\beta = 0$ recovers the shortest path length. In comparison with the shortest path routing strategy average path length slightly increases with β ; cf. Fig. 4(b).

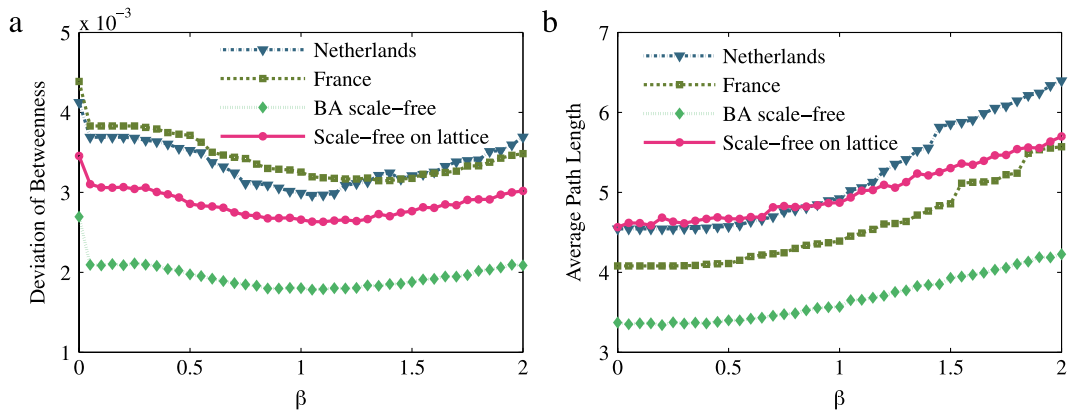


Fig. 4. (a) Deviation of node betweenness vs. parameter β from Eq. (3) and (b) average path length vs. parameter β for different network models.

4.2. Dynamic routing

In this paper, we also analyze the effectiveness of the dynamic deflection routing strategy. We assume each node has a knowledge about the load of its neighbors. If a package is about to arrive to a congested node, it will be deflected, i.e., its path will be dynamically extended. In other words, if we denote a path between nodes i and j as $P(i \rightarrow j) := x_0, x_1, \dots, x_{n-1}, x_n$, where $x_0 = i$ and $x_n = j$. The node x_m , instead forwarding packet to the congested node x_{m+1} , deflects the packet back to the node x_{m-1} . In the moment $t + 1$, the node x_{m-1} sends the packet to the node x_m , and in the moment $t + 2$ the packet is sent to the node x_{m+1} . The deflection of the packet is tried only once.

The total number of packets in the network is equal to $pN\langle d \rangle$. Since node betweenness represents the number of paths going through the node, the probability that a packet in one time step is going through node i is $B_i = \frac{b_i}{N(N-1)\langle d \rangle}$. Therefore, the average number of packets going through node i per step is $q_i = pN\langle d \rangle \times B_i = p \frac{b_i}{N-1}$. In order to avoid a reduction of the system performance due to the deflection traffic, we introduce a condition that a node can send packet back only if $q_i < 0.5C$. The localized congestion events are more likely to occur at the nodes with higher betweenness and that in the regime when traffic is not heavily congested. We find that it is sufficient to implement the deflection routing only in the nodes with highest betweenness (about 10% of all nodes). A higher amount of the nodes with deflection routing capability does not improve significantly the network performance.

5. Simulation results

To compare different routing strategies, we apply previously described routing algorithms and measure packet drop probability η in the two generic networks and NRENs of France and the Netherlands. As shown in Fig. 5, the scale-free network on lattice under the shortest path routing strategy reproduces well behavior of both NREN topologies, for both simulated node transfer capacities $C = 2$ and 4. We can observe that in the real-world networks and scale-free network on lattice for $p > 0.1$ and for node capacity $C = 2$ more than 40% of generated packets are dropped. Also, when amount of the traffic in the network is small ($p < 0.1$), there is a significant amount of loss, i.e., more than 5%. For the higher node transfer capacity ($C = 4$) the number of deleted packets significantly decreases and there is almost no packet loss when the network traffic is small. The BA scale-free network is less prone to the congestion compared to the scale-free network on lattice and two NRENs, which is expected since this network has a shorter average path length.

Impact of the static and dynamic routing strategies is analyzed in Figs. 6 and 7, respectively. The results confirm that the static weighted routing using the topological information can greatly improve the traffic flow for small p values ($p < 0.1$), in comparison with the shortest path routing mechanism. The static weighted routing method proves to be better for NREN topologies and scale-free model on lattice. The dynamic routing strategy is better in case of BA scale-free networks due to shorter paths; cf. Fig. 7. The shorter paths between nodes result in a lower amount of the packages being transferred through the network at any instance of time compared to other studied networks, allowing for the dynamic deflection algorithm to be more efficient.

The dynamic and static routing strategies are complementary and can be combined. The system performance for combined static and dynamic routing and the static routing is compared in Fig. 8. The redistribution of the traffic load from the most congested nodes enables more efficient dynamic deflection routing. Improvement is in the range of 5%–20%. In NREN of Netherlands four most congested nodes after path redistribution were in the proximity, and combined effect of dynamic–static routing strategy is rather weak; cf. Fig. 8(a). On the other hand, the French NREN shows a considerable improvement due to the larger distance between the most congested nodes; cf. Fig. 8(b). This is the consequence of the fact that both studied networks are small and the difference in the performance of the routing algorithms is a result of the variations in the topology. The scale-free model on lattice allows us to evaluate and compare routing strategy on large

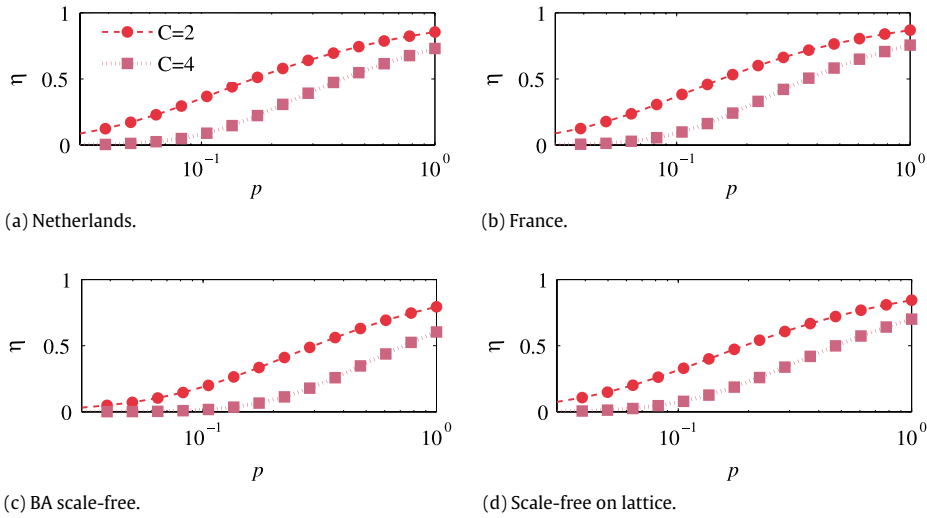


Fig. 5. Packet drop probability η vs. the packet-generating rate p for the shortest path routing algorithm. We have set node transfer capacity $C = 2$ (circles) and $C = 4$ (squares).

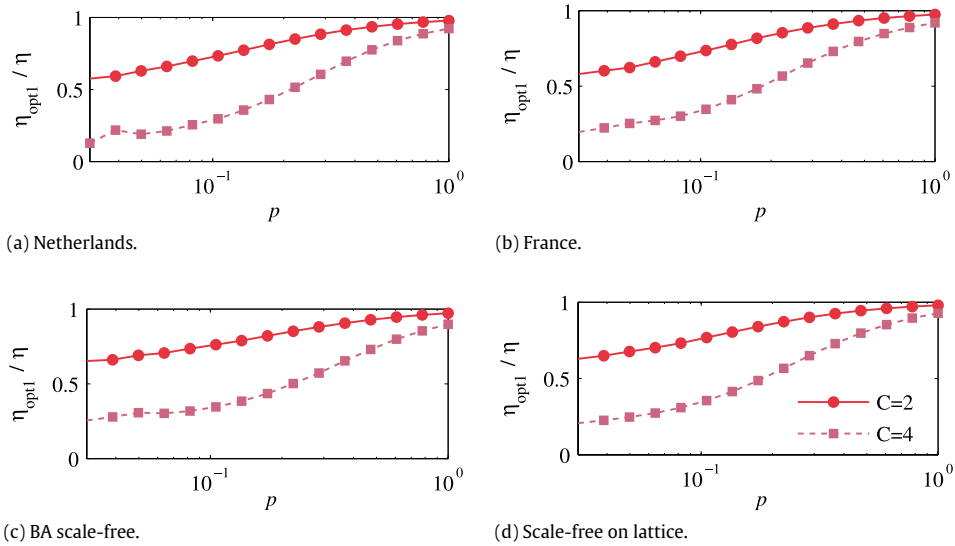


Fig. 6. Comparison between packet drop probability η as a function of packet-generating rate p for the shortest path routing and static weighted routing strategy with β_{opt} (η_{opt}). We have set node transfer capacity $C = 2$ (circles) and $C = 4$ (squares).

number of artificially generated networks. In this way we obtain information, how whole class of the networks responds to the routing strategy in average and we observe a considerable improvement in performance; cf. Fig. 8(d).

Until now, we have tested the routing strategies for small networks (with $N \leq 64$ nodes). At this point, we want to show how the static and dynamic strategies behave in the larger networks, i.e., $N = 256, 1024,$ and 2500 . In Fig. 9, we show the relationship between packet drop probability η and network size N under different routing strategies on scale-free networks on lattice. In the case of the shortest path routing, the results are consistent with our intuition: with the increase of the system size N the congestion starts at smaller packet generation rates p . In order to understand where the packet drop first occurs and how congestion develops, we need to look into the properties of the single network nodes. The probability of information drop (loss) on certain node depends on the number of packets going through it. The average number of packets going through node i per step is $q_i = p \frac{b_i}{N-1}$ and the largest values of node betweenness, b_{max} , increase faster than linear with system size; cf. Ref. [34] and Fig. 3(b). Therefore increase of system size N results in the increase of drop probability. Fig. 9(b) compares the relation of the packet drop probability η vs. packet-generating rate p under the static weighted routing and the shortest path routing with different network size N . The efficiency of static weighted routing strategy increases with the system size N . Also, the larger networks compared to very small ones, i.e., $N = 64$, offer more possibility to redistribute load and therefore the strategy is more efficient. On the other hand, dynamic routing algorithm changes alone is less efficient in larger networks; see Fig. 9(c). The reason for this is the existence of the nodes with the large

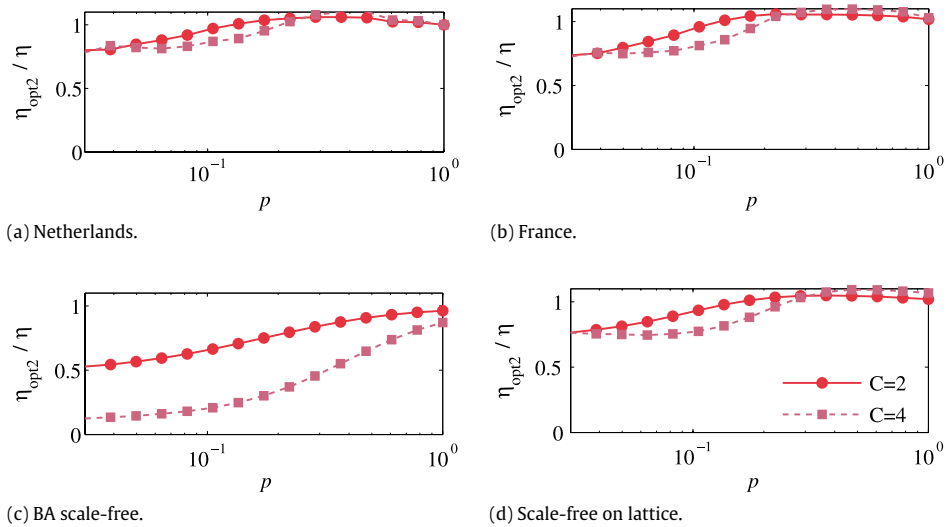


Fig. 7. Comparison between packet drop probability η as a function of packet-generating rate p for the shortest path routing and dynamic routing algorithm (η_{opt2}). We have set node transfer capacity $C = 2$ (circles) and $C = 4$ (squares).

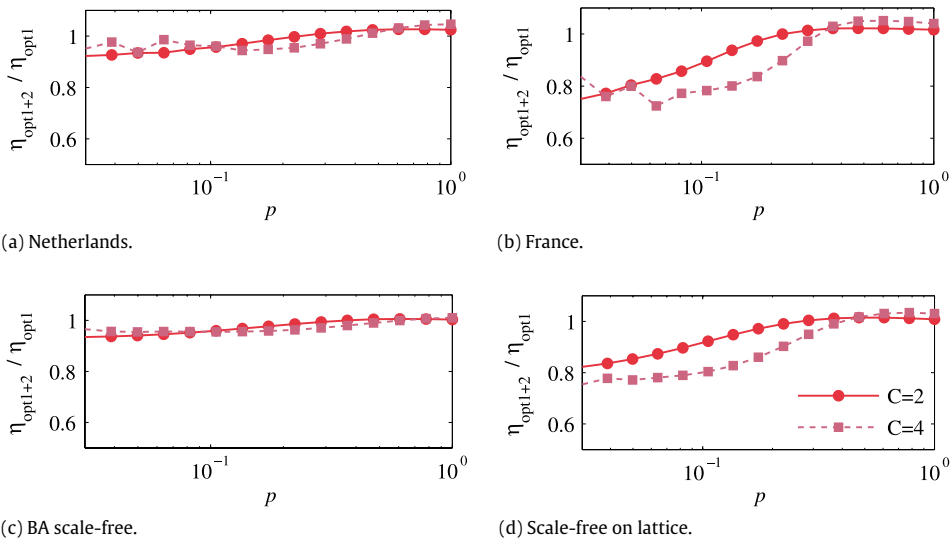


Fig. 8. Comparison between packet drop probability η as a function of packet-generating rate p for static weighted routing strategy, β_{opt} (η_{opt1}) and combined static and dynamic routing (η_{opt1+2}). We have set node transfer capacity $C = 2$ (circles) and $C = 4$ (squares).

node degree (comparable to the size of the network); cf. Ref. [36]. As result the number of deflected packets increases with the network size increasing congestion in the surrounding nodes. This problem is alleviated in the combination of the static weighted and dynamic routing algorithms; cf. Fig. 9(d). Since in this case the most congested nodes are avoided and the load is more evenly distributed, the dynamic routing can therefore further reduce congestion. We observe that improvement of network performance due to combined dynamic–static routing strategy is increasing with system size.

6. Conclusions

In summary, we have introduced an information flow model for networks without buffering capacity. We have shown that scale-free model on lattice reproduces well both topological and information transport characteristics of the small national research and educational networks of the Netherlands, France, Norway and Spain. In small networks, the point of the network congestion and amount of information lost strongly depends on the underlying network structure. We have further described a dynamic deflection routing strategy suitable for the networks without buffers. The proposed strategy dynamically extends packet path before it reaches a congested node. Simulations on two real world networks and two generic networks with small diameter confirmed the highest traffic capacity under combined static and dynamic routing is achieved when compared with the shortest path routing strategy and the static weighted routing strategy. Moreover, we

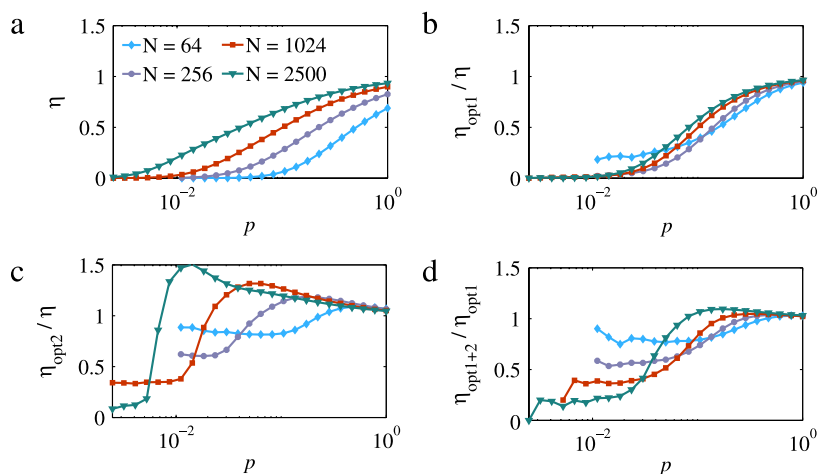


Fig. 9. Packet drop probability (η) in four scale-free on lattice networks with different sizes $N = 64, 256, 1024$ and 2500 . Results for different routing strategies are compared: (a) the shortest path routing algorithm, (b) static weighted routing strategy and shortest path routing (η_{opt1}/η), (c) dynamic routing algorithm and shortest path routing (η_{opt2}/η), and (d) combined static and dynamic routing strategies and static weighted routing strategy ($\eta_{opt1+2}/\eta_{opt1}$). The value of node transfer capacity used in this figure is fixed ($C = 4$).

found that the dynamic deflection routing can further improve network information transport capacity when combined with the efficient path routing strategy both in small and large networks.

Acknowledgments

Discussions with Jovan Radunović and Vladica Tintor helped to motivate this work. The authors acknowledge support by the Ministry of Science of the Republic of Serbia, under project No. ON171017. Numerical simulations were run on the AEGIS e-Infrastructure, supported in part by FP7 projects EGInSPIRE, PRACE-1IP, PRACE-2IP, and HP-SEE. The authors also acknowledge support received through SCOPES grant IZ73Z0-128169 of the Swiss National Science Foundation.

References

- [1] D.-H. Kim, A.E. Motter, *J. Phys. A: Math. Theor.* 41 (2008) 224019.
- [2] D.-H. Kim, A.E. Motter, *New J. Phys.* 10 (2008) 053022.
- [3] B. Danila, Y. Yu, J.A. Marsh, K.E. Bassler, *Phys. Rev. E* 74 (2006) 046106.
- [4] B. Danila, Y. Sun, K.E. Bassler, *Phys. Rev. E* 80 (2009) 066116.
- [5] G. Yan, T. Zhou, B. Hu, Z.-Q. Fu, B.-H. Wang, *Phys. Rev. E* 73 (2006) 046108.
- [6] M. Tang, T. Zhou, *Phys. Rev. E* 84 (2011) 026116.
- [7] R. Guimerà, A. Arenas, A. Díaz-Guilera, F. Giralt, *Phys. Rev. E* 66 (2002) 026704.
- [8] W.-X. Wang, B.-H. Wang, C.-Y. Yin, Y.-B. Xie, T. Zhou, *Phys. Rev. E* 73 (2006) 026111.
- [9] C.-Y. Yin, B.-H. Wang, W.-X. Wang, T. Zhou, H.-J. Yang, *Phys. Lett. A* 351 (2006) 220.
- [10] S. Sreenivasan, R. Cohen, E. Lopez, Z. Toroczkai, H.E. Stanley, *Phys. Rev. E* 75 (2007) 036105.
- [11] C.-L. Pu, S.-Y. Zhou, K. Wang, Y.-F. Zhang, W.-J. Pei, *Physica A* 391 (2012) 866.
- [12] W.-X. Wang, C.-Y. Yin, G. Yan, B.-H. Wang, *Phys. Rev. E* 74 (2006) 016101.
- [13] B. Kujawski, G.J. Rodgers, B. Tadić, *Lecture Notes in Comput. Sci.* 3993 (2006) 1024.
- [14] P. Echenique, J. Gómez-Gardeñes, Y. Moreno, *Phys. Rev. E* 70 (2004) 056105.
- [15] P. Echenique, J. Gómez-Gardeñes, Y. Moreno, *Europhys. Lett.* 71 (2005) 325.
- [16] X. Ling, M.-B. Hu, R. Jiang, Q.-S. Wu, *Phys. Rev. E* 81 (2010) 016113.
- [17] V. Tintor, J. Radunović, *Photonic Network Communications* 18 (2009) 55.
- [18] D. Wischik, In *ECOC, Scotland* (2005).
- [19] D. Wischik, N. McKeown, *ACM CCR* 35 (2) (2005) 75.
- [20] G. Raina, D. Towsley, D. Wischik, *ACM CCR* 35 (2) (2005) 79.
- [21] M. Enachescu, et al., *ACM CCR* 35 (2) (2005) 83.
- [22] M. Enachescu, et al. In *IEEE INFOCOM, Spain* (2006).
- [23] S. Gorinsky, A. Kantawala, J. Turner, In *ISCC, Spain* (2005).
- [24] S. Gorinsky, A. Kantawala, J. Turner, *Simulation* 83 (3) (2007) 245.
- [25] E.W.M. Wong, L.L.H. Andrew, T. Cui, B. Moran, A. Zalesky, R.S. Tucker, M. Zukerman, *J. Lightwave Technol.* 27 (14) (2009) 2817.
- [26] A.-L. Barabási, R. Albert, *Science* 286 (1999) 509.
- [27] S.N. Dorogovtsev, J.F.F. Mendes, *Evolution of Networks*, Oxford University Press, Oxford, 2003.
- [28] A.F. Rozenfeld, R. Cohen, D. ben-Avraham, S. Havlin, *Phys. Rev. Lett.* 89 (2002) 218701.
- [29] <http://www.surfnet.nl>.
- [30] <http://www.renater.fr>.
- [31] <https://www.uninett.no>.
- [32] <http://www.rediris.es>.
- [33] E.W. Dijkstra, *Numer. Math.* 1 (1959) 269.
- [34] K.-I. Goh, B. Kahng, D. Kim, *Phys. Rev. Lett.* 90 (2003) 058701.
- [35] S. Havlin, R. Cohen, *Complex Networks—Structure, Robustness and Function*, Cambridge University Press, Cambridge, 2010.
- [36] R. Cohen, S. Havlin, *Phys. Rev. Lett.* 87 (2001) 278701.