

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 Јелена Радовановић, редовни професор  
Универзитет у Београду - Електротехнички факултет

Чланови комисије:

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

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

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: \_\_\_\_\_

## 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

<b>1</b>	<b>Увод</b>	<b>1</b>
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
<b>2</b>	<b>Квантитативне методе</b>	<b>10</b>
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
<b>3</b>	<b>Подаци</b>	<b>33</b>
3.1	Научне конференције . . . . .	34
3.2	Meetup групе . . . . .	39
<b>4</b>	<b>Обрасци учешћа</b>	<b>45</b>
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
<b>5</b>	<b>Структура социјалних мрежа под утицајем учешћа на догађајима</b>	<b>65</b>
5.1	Анализа Meetup социјалних мрежа . . . . .	66
5.2	Утицај појединачних догађаја на карактеристике социјалне мреже	73
<b>6</b>	<b>Закључак</b>	<b>79</b>
	<b>Додатак А</b>	<b>84</b>
A.1	Аналитичко решавање модела у случају Бернулијевог процеса . . .	84
A.2	Аналитичко решавање модела у случају хомогеног Марковског ланца . . . . .	85
	<b>Литература</b>	<b>88</b>

# 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  $\theta$ . Код експоненцијалне расподеле то је параметар  $\lambda$ , 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  $\theta$  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  $\theta$  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  $\theta$ , 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  $\theta$ . 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  $\epsilon$ , одбацује се хипотеза да подаци потичу из посматране теоријске расподеле. За ниво значајности  $\epsilon$  најчешће се узимају вредности 0.1, 0.05 или 0.01. Препорука је да се у зависности од изабране вредности  $\epsilon$  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  $\epsilon$ , 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  $\theta_1$  и  $\theta_2$ , 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 <sup>1</sup>, 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  $\sigma^2$  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-

---

<sup>1</sup>nested 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  $\pi$  мања од одговарајуће вредности  $\epsilon$ , 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  $\chi^2$ . 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.

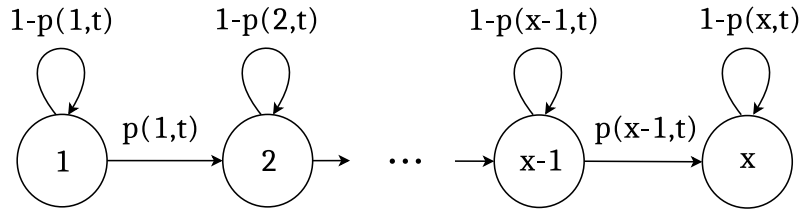
Један од начина за оцену параметара  $v = (v_1, v_2, \dots)$  математичког модела јесте поређење емпиријске расподеле  $p_N(x)$  и расподеле  $p(x|v)$  која се добија када се за опис система користи дати математички модел. Расподела  $p(x|v)$  се може добити аналитичким или нумеричким решавањем модела, 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  $\mu$  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  $\gamma$  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 убаци 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  $\rho$ , 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 монопартиитне мреже. У једној пројекцији бипартитне мреже чворови ће припадати скупу  $\mathcal{V}$ , а у другој скупу  $\mathcal{U}$ . У монопартиитној 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  $\pi_{ij}$  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  $\pi_{ij}$  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  $\pi_{ij}$ . Описани поступак за израчунавање вероватноће  $\pi_{ij}$  помоћу симулација предложен је у [105]. Недавно је у радовима [13, 103] приказан супериорнији метод помоћу кога се вероватноћа  $\pi_{ij}$  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  $\lambda_i$  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  $\lambda_i$ .

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  $\alpha_i$  и  $\beta_l$  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  $\pi_{ij}$  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  $\pi_{ij}$  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 сарђжати информације из само једног узорка из кога је узет почетни потпис. Онда се претражују сви групписани 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  $\tau_i$  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  $\tau_i$ .

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

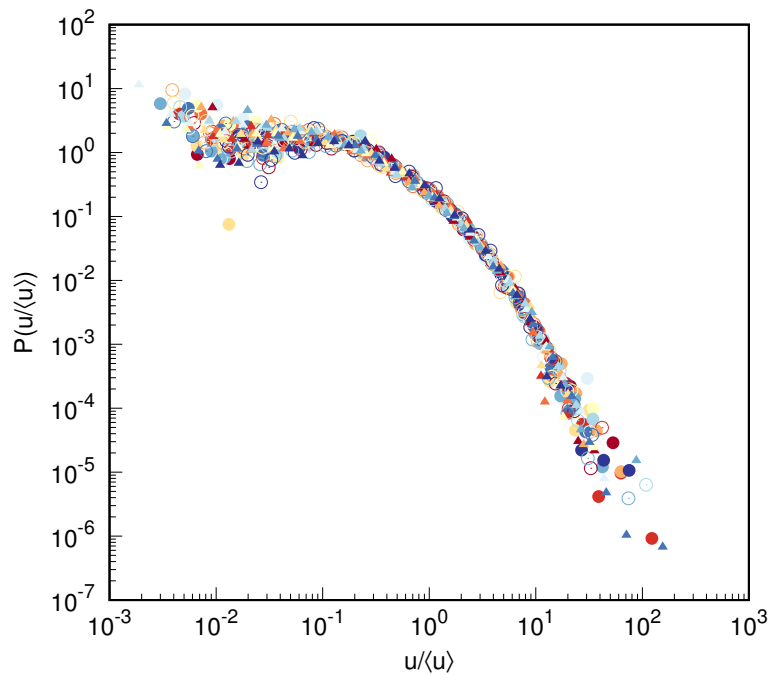
Табела 3.3: Преглед Meetup категорија.

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

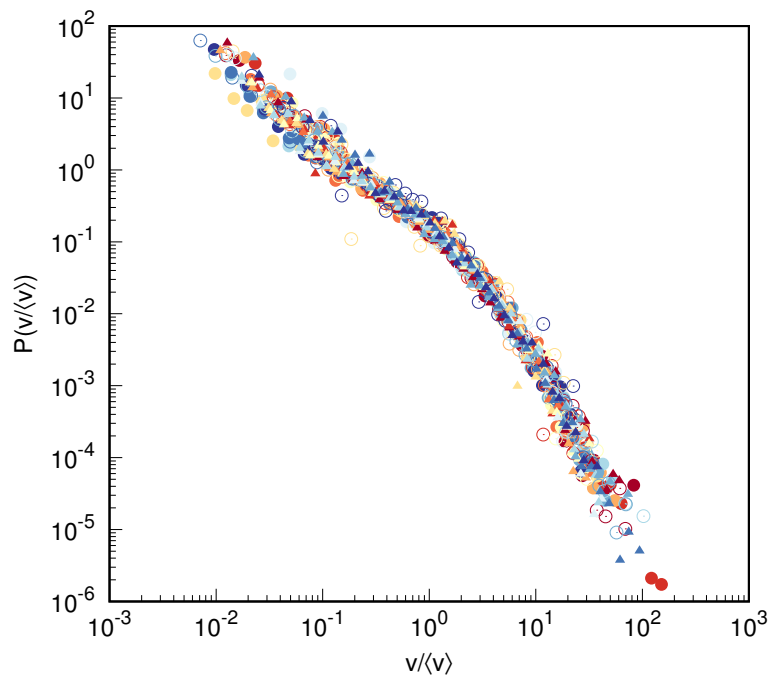
веома популаран ако већина чланова одлучи да учествује, али исто тако постоји могућност да нико од чланова не потврди учествовање на неком догађају. Број чланова и учестаност догађаја се мењају током времена и разликују се од групе до групе. Приликом избора групе, веома битну улогу има географска локација. 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.

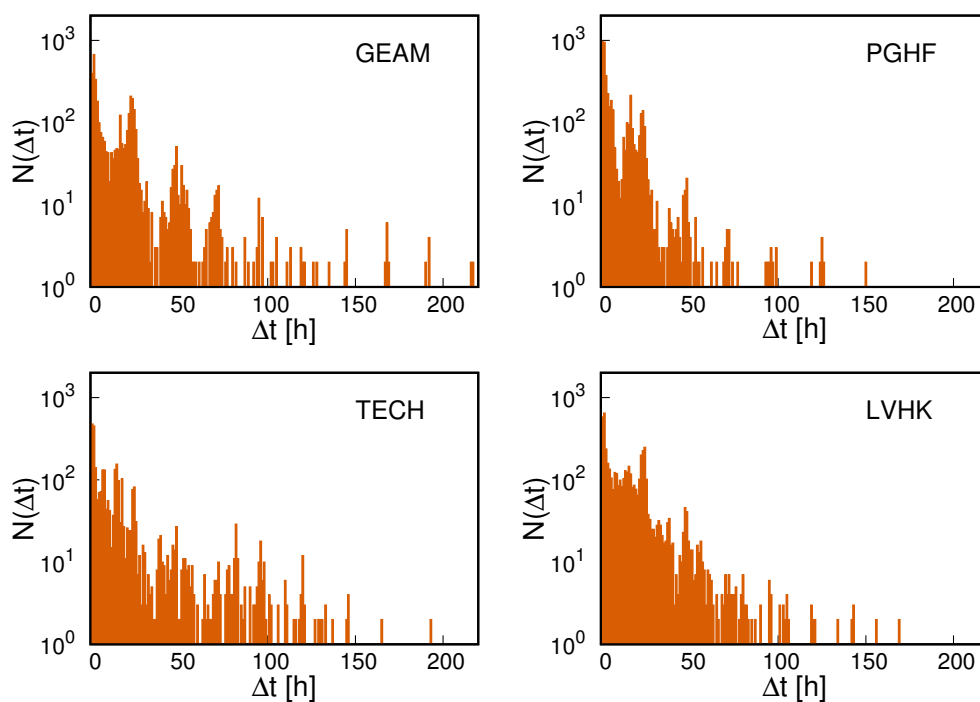
Табела 3.4: Приказ четири изабране Meetup групе.

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

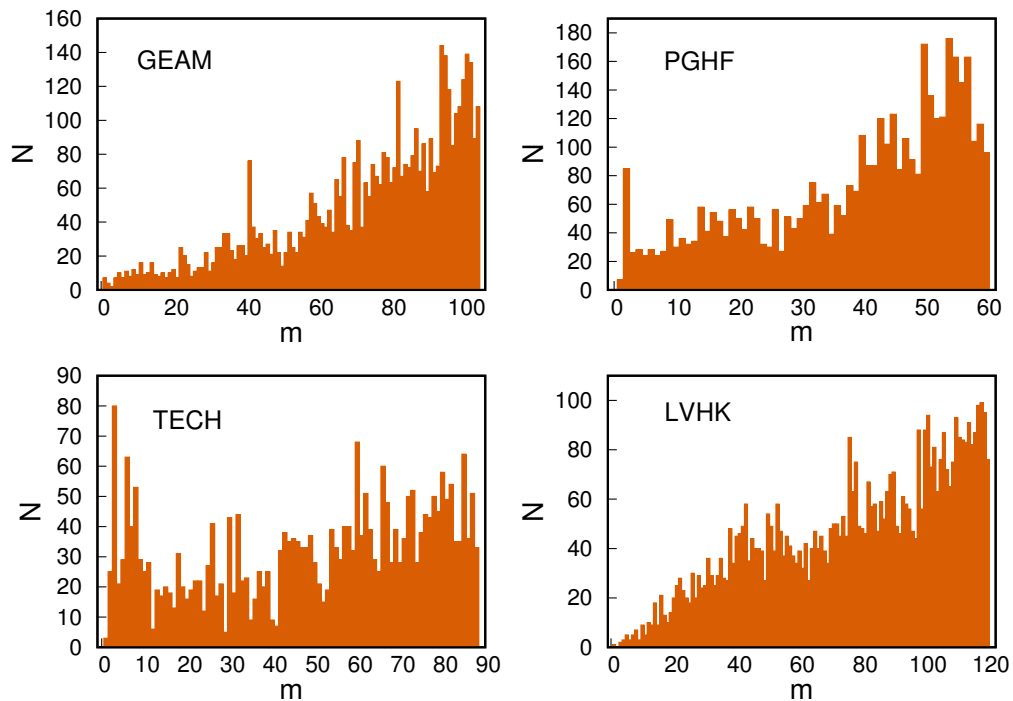
Група geamclt (GEAM) регистрована је у Шарлоту у Сједињеним Америчким државама. Њени чланови се окупљају у ресторанима и баровима како би пробали различита јела и коктеле. У групи pittsburgh-free (PGHF), чланови заједно иду на дешавања у Питсбургу, као што су концерти, фестивали, обиласци музеја и код 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,  $\Delta t$  представља време мерено у сатима. На y-оси је дат број пауза које трају  $\Delta t$ .



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.

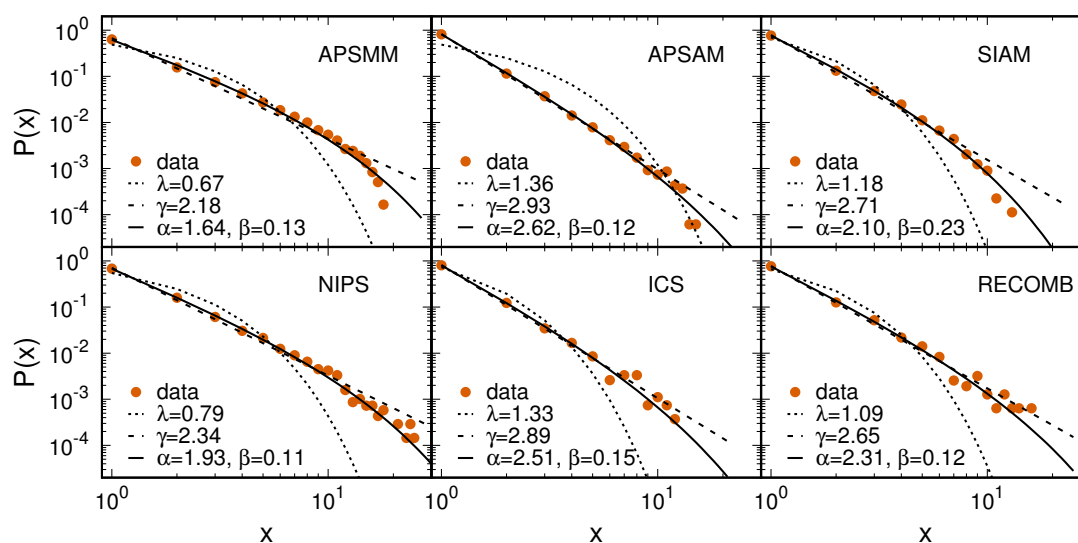
### 4.1 Емпиријска анализа

#### 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  $\lambda$ . 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  $\gamma$ , а пуна линија степеној расподели са експоненцијалним репом и параметрима  $\alpha$  и  $\beta$ . 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



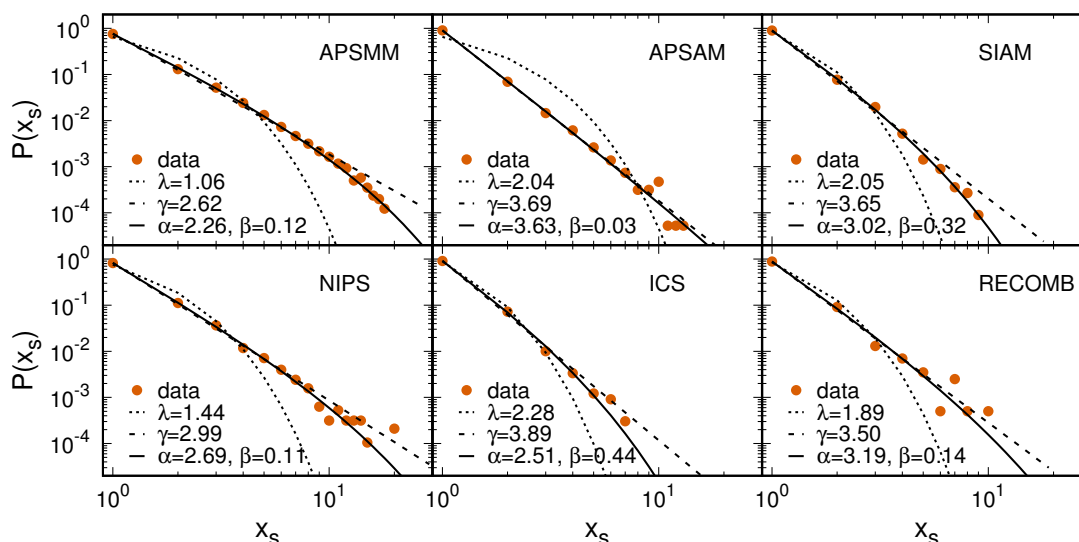
флуктуација, као што је објашњено у поглављу 2. Ова хипотеза је одбачена, пошто је вероватноћа  $\pi \simeq 0$ , као што је приказано у табели 4.1. Из приказаних резултата, може се закључити да вероватноћа да учесник има  $x$  учествовања на почетку опада по степеном закону са параметром  $\alpha > 1.5$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

Табела 4.1: Вредности параметара  $\mathcal{R}$  и  $\pi$  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	$\mathcal{R}$	$\pi$
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



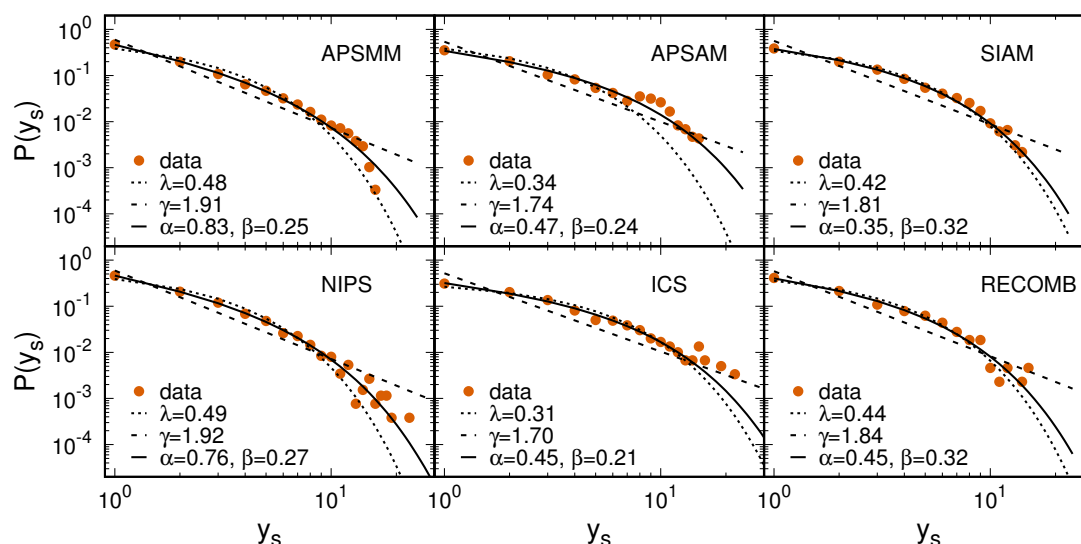
Слика 4.2: Расподела броја узастопних учествовања без пауза на научним конференцијама. Наранцастим тачкама приказана је емпиријска расподела. Тачкаста линија одговара експоненцијалној расподели, испрекидана линија степеној расподели и пуна линија одговара степеној расподели са експоненцијалним репом.

расподеле  $\alpha$  и  $\beta$  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  $\pi$  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	$\mathcal{R}$	$\pi$
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.

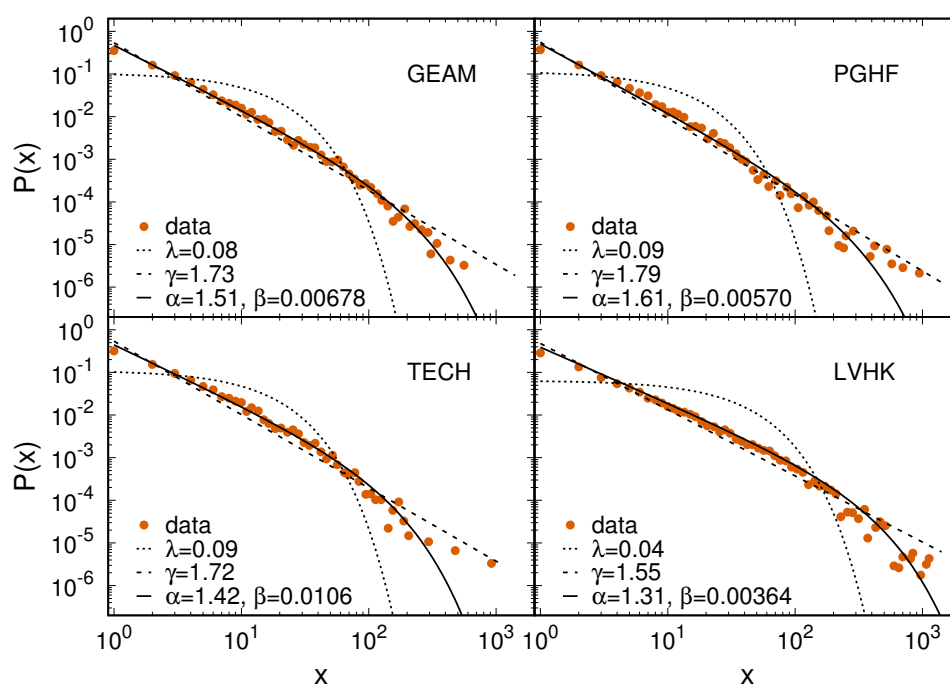
Табела 4.3: Вредности параметара  $\mathcal{R}$  и  $\pi$  након примене метода количника веродостојности за експоненцијалну расподелу и степену расподелу са експоненцијалним репом када се посматрају узорци који одговарају трајању паузе између два учествовања на конференцији.

Конференција	$\mathcal{R}$	$\pi$
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



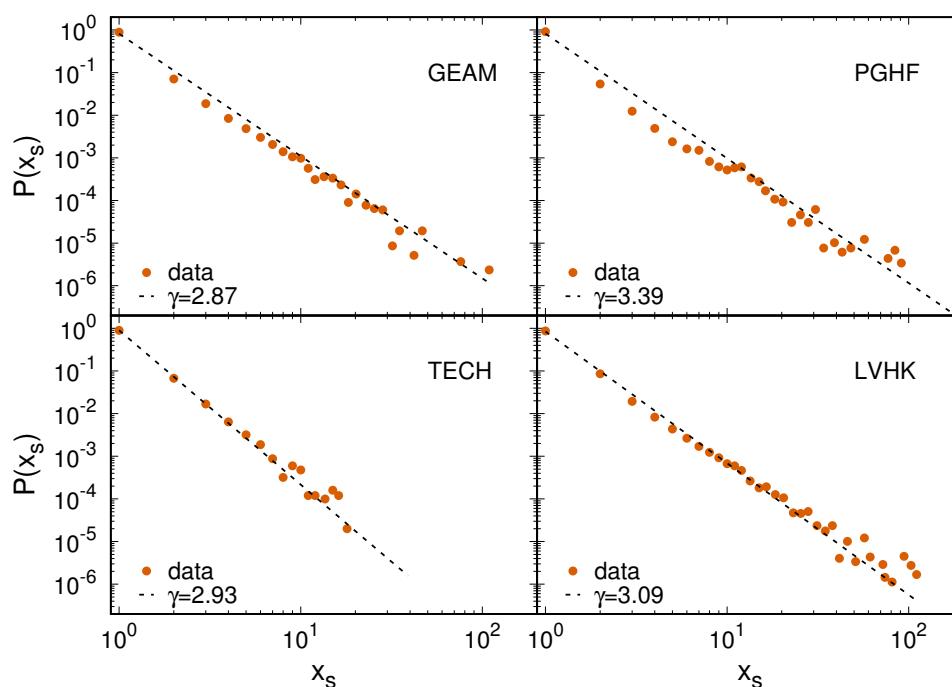
Слика 4.4: Расподела укупног броја учествовања на Meetup догађајима. Наранџастим тачкама приказана је емпиријска расподела. Тачкаста линија одговара експоненцијалној расподели, испрекидана линија степеној расподели и пуна линија одговара степеној расподели са експоненцијалним репом.

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  $\beta$  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  $\pi$  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.

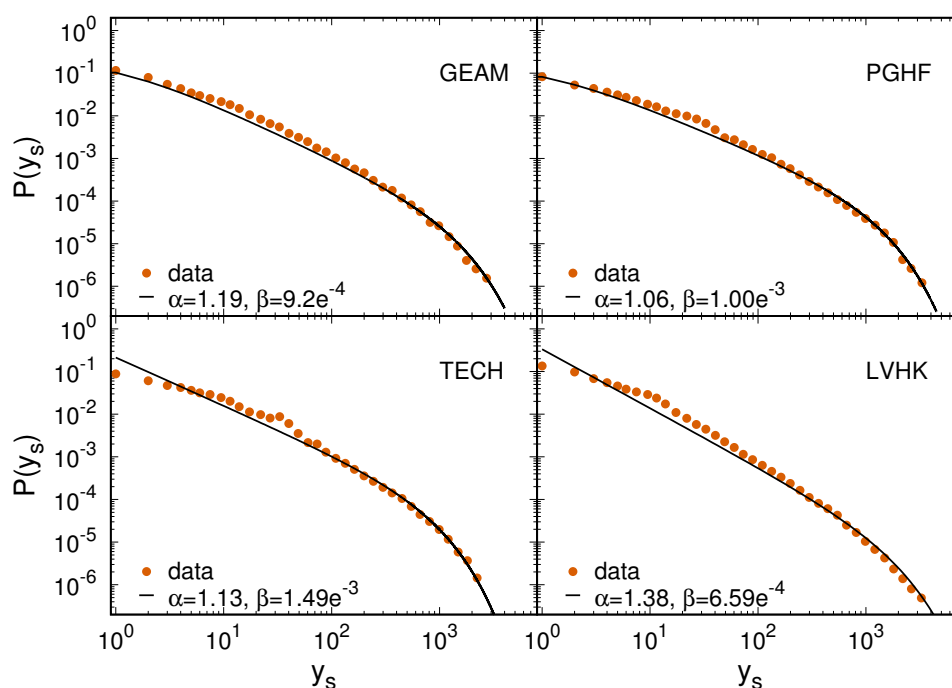
	$\mathcal{R}$	$\pi$
GEAM	-97.70	0.0
PGHF	-58.04	0.0
TECH	-90.84	0.0
LVHK	-236.52	0.0

Када се погледа расподела броја узастопних догађаја без изостанака на слици 4.5, може се видети да она опада по степеном закону са параметром  $\gamma$  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 расподели.

У склопу анализе динамике учествовања на Meetup догађајима, одређена је

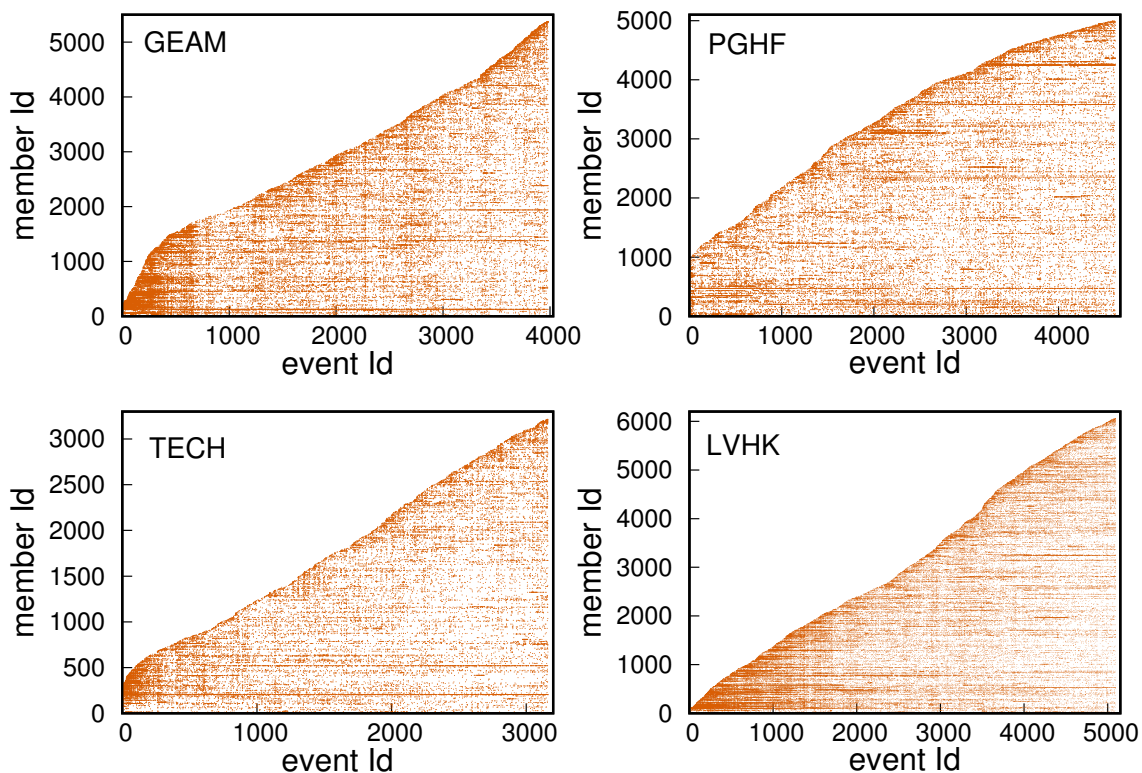


Слика 4.6: Расподела дужине паузе између два учествовања на Meetup догађајима. Наранџастим тачкама приказана је емпиријска расподела, док пуна линија одговара степеној расподели са експоненцијалним репом.

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



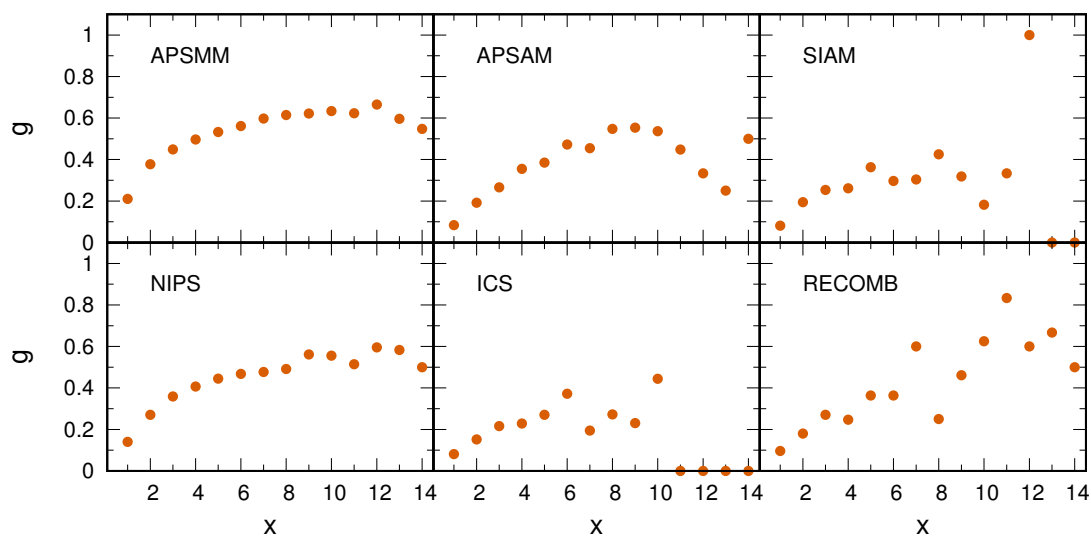
многобројни примери из реалног света где постоје изразито неједнаке расподеле код којих је одступање од средње вредности веома велико, као што су степена расподела и степена расподела са експоненцијалним репом [114], што је случај и код података који су овде посматрани. Ова појава се објашњава Матејевим 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.

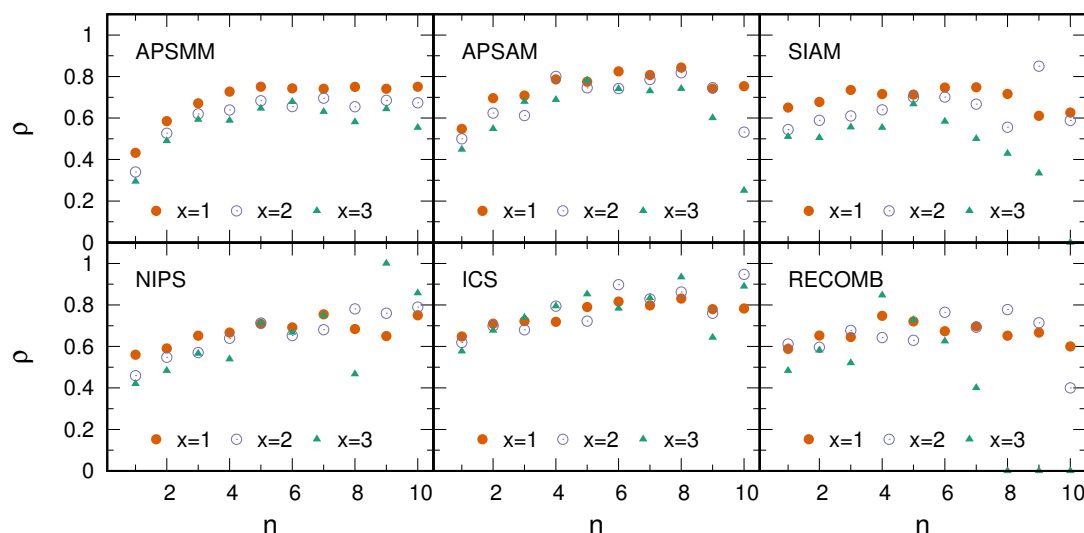
Пошто је утврђено да је претпоставка да вероватноћа појављивања на конференцији зависи од претходних долазака оправдана, размотрен је математич-



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  $\rho$  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.

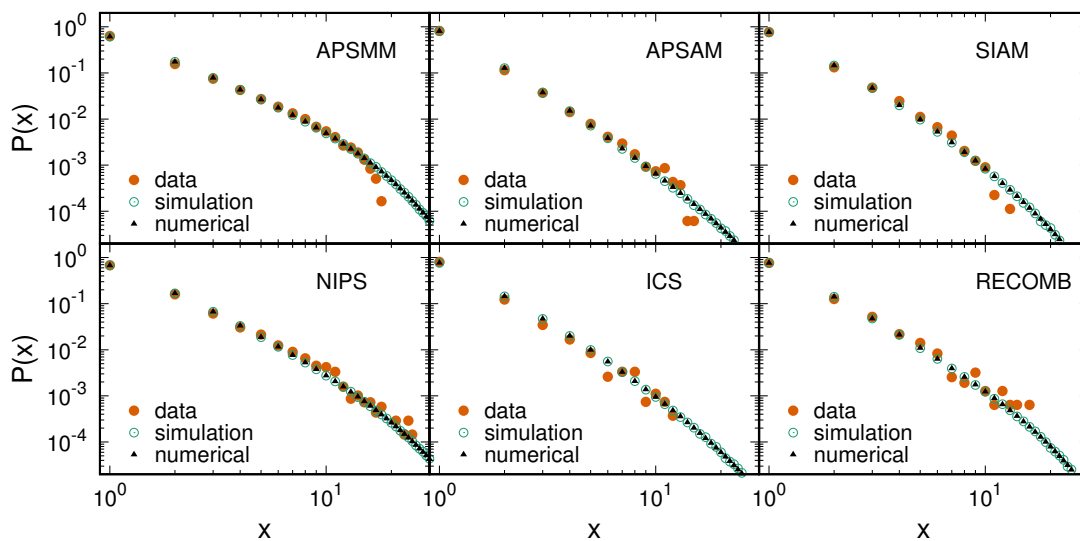
Оптималне вредности параметара за шест посматраних конференција дате су у табели 4.5. Као што се може видети, свакој конференцији одговарају различити параметри. Треба приметити да је вредност параметра  $a > 1.5$ , што значи да вероватноћа учествовања брзо расте када учесник редовно долази на конференцију, 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%.

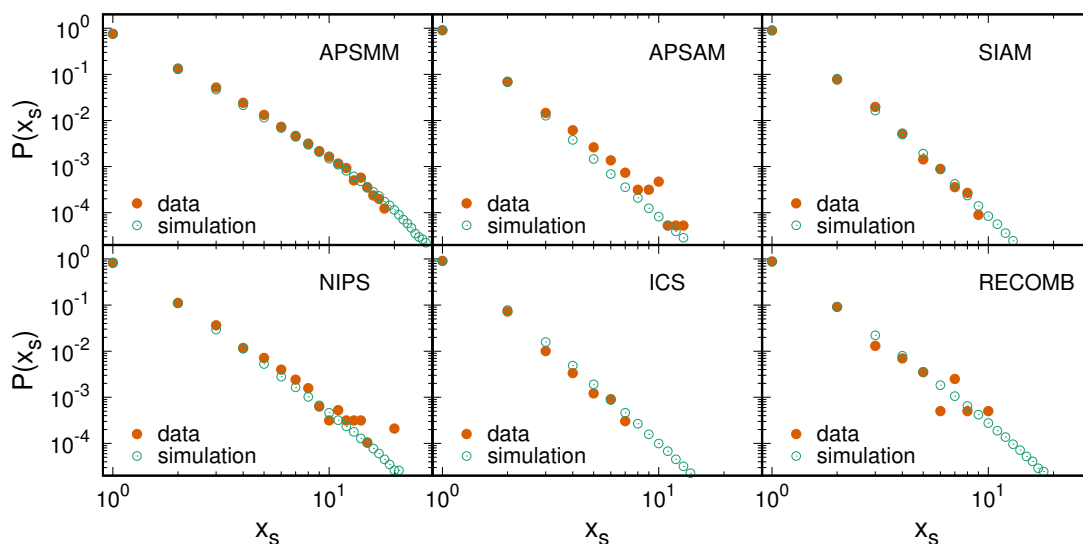
Табела 4.5: Оптималне вредности параметара модела за сваку од посматраних конференција.

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

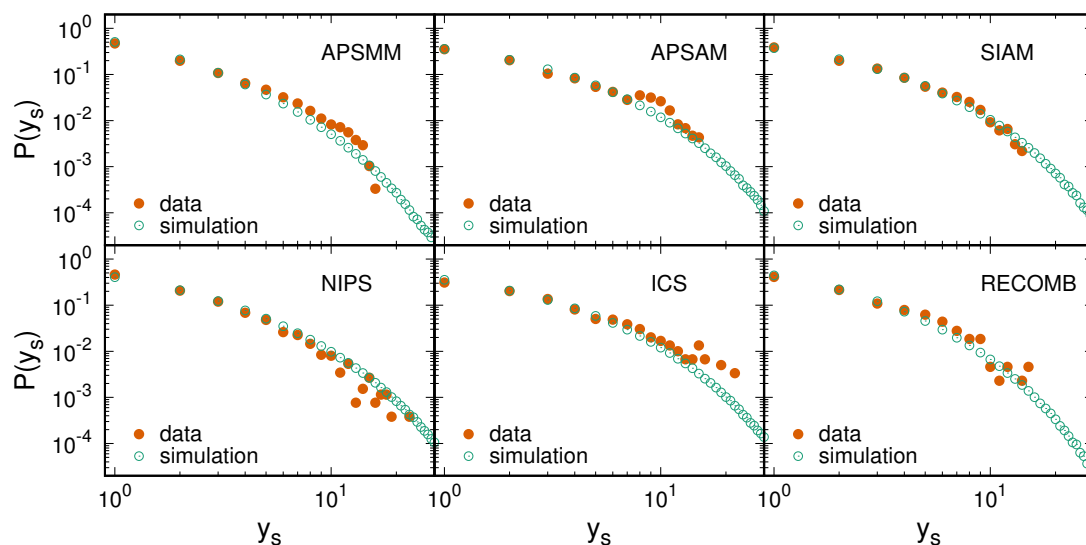
Са слике 4.10 се види да су расподела укупног броја учествовања на конференцији на основу прикупљених емпиријских података и расподела применом математичког модела са оптималним параметрима веома сличне. Исто важи и за расподелу броја узастопних учествовања без пауза, која је приказана на слици 4.11, 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. Наранцастим 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-

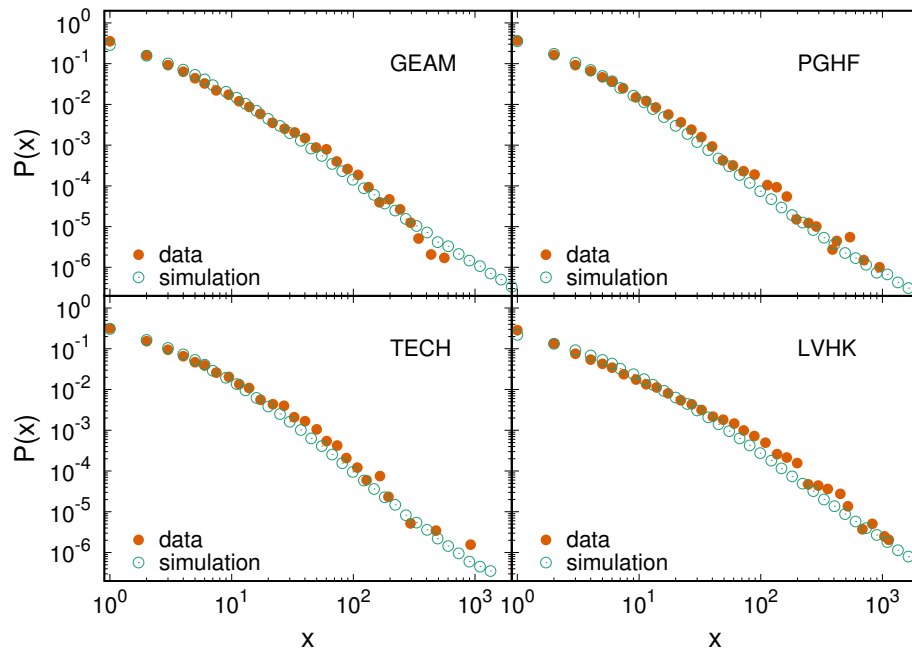
рализованим бинарним Појиним моделом. Параметри модела  $a$  и  $y_0$ , одређени су на потпуно исти начин као и код научних конференција. У свакој симулацији посматрано је 5000 догађаја и укупно  $10^5$  различитих учесника. Оптималне вредности параметара, које су добијене као резултат, приказане су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

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

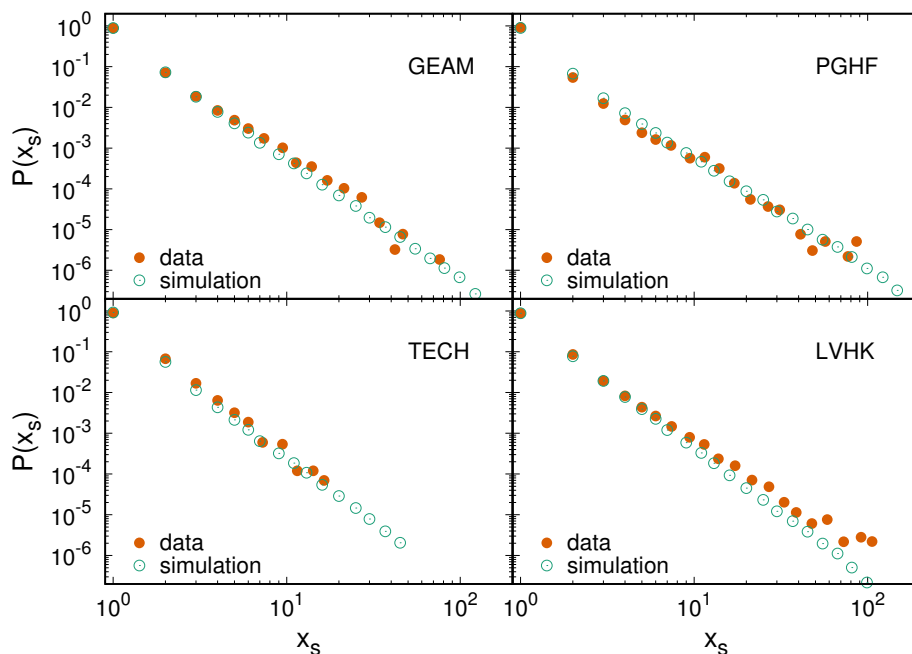
Конференција	$a$	$y_0$
GEAM	1.35	7
PGHF	1.40	7
TECH	1.35	8
LVNK	1.30	7

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

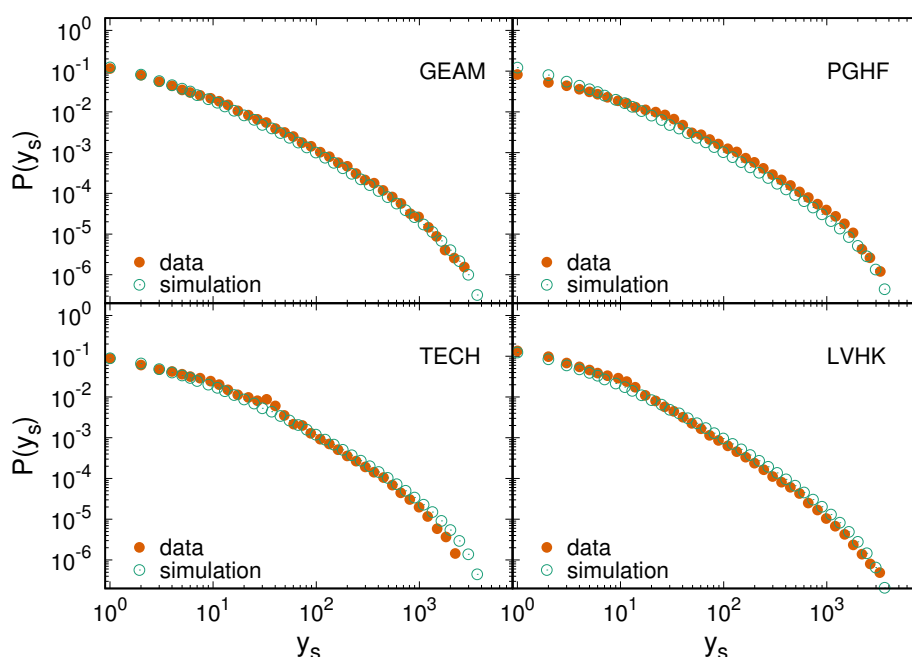


Слика 4.13: Расподела укупног броја учествовања на Meetup догађајима. Наранџастим тачкама приказана је расподела на основу података прикупљених са Meetup портала, а зеленим кружићима на основу резултата симулација.



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-

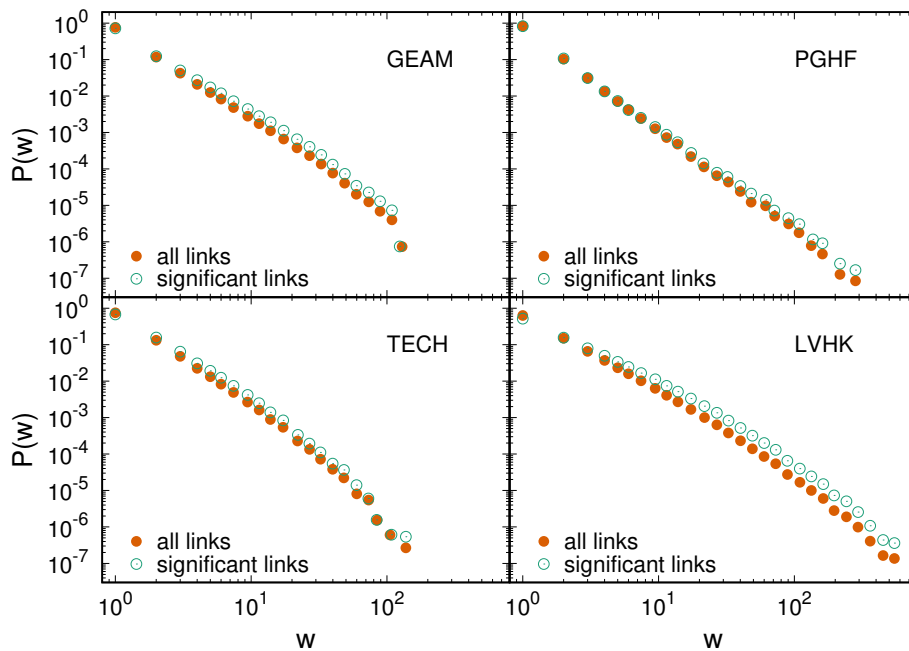
## **5. Структура социјалних мрежа под утицајем учешћа на догађајима**

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

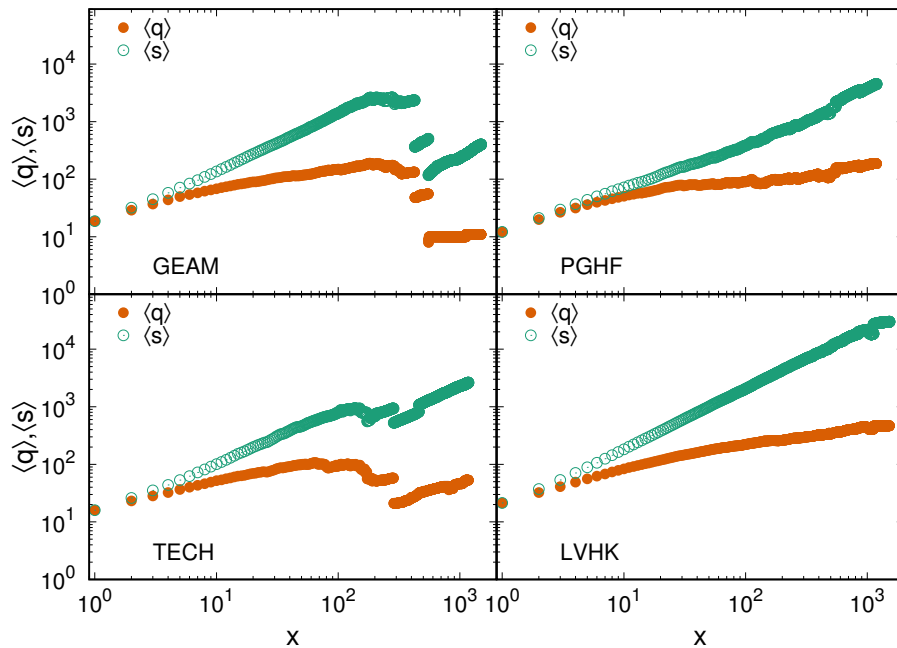


Слика 5.1: Расподела тежина грана у пројектованој монопартитној мрежи пре филтрирања (наранџасте тачке) и након филтрирања (зелени кружићи).

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



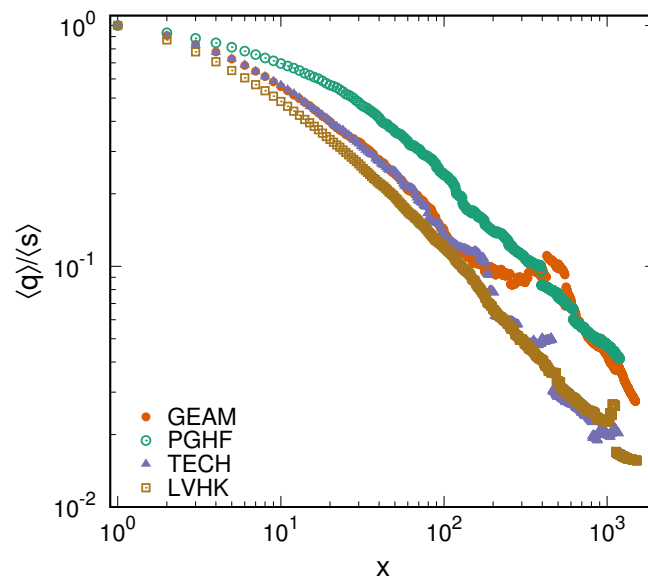
Слика 5.2: Зависност средњег степена  $\langle q \rangle$  и средње снаге чвора  $\langle s \rangle$  од броја учествовања  $x$ .

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

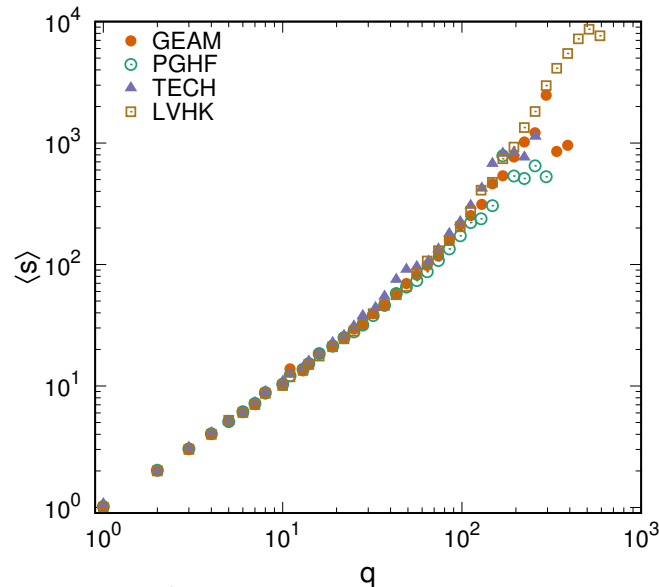


Слика 5.3: Промена односа између средњег степена  $\langle q \rangle$  и средње снаге чвора  $\langle s \rangle$  са бројем учествовања  $x$ .

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.

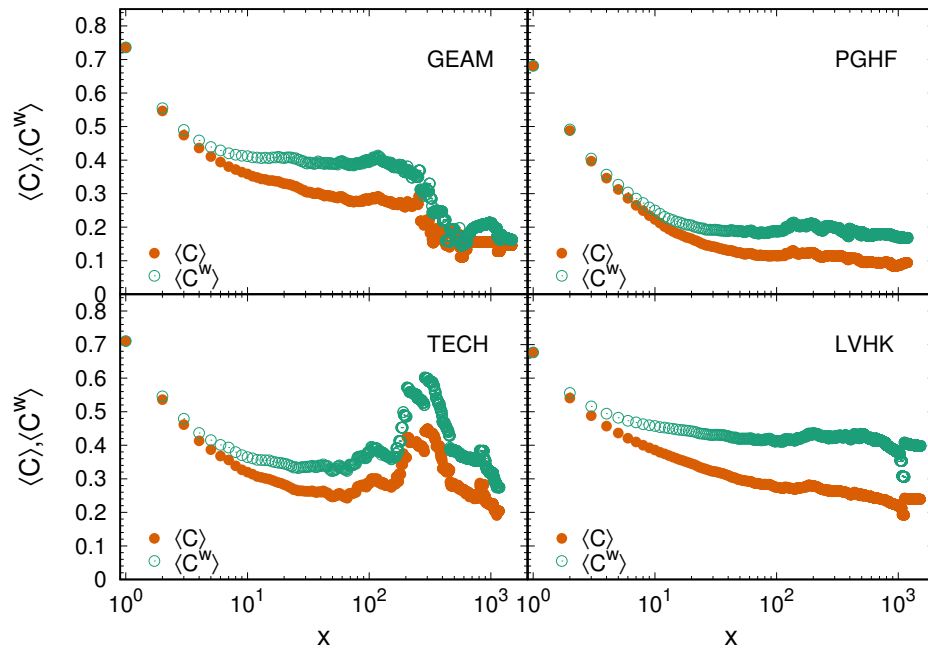
После првог учествовања на догађају, везе корисника у социјалној мрежи су слабе и вероватноћа учествовања на следећим догађајима је мала. Активним учествовањем корисник ће истовремено учврстити своје везе у мрежи и повећати изгледе за долазак на следеће окупљање, при чему та одлука зависи и од одлуке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

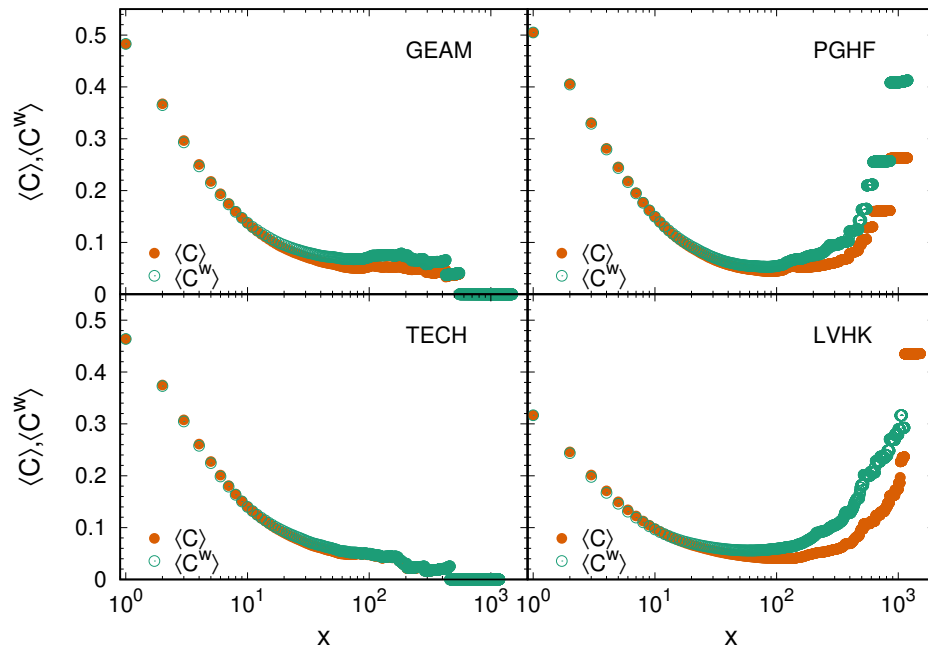
Да би се додатно истакао значај резултата који су добијени за вредности

## 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 монопартидне мреже која се добија пројекцијом бипартидне 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 бипартидне 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.

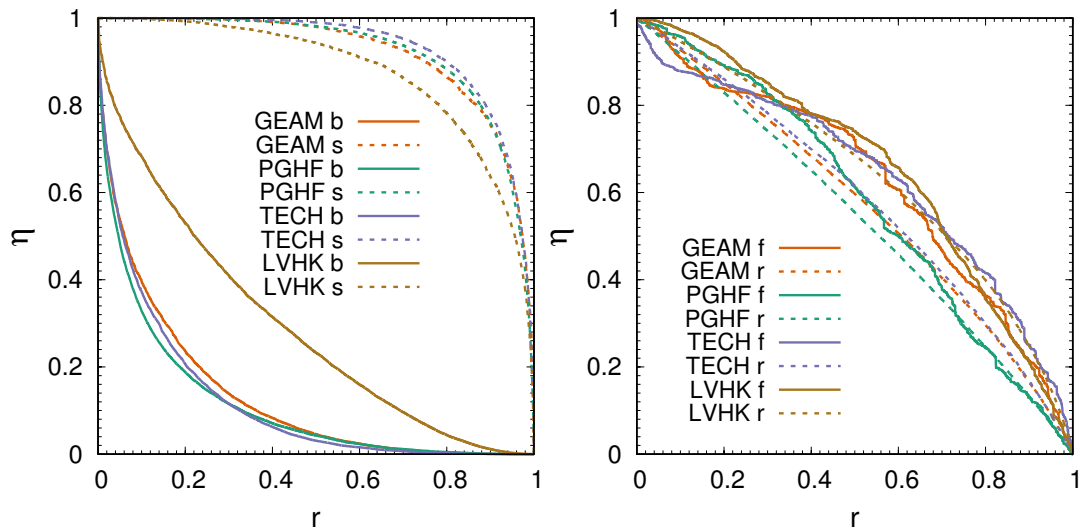
Посматрана су два параметра. Први се односи на проценат грана у мрежи  $\eta$ , које би постојале и да поједини догађаји нису организовани. Овај параметар је дефинисан као однос броја грана у мрежи  $E^r$  који преостане када се одстрани удео од  $r$  догађаја и укупног броја грана у комплетној мрежи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

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

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

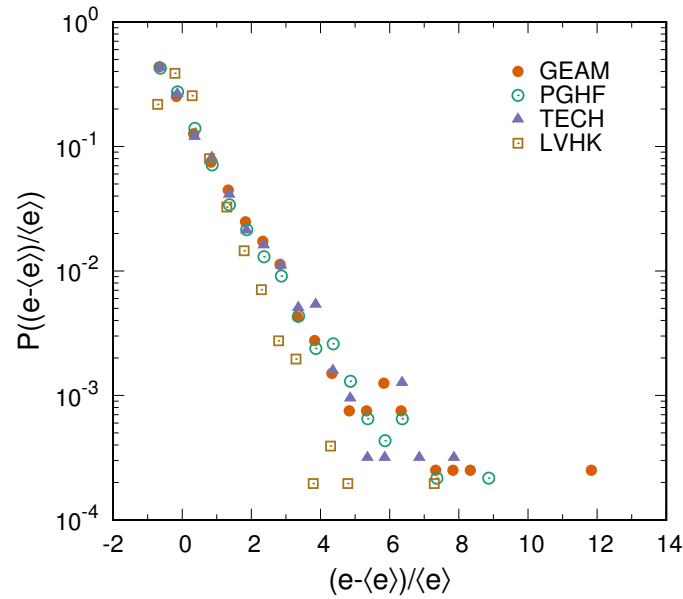


Слика 5.7: Утицај појединачних догађаја на међусобну повезаност чворова у мрежи. Слика лево: догађаји су уклањани по величини, од највећег ка најмањем (b) и обрнуто, од најмањег ка највећем (s). Слика десно: догађаји су уклањани по редоследу одржавања (f) и насумично (r).

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  $\eta$  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



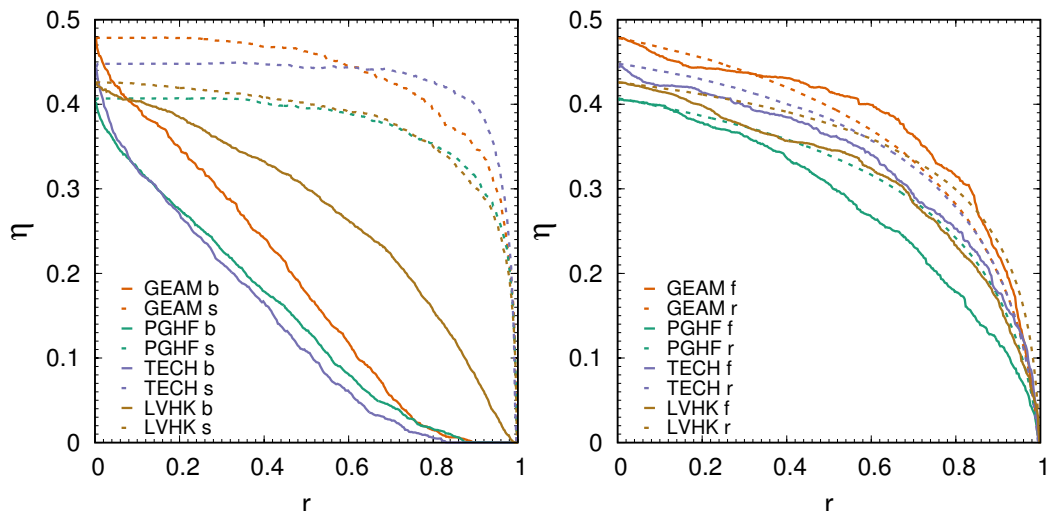
Слика 5.8: Расподела релативних флукуација величина догађаја  $\frac{e-\langle e \rangle}{\langle e \rangle}$ , где је са  $e$  означена величина појединачних догађаја, а са  $\langle e \rangle$  просечна величина догађаја у групи.

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  $\eta$  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  $\eta$  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  $\eta$  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 уклањају по величини, од највећег ка најмањем (b) и обрнуто, од најмањег ка највећем (s). Слика десно: догађаји су уклањају 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 бипартидне мреже зависи од његове величине (слика 5.9 лево), важиће слични закључци као и када је посматран релативан број грана  $\eta$ . 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. Поглавље 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 учешћа 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.** Дозвољаваате 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.** Дозвољаваате 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.** Дозвољаваате 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.** Дозвољаваате 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.** Дозвољаваате 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.** Дозвољаваате 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.