

Математички факултет
Универзитет у Београду

Мастер рад:

**Асимптотска својства оцена неких обележја случајног графа као
модела мреже**

Студенткиња:

Елеонора Креачић

Ментор:

Проф. др Павле Младеновић

Октобар 2012. године

Садржај

Увод	1
I Концепт случајног графа	3
I.1. Мотиви проучавања случајних графова	3
I.2. Основе теорије графова	5
I.3. Концепт случајног графа и неке важније статистике случајних графова	10
II Erdős-Rényi-ев модел случајног графа и нека асимптотска својства	14
III Неки аспекти непараметарског приступа моделу случајног графа	17
III.1. Параметарски и непараметарски приступ моделу случајног графа	17
III.2. Заменљивост случајних величина	18
III.3. Поставка модела	21
III.4. Нека асимптотска својства U -статистика	26
III.5. Асимптотска својства модела-доказ <i>Теореме III.3.1.</i>	30
IV Неки актуелни изазови проучавања случајних графова	39
Списак литературе	41

Увод

Основни циљ овог мастер рада је приказ неких резултата из рада *The method of moments and degree distributions for network models* аутора *P. J. Bickel-a, A. Chen-a, и E. Levina-e*, објављеног у *The Annals of Statistics, 2011, Vol. 39, No. 5, 2280-2301*. Резултат који ће бити приказан односи се на део *Теореме 1* овог рада. Та теорема утврђује асимптотска својства неких статистика случајних графова из посматране класе случајних графова, када број чворова графа тежи бесконачности. Статистике које се посматрају представљају оцену вероватноће гране, и (скалирану) оцену вероватноће појаве одређеног подграфа у графу. Класа случајних графова која се посматра задата је у непараметарском контексту, односно одговарајућим скупом услова које њени елементи задовољавају.

Рад сачињавају четири целине.

Прва целина објашњава концепт случајног графа. У првом одељку ове целине објашњена је мотивација проучавања случајних графова кроз приказ неких система у природи за чије је проучавање значајно посматрати везе, односно интеракције између елемената који сачињавају систем. У другом одељку ове целине приказане су основе теорије графова неопходне за развијање теорије случајних графова. У последњем одељку ове целине уведен је појам случајног графа, и наведене неке од статистика које се најчешће користе при карактерисању особина случајних графова.

Друга целина рада односи се на нека асимптотска својства Erdős-Rényi-евог модела. Асимптотска својства овог модела значајно је приказати из неколико разлога. Прво, историјски посматрано, Erdős-Rényi-ев модел је први описани модел случајног графа, на основу кога је ова, данас богата теорија, почела да се развија. Својства која ћемо посматрати односе се на присутност одређених подструктура у графу. Друго, асимптотска својства која ће бити посматрана имају једну особину типичну за случајне графове уопште: она се нагло мењају. Ова тенденција ка наглој промени карактеристика огледа се у чињеници да, након „пробијања“ одговарајуће асимптотске баријере, када број чворова графа тежи ка бесконачности, вероватноћа да граф поседује одговарајуће својство се мења са нула на један. Треће, проучавање Erdős-Rényi-евог модела је пример посматрања конкретного модела случајног графа, које представља полазну тачку тзв. параметарског приступа моделирању мрежа.

Трећа целина представља главни део рада. У њеном првом одељку објашњава се разлика између параметарског, и непараметарског приступа моделирању мрежа, који је коришћен у раду чији резултат желимо да прикажемо. У другом одељку, уведен је концепт заменљивости случајних величина који је неопходан за резумавање модела који ће бити уведен у трећем одељку ове целине. У трећем одељку постављен је модел случајног графа у непараметарском контексту, и уведене су статистике које мере вероватноћу реализације гране између два чвора графа, као и вероватноћу појаве одређеног подграфа у графу. На крају одељка, дат је исказ теореме која утврђује нека асимптотска својства ових статистика. Како је ради разумевања доказа теореме потребно познавати нека својства U -статистика, у наредном,

четвртом одељку ове целине, уведен је појам U -статистике и показана су нека од њених својстава. Последњи, пети одељак ове целине, представља доказ поменуте теореме.

У четвртој целини поменути су неки од недавно добијених резултата проучавања случајних графова, у циљу наговештаја неких од актуелних изазова који се сусрећу при истраживањима у овој области статистике.

I Концепт случајног графа

I.1. Мотиви проучавања случајних графова

Многе системе у природи карактерише сложена структура, која се огледа у интеракцији великог броја ентитета који сачињавају тај систем. Интеракција, или повезаност тих ентитета, није детерминистичке природе. То значи да постоје ентитети система за које је могуће да међусобно интерагују, али то није неопходно. Зато је, при проучавању таквих система, њихових особина и понашања, потребно посебну пажњу посветити међусобном односу ентитета који тај систем сачињавају. Такви системи називају се мрежама, и одређени су одговарајућим скупом ентитета, односно чворова, и скупом веза или интеракција, међу тим чворовима. Проучавање мрежа заснива се на идентификовању модела који их, у извесном смислу, довољно добро описују. Како су везе у мрежама стохастичке природе, проблем идентификације одговарајућег модела природно спада у домен математичке статистике.

Ради сагледавања значаја, као и сложености проблема идентификовања модела мрежа, наведимо неке примере структура у природи које је погодно описати као мреже, у смислу претходне дискусије.

Међусобна интеракција протеина карактерише скоро све процесе који се одвијају у ћелији- ћелијски метаболизам, репликација ДНК, и раст ћелије само су неки од примера ових процеса¹. Природан приступ проучавању ових процеса, из перспективе статистике, представља посматрање мреже чији су чворови одговарајући протеини, а везе међу чворовима одговарајуће интеракције.

Екосистеми, односно врсте које у њима живе и одговарајуће релације исхране које важе међу врстама, одређују мрежу чијим се статистичким својствима можемо бавити². Проблем који се, међутим, често јавља при моделирању оваквих мрежа, јесте премали број врста у екосистему. Уочимо, такође, да у овој мрежи повезани чворови нису “равноправни”- позиција врсте која представља плен, и врсте која лови, суштински се разликују. Тако долазимо до концепта усмерених мрежа- мрежа чији актери интеракције имају суштински различите позиције.

World Wide Web је, по својој природи, мрежа чије чворове представљају *HTML* документи, а везе линкови међу њима. Пошто је ово огроман систем, математичка статистика је неопходна при његовом проучавању.

Интересантан пример система који је могуће посматрати као мрежу, јесте језик. Наиме, чворовима можемо сматрати речи језика, а одговарајућу везу између речи можемо дефинисати на разне начине. Рецимо, могуће је дефинисати следећу везу- две речи су повезане ако постоји било која (смислена) реченица тог језика у којој су посматране две речи суседне³. Мада је, на први поглед, овако креирана мрежа вештачка творевина, њено проучавање може послужити у тестирању одговарајућих хипотеза из домена лингвистике.

¹ Погледати у [21].

² Погледати у [10].

³ Погледати у [10].

Банкарски сектор једне земље, односно банке, као чворови, и токови новца између њих, као одговарајуће интеракције, одређују мрежу⁴. Посматрано из перспективе финансијске стабилности, значајно је утврдити својства ове мреже која се односе на њену рањивост на шокове, на пример, идентификовати системски значајне банке, односно чворове који наглим променама својстава интеракција са осталим чворовима могу променити особине система, односно мреже, као целине.

Значајан социолошки феномен данашњице свакако су друштвене мреже, чији је *Facebook* вероватно најраспрострањенији представник. Идентификација кључних особина ове мреже свакако ће имати плодноне резултате у развоју социолошких теорија.

Из наведених примера јасно се види да моделирање мрежа представља моћан алат у трагању за одговорима на многа од најинтригантнијих питања савремене научне мисли. Зато чињеница да је ова област једно од најживљих поља развоја математичке статистике данас, нимало не чуди.

Моделирање мрежа заснива се на теорији случајних графова. Концепт случајног графа биће објашњен у даљем тексту.

⁴ Погледати у [20].

I.2. Основе теорије графова⁵

Означимо са $\mathcal{P}_k(V)$ скуп подскупова од k елемената скупа V .

Деф I.2.1. Тривијални граф G је пар $G = (V, E)$, при чему је:

- 1) V коначан скуп чије елементе називамо чворовима
- 2) E подскуп скупа $\mathcal{P}_2(V)$, чије елементе називамо гранама.

Деф I.2.2. Графом називамо тројку $G = (V, E, \varphi)$ где је:

- 1) V коначан скуп чије елементе називамо чворовима
- 2) E коначан скуп, чије елементе називамо гранама
- 3) φ функција са доменом E и кодоменом $\mathcal{P}_2(V)$, односно $\varphi: E \rightarrow \mathcal{P}_2(V)$.

Уколико за $e \in E$ и $u, v \in V$ важи $\varphi(e) = \{u, v\}$, кажемо да су чворови u и v спојени, или повезани граном e .

Деф I.2.3. Ред графа представља број његових чворова.

Деф I.2.4. Величина графа јесте број његових грана.

Деф I.2.5. Степен чвора је број грана које из њега полазе.

Уочимо да су, првом дефиницијом, гране одређене као парови чворова, док је другом дефиницијом свакој грани додељен пар чворова. Из овога следи да *Деф I.2.1.* дозвољава највише једну грану између било која два чвора, док *Деф I.2.2.* дозвољава да су два чвора повезана произвољним бројем грана. Уочимо да, уколико за тривијални граф $G = (V, E)$ дефинишемо функцију $\varphi(e) = e$ за $e \in E \subseteq \mathcal{P}_2(V)$, тада је тако одређена тројка $G = (V, E, \varphi)$ граф у смислу друге дефиниције, па је тривијални граф заправо специјалан случај графа.

Нагласимо да овом дефиницијом нису предвиђени графови са тзв. петљама, односно гранама које повезују чвор са самим собом. Заиста, тако одређена “грana” $\{u, u\}$ није елемент скупа $\mathcal{P}_2(V)$, због $\{u, u\} = \{u\}$, што није у складу са *Деф I.2.1.* Такође, у контексту *Деф I.2.2.*, за тако одређену грану било би $\varphi(e) \notin \mathcal{P}_2(V)$ што противречи дефиницији. Овај проблем се, међутим, може превазићи проширењем кодоменског скупа функције φ (односно скупа грана, у смислу *Деф I.2.1*) на $\mathcal{P}_2(V) \cup \mathcal{P}_1(V)$. Пошто се нећемо бавити графовима с петљама, остаћемо при датим дефиницијама.

Деф I.2.6. Матрица повезаности тривијалног графа $G = (V, E)$, где је $|V| = n$, је матрица A одређена са:

$$A_{ij} = \begin{cases} 1, & (i, j) \in E \\ 0, & (i, j) \notin E \end{cases} \text{ за } \forall i, j \in V$$

Приметимо да смо, у претходној дефиницији, имплицитно користили обележавање темена графа природним бројевима $\{1, \dots, n\}$. Уочимо да бисмо, другачијим обележавањем темена, односно одабиром друге бијекције скупа V на скуп $\{1, \dots, n\}$, добили различиту матрицу повезаности. Ствар је договора да

⁵ Одељак се највећим делом ослања на ставку [3] литературе.

ли ћемо тако добијене графове сматрати истоветним или ћемо их разликовати. Уколико одлучимо да их сматрамо истоветним, посматрамо заправо одређену класу еквиваленције. Уведимо одговарајућу релацију:

$$G_1 = (V(G_1), E(G_1)) \sim G_2 = (V(G_2), E(G_2)) \Leftrightarrow_{(\text{деф})}$$

$$1) \quad V(G_1) = V(G_2) \quad (\text{као скупови})$$

$$2) \quad \text{постоји коначна пермутација } \pi: V(G_1) \rightarrow V(G_2) \text{ таква да:}$$

$$(\forall v_1, v_2 \in V(G_1)) \quad ((v_1, v_2) \in E(G_1) \Leftrightarrow (\pi v_1, \pi v_2) \in E(G_2))$$

Проверимо да је овако уведена релација заиста релација еквиваленције. Рефлексивност следи тривијално, док симетричност и транзитивност следе из чињенице да пермутације чине групу у односу на композицију функција.

Деф I.2.7. Међусобно еквивалентне графове, у односу на претходно уведену релацију еквиваленције, називамо изоморфним.

Деф I.2.8. Уколико представнике одређене класе еквиваленције, у односу на претходно уведену релацију еквиваленције, сматрамо истоветним, појединачна класа еквиваленције представља необележени граф. У супротном, говоримо о обележеном графу.

На основу претходног, релацију еквиваленције можемо увести на скупу матрица повезаности графова:

$$A \sim_m B \Leftrightarrow G_A \sim G_B, \text{ где су } G_A, \text{ односно } G_B, \text{ графови које одређују дате матрице.}$$

Како бисмо увели појмове који ће нам бити потребни касније, изложићемо сада неколико дефиниција које се односе на различите начине “кретања по графу”.

Деф I.2.9. Низ грана $e_1, \dots, e_{n-1} \in E$ графа G , за који постоји низ различитих чворова $a_1, \dots, a_n \in V$ таквих да, за $i = 1, \dots, n-1$ важи $\varphi(e_i) = \{a_i, a_{i+1}\}$ називамо стазом.

Деф I.2.10. Низ различитих грана $e_1, \dots, e_{n-1} \in E$ графа G , за који постоји низ чворова $a_1, \dots, a_n \in V$ таквих да, за $i = 1, \dots, n-1$ важи $\varphi(e_i) = \{a_i, a_{i+1}\}$ називамо трагом.

Уочимо да самим тим што стаза захтева различите чворове, она садржи различите гране. Са друге стране траг може “пролазити” истим чворовима, али се гране којима се иде не смеју понављати.

Важан појам, у контексту статистичке анализе случајних графова, јесте појам подграфа. Зато наводимо следећу дефиницију.

Деф I.2.11. Граф $G' = (V', E', \varphi')$ је подграф графа $G = (V, E, \varphi)$ ако $V' \subseteq V$, $E' \subseteq E$ и φ' је рестрикција φ на E' .

Обратићемо, сада, пажњу на то како подскуп скупа грана, односно подскуп скупа чворова, индукују подграф датог графа. Наиме, нека је $G = (V, E, \varphi)$, и $E' \subseteq E$. Ако V' дефинишемо као $V' = \bigcup_{x \in E'} \varphi(x)$ и φ' као одговарајућу рестрикцију функције φ , тада је добијени граф $G' = (V', E', \varphi')$ подграф графа $G = (V, E, \varphi)$ индукован скупом грана E' . Слично, ако је V' подскуп скупа V , тада E' дефинишемо са $E' = \{x \in E : \varphi(x) \subseteq V'\}$, а φ' као одговарајућу рестрикцију функције φ . Добијени подграф $G' = (V', E', \varphi')$ називамо подграфом индукованим скупом V' .

На основу претходних појмова и дискусије, уводимо следећу дефиницију:

Деф 1.2.12. Нека је $e_1, \dots, e_n$ траг графа $G = (V, E, \varphi)$ који се завршава у свом почетном чвору, односно нека је $a_1, \dots, a_n, a_1$ низ чворова који одређују траг. Подграф $G' = (V', E', \varphi')$ графа $G = (V, E, \varphi)$ индукован скупом грана $E' = \{e_1, \dots, e_n\}$ назива се циклом графа G . Уколико је једини чвор, који се понавља у низу чворова $a_1, \dots, a_n, a_1$ који одређују траг, чвор a_1 , онда подграф G' називамо тривијалним циклом графа G .

Деф 1.2.13. Граф G који не садржи ниједан цикл називамо ацикличним.

Даље, уколико желимо да одредимо доступност одређеног чвора, кретањем по гранама полазећи из неког другог чвора, потребан нам је појам повезаности који се формално уводи на следећи начин:

Деф 1.2.14. Ако за било која два различита чвора $u, v \in V$ графа $G = (V, E, \varphi)$ постоји стаза са почетком у u и крајем у v , онда је граф G повезан. Такође, ако је $|V| = 1$, тада је G повезан.

Захваљујући претходној дефиницији, могуће је увести следећу релацију на скупу чворова V одређеног графа:

$$u \sim v \Leftrightarrow u = v \text{ или постоји стаза која повезује } u \text{ и } v$$

Покажимо да је уведена релација, релација еквиваленције. Рефлексивност и симетричност следе тривијално. Како бисмо показали транзитивност, претпоставимо да је $u \sim v$ и $v \sim w$. Уколико w “лежи” на стази између u и v , тада постоји стаза између u и w , као део стазе између u и v . Ако w ипак “не лежи” на стази између u и v , поступићемо на следећи начин. Нека је $a_k = b_l$ први пресечан чвор стазе $a_1, \dots, a_n$ (између u и v) и $b_1, \dots, b_m$ (између v и w), када се иде од u ка v . Тада $a_1, \dots, a_k = b_l, \dots, b_m$ одређују стазу између u и w . Заиста, ако би за неко $b_j \in \{b_{l+1}, \dots, b_m\}$ постојало $a_i \in \{a_1, \dots, a_{k-1}\}$ тако да је $a_i = b_j$, тада $a_k = b_l$ не би био први пресечни чвор између стазе које повезују u и v , и v и w , супротно његовој дефиницији. Дакле, низом чворова $a_1, \dots, a_k = b_l, \dots, b_m$ одређена је стаза. Тривијално, она повезује u и w , одакле следи транзитивност релације.

Деф 1.2.15. Подграфике посматраног графа $G = (V, E, \varphi)$ индуковане класама еквиваленције претходно уведене релације називамо компонентама повезаности графа G .

Посебно дефинишемо повезане, ацикличне графике, односно:

Деф 1.2.16. Дрво је повезан, ацикличан граф. Ако је $|V| = 1$, тада је G дрво.

Од посебног значаја ће бити да утврдимо однос броја грана и чворова ацикличног графа. Због тога нам је важна наредна теорема.

Теорема I.2.1. Ако је G повезан граф, тада су следећа тврђења еквивалентна:

- 1) G је дрво.
- 2) G је ацикличан.
- 3) За сваки пар различитих чворова u и v постоји тачно једна стаза од u до v .
- 4) Уклањањем произвољне гране графа G добија се граф који није повезан.
- 5) Ако је $|V|$ број чворова графа G , и $|E|$ број грана тог графа, тада је $|V| = |E| + 1$.

Доказ: По дефиницији дрвета и услови теореме следи еквиваленција тврђења 1) и 2). Показаћемо сада $2) \Rightarrow 3) \Rightarrow 4) \Rightarrow 2)$. Покажимо прво да за сваки пар различитих чворова повезаног ацикличног графа постоји тачно једна стаза која их повезује. Претпоставимо супротно, да за неки пар чворова u и v постоје бар две стазе које их повезују. Крећући се од u до v једном стазом, а затим од v до u другом, пролазимо циклом, односно граф није ацикличан. Дакле, важи $2) \Rightarrow 3)$. Претпоставимо, сада, да важи 3). Уклонимо ли произвољну грану, тада ће чворови које је повезивала остати неповезани, јер између свака два чвора постоји само једна стаза која их повезује, односно следи $3 \Rightarrow 4)$. Како бисмо показали $4) \Rightarrow 2)$, претпоставимо да G садржи цикл, и покажимо да тада постоји грану коју можемо уклонити, тако да граф остане повезан. Заиста, нека је $a_1, \dots, a_n, a_1$ произвољан тривијални цикл. Уклонимо ли грану која повезује a_n и a_1 , ова два чвора остају повезана, а тиме и сви чворови који су били повезани “преко њихове гране”. Дакле, до сада смо показали $2) \Rightarrow 3) \Rightarrow 4) \Rightarrow 2)$ и $1) \Leftrightarrow 2)$, односно показана је еквиваленција прва четири тврђења.

Показаћемо, сада, математичком индукцијом по броју чворова графа G , импликацију $2) \Rightarrow 5)$. Покажимо, пре тога, да повезан ациклични граф има чвор од кога полази само једна грану. Претпоставимо супротно, да сваки чвор „лежи“ на најмање две гране. Одаберимо произвољни чвор v_1 за полазиште, и крећимо се по графу на следећи начин: произвољном граном из v_1 стижемо до неког чвора v_2 , а затим из v_2 крећемо неком другом граном, и стижемо до v_3 , и на исти начин настављамо кретање. Пошто је за овакво кретање потребно само имати више од једне гране из сваког чвора, у једном тренутку ћемо доспети у чвор у коме смо већ били, што је у контрадикцији са ацикличношћу графа. Дакле, за ациклични граф постоји чвор из кога полази само једна грану. Покажимо сада жељено тврђење. Ако G има један чвор, тада он нема грана па је једнакост задовољена. Претпоставимо, сада, да за сваки повезан ацикличан граф који има n чворова важи $|E| = n - 1$. Посматрајмо произвољан ациклични граф са $n + 1$ чворова. Тада постоји чвор из кога полази само једна грану. Уклонимо ли тај чвор и њему припадајућу грану, преостаће ацикличан граф са n чворова који, по индукцијској претпоставци има $n - 1$ грана. Како полазни граф има тачно један чвор и тачно једну грану више, то је број грана ацикличног графа са $n + 1$ чворова једнак n , односно тврђење је показано.

Покажимо, даље, $5) \Rightarrow 1)$. Претпоставимо да не важи 1), односно да G није дрво. Како $1) \Leftrightarrow 4)$, то можемо уклонити неку грану графа G тако да он остане повезан. Процес уклањања грана настављамо све док не добијемо дрво. Како је $1) \Leftrightarrow 2)$ и $2) \Rightarrow 5)$, то је број чворова дрвета за један већи од броја грана, а пошто

смо до тог дрвета дошли задржавањем свих чворова и уклањањем најмање једне гране, број чворова полазног графа је мањи од броја његових грана, односно не важи 5). Овим је показано $5) \Rightarrow 1)$, чиме је теорема коначно доказана.

Теорема I.2.2. Нека је c број компонената повезаности ацикличног графа, $|V|$ број његових чворова и $|E|$ број грана тог графа. Тада важи $|V| - c = |E|$.

Доказ: Свака компонента повезаности ацикличног графа је дрво, па задовољава $|V_i| - 1 = E_i$. Како је $|V| = V_1 + \dots + V_c$, као и $|E| = E_1 + \dots + E_c$, то је $|V| = V_1 + \dots + V_c = E_1 + \dots + E_c + c = |E| + c$, што је и требало доказати.

Последица I.2.2.1: Ако је G ацикличан граф, тада је $|E| + 1 \leq |V|$. Заиста, како је број компонената повезаности најмање један, то је $|E| + 1 \leq |E| + c = |V|$.

I.3. Концепт случајног графа и неке важније статистике случајних графова

Посматрајмо неки скуп графова. Задавањем вероватноће реализације сваког појединачног графа датог скупа, дефинисан је случајни граф. Другим речима, реализација појединачног графа, у складу са задатим законом вероватноће, представља случајни граф.

Навешћемо сада неке од статистика које квантификују својства случајних графова. Нагласимо да постоје многе статистике које се користе при проучавању случајних графова, и да ћемо овде објаснити само неколико њих. Избор статистика које се посматрају зависи од природе система који се посматра, односно особина система које је значајно утврдити.

Посматраћемо графове са n чворова, без петљи.

1. Степен чвора и просечан степен чворова⁶; Расподела степена чворова графа

Једноставно речено, степен чвора представља број грана које полазе из тог чвора. Односно:

Деф I.3.1. Нека је i чвор случајног графа и A одговарајућа матрица повезаности. Тада је са:

$$k_i = \sum_{j=1}^n A_{ij}$$

одређен степен чвора i .

Можемо дефинисати просечан степен чворова случајног графа:

Деф I.3.2. Нека је $\{1, \dots, n\}$ скуп чворова случајног графа, и k_i степен чвора i . Тада је:

$$k = \frac{1}{n} \sum_{i=1}^n k_i$$

просечан степен чворова посматраног случајног графа.

Очигледно, просечан степен чворова графа даје информацију о томе колико просечно грана полази из чвора тог графа.

Веома важна карактеристика случајног графа јесте расподела степена:

Деф I.3.3. Нека је p_k вероватноћа да је степен случајно изабраног чвора посматраног графа једнак k . Тада низ $\{p_k\}_{k=0}^{\infty}$ одређује расподелу степена графа.

2. Просечна дужина најкраће стазе између чворова⁷

⁶ Погледати у [10].

⁷ Погледати у [10].

Присетимо се појма стазе, уведеног у одељку 1.2. Нека је дужина стазе број грана које је одређују, и нека је стаза a краћа од стазе b ако је дужина стазе a мања од дужине стазе b . Посматрајмо повезани граф. Уводимо следећу дефиницију:

Деф 1.3.4. Дужина најкраће стазе која повезује чворове i и j датог графа назива се растојањем чворова i и j .

Аналогно се може посматрати дужина најдуже стазе.

Статистика која се често посматра јесте просечна дужина стазе између чворова случајног графа. Она је дефинисана у складу са:

Деф 1.3.5. Просечна дужина најкраће стазе између чворова повезаног случајног графа G одређена је са:

$$\bar{l} = \frac{1}{n(n-1)} \sum_{i \neq j} l(i, j)$$

где је са $l(i, j)$ означено растојање између i и j .

Опет очигледно, просечна дужина најкраће стазе даје информацију о средњем броју грана које очекујемо да морамо прећи да бисмо дошли из неког, произвољног полазног чвора, у произвољни одредишни чвор.

3. Коефицијент груписаности⁸

Посматрајмо чвор i графа G , и чворове који су са њим повезани. Нека је $z^{(i)}$ број чворова повезаних са чвором i . Концентрисаћемо се на евентуалне гране између тих чворова- суседа чвора i . Уочимо да је максимални број грана међу тим чворовима $\frac{z^{(i)}(z^{(i)}-1)}{2}$. Уводимо следећу дефиницију:

Деф 1.3.6. Коефицијент груписаности чвора i одређен је са:

$$C^{(i)} = \frac{y^{(i)}}{\frac{z^{(i)}(z^{(i)}-1)}{2}}$$

где је $z^{(i)}$ број чворова повезаних са чвором i , а $y^{(i)}$ број грана међу тим чворовима.

Уочимо да је $0 \leq C^{(i)} \leq 1$. Истакнимо да мотивација за посматрање ове статистике лежи у, за неке мреже, интуитивно јасној претпоставци. Претпоставимо да знамо да двоје корисника *Facebook*-а имају заједничког *Facebook*-пријатеља. Тада је очекивање да су и они у релацији *Facebook*-пријатељства веће него у случају супротне претпоставке.

Коефицијент груписања целог случајног графа одређен је са:

⁸ Погледати у [10].

Деф 1.3.7. Нека је $C^{(i)}$ коефицијент груписања чвора i случајног графа G . Коефицијент груписања графа G одређен је са:

$$C = \frac{1}{n} \sum_{i=1}^n C^{(i)}$$

Просечан степен чвора, расподела степена графа, просечна дужина најкраће стазе између чворова и коефицијент груписаности су неке од најчешће коришћених статистика случајних графова.

4. Број троуглова и 2-звезда као подграфа датог графа

Троуглом називамо подграф посматраног графа G који је одређен са три чвора и свим могућим гранама између њих (три гране су могуће). Подграф графа G одређен са три његова чвора, нпр. i, j , и k , који сигурно има гране одређене паровима (i, j) и (i, k) називамо 2-звездом са центром у i . Уочимо да су троуглови, у смислу ове дефиниције, подскуп скупа 2-звезда са центром у једном од његових темена.

Број троуглова у графу G , који садрже чвор i одређен је са:

$$\sum_{\substack{j \in V \\ j \neq i}} \sum_{\substack{k \in V \\ k \neq i, k \neq j}} A_{ij} A_{ik} A_{jk}$$

Број 2-звезда са центром у i је:

$$\sum_{\substack{j \in V \\ j \neq i}} \sum_{\substack{k \in V \\ k \neq i, k \neq j}} A_{ij} A_{ik}$$

Укупан број троуглова у графу је:

$$\frac{1}{3!} \sum_{i, j, k \in V} A_{ij} A_{ik} A_{jk}$$

док је број 2-звезда:

$$\frac{1}{2!} \sum_{i, j, k \in V} A_{ij} A_{ik}$$

Степен присутности троуглова и 2-звезда даје информацију о присутности груписања у посматраном графу. Наиме, ако посматрамо број троуглова који садрже дати чвор, тада заправо знамо колико је суседа посматраног чвора међусобно повезано. Дакле, коефицијент груписања чвора i одређен је бројем троуглова и 2-звезда које га садрже, односно:

$$C^{(i)} = \frac{\sum_{j \in V, j \neq i} \sum_{k \in V, k \neq i, k \neq j} A_{ij} A_{ik} A_{jk}}{\sum_{j \in V, j \neq i} \sum_{k \in V, k \neq i, k \neq j} A_{ij} A_{ik}}$$

На основу присутности броја чворова и 2-звезда у графу може се увести алтернативна дефиниција коефицијента груписања графа:

Деф I.3.8. Коефицијент груписања графа дефинисан је са⁹:

$$C^* = \frac{\sum_{i,j,k} A_{ij}A_{ik}A_{jk}}{\sum_{i,j,k} A_{ij}A_{ik}}$$

Нагласимо да дефиниције *Деф I.3.7.* и *Деф I.3.8.* нису еквивалентне.

Број троуглова и 2-звезда у графу навели смо као пример статистика које мере присутност неких подструктура посматраног графа. Као што се може уочити на примеру њихове повезаности са присутношћу груписања у графу, учесталост неке структуре може дати корисну информацију о некој особини графа. У наредним поглављима бавићемо се, између осталог, оваквим статистикама.

⁹ Погледати у [24].

II Erdős-Rényi-ев модел случајног графа и нека асимптотска својства¹⁰

Историјски посматрано, Erdős-Rényi-ев модел представља први описани модел случајног графа. У овом одељку приказаћемо нека од његових асимптотских својстава. Концентрисаћемо се на она својства која се односе на постојање одређених подструктура у случајном графу. Управо таквим својствима бавићемо се касније, у непараметарском контексту.

Посматрајмо скуп графова разапетих на (означеним) чворовима $V_1, \dots, V_n$ са N грана, у ознаци $E_{n,N}$. Како на датих n чворова, N грана можемо одабрати на $\binom{n}{2}$ начина, то скуп $E_{n,N}$ садржи $\binom{n}{2}$ елемената.

Деф II.1. Случајни граф $\Gamma_{n,N}$ Erdős-Rényi модела је случајна величина са простором исхода $E_{n,N}$ која сваки од могућих исхода узима са једнаком вероватноћом $1/\binom{n}{2}$.

Постоји алтернативна дефиниција Erdős-Rényi модела, која је посебно погодна посматрано са асимптотског аспекта, те ћемо је стога приказати. Она третира граф са N грана као реализацију случајног процеса након N временских тренутака, односно као резултат процеса еволуције који траје N временских тренутака. Наиме, у почетном тренутку имамо n неповезаних чворова. У тренутку 1, успостављамо једну грану, и то случајним избором од $\binom{n}{2}$ могућих грана са једнаким вероватноћама $1/\binom{n}{2}$. У сваком следећем тренутку, закључно са тренутком N , бирамо, са једнаким вероватноћама, једну од преосталих могућих грана. Односно, у тренутку $k \leq N$ бирамо, са вероватноћом $1/\left(\binom{n}{2} - (k-1)\right)$ једну од могућих $\left(\binom{n}{2} - (k-1)\right)$ грана. Случајни граф $\Gamma_{n,N}$ можемо дефинисати као претходно описани случајни процес у тренутку N .

Формално уведена дефиниција и описано поимање случајног графа из перспективе еволуције, су еквивалентни. Ово једноставно следи из чињенице да је вероватноћа реализације конкретног графа, као резултата описаног случајног процеса, једнака $\frac{1}{\binom{n}{2}} \cdot \frac{1}{\binom{n}{2}-1} \cdot \dots \cdot \frac{1}{\binom{n}{2}-(N-1)} \cdot N!$ при чему је множење са $N!$ последица чињенице да није битно у ком је тренутку одређена грана настала, већ је важна само одговарајућа комбинација грана.

Посматрамо ли описани процес еволуције у току $\binom{n}{2}$ временских тренутака, можемо тврдити да ће његов резултат бити граф чија су свака два чвора повезана граном. У средишту нашег интересовања јесу особине случајног графа када број чворова тежи бесконачности, односно $n \rightarrow \infty$. Уочимо да, у том случају, еволутивни процес, који траје $\binom{n}{2}$ временских тренутака, пролази кроз различите фазе у смислу асимптотског односа броја чворова n и броја грана $N(n)$. Сваку од тих фаза одликују специфичности у

¹⁰ Одељак припремљен на основу ставке [12] литературе.

погледу подструктура које случајни граф садржи. Управо те одлике приказаћемо сада, без доказа, за неке од фаза еволуције.

$$1) N(n) = o(n)$$

У овој фази $\Gamma_{n,N}$ се, скоро сигурно, састоји искључиво из компонената које су дрвета. При томе, уколико $N(n)$ не досеже ред величине $\frac{k-2}{n^{k-1}}$, односно ако је $\frac{N(n)}{\frac{k-2}{n^{k-1}}} \rightarrow 0$, при $n \rightarrow \infty$, тада $\Gamma_{n,N}$, са вероватноћом 1, не садржи дрвета реда k . Ако је $N(n) \sim \rho \cdot \frac{k-2}{n^{k-1}}$ за неко $\rho > 0$, тада број дрвета реда k у $\Gamma_{n,N}$ има, при $n \rightarrow \infty$, Пуасонову расподелу са очекивањем једнаким $\frac{(2\rho)^{k-1} k^{k-2}}{k!}$. Ако је, пак, $\frac{N(n)}{\frac{k-2}{n^{k-1}}} \rightarrow \infty$ тада број дрвета реда k у $\Gamma_{n,N}$ има, при $n \rightarrow \infty$, нормалну $\mathcal{N}(M_n, M_n)$ расподелу где је $M_n = n \cdot \frac{k^{k-2}}{k!} \cdot \left(\frac{2N(n)}{n}\right)^{k-1} \cdot e^{\frac{-2kN(n)}{n}}$.

$$2) N(n) \sim c \cdot n \text{ за } 0 < c < \frac{1}{2}$$

Док се у првој фази $\Gamma_{n,N}$ састојао скоро сигурно, од дрвета, односно ацикличних структура, у овој фази расподела тривијалних циклора реда k тежи, при $n \rightarrow \infty$, Пуасоновој расподели са очекивањем $\frac{(2c)^k}{2k}$. У овој фази, скоро све компоненте графа су дрвета, или имају једнак број чворова и грана (уочимо да тривијални цикл има једнак број чворова и грана). Расподела броја компонената са k чворова и k грана асимптотски тежи Пуасоновој расподели са очекивањем $\frac{(2ce^{-2c})^k}{k!} \left(1 + k + \frac{k^2}{2!} + \dots + \frac{k^{k-3}}{(k-3)!}\right)$. Највећа компонента графа је дрво и има, са вероватноћом која, при $n \rightarrow \infty$, тежи 1, приближно $\frac{1}{\alpha} \left(\log n - \frac{5}{2} \log \log n\right)$ чворова за $\alpha = 2c - 1 - \log 2c$.

$$3) N(n) \sim \frac{1}{2} \cdot n$$

Овај случај је детаљно проучен у ставци [6] литературе. За $N = \frac{1}{2} \cdot n + s$ где је $s = o(n)$, и $s \geq (\log n)^{1/2} n^{2/3}$, граф се састоји из једне ‘велике’ и осталих ‘малих’ компоненти, при чему велика компонента има најмање $n^{2/3}$ чворова, док свака од осталих компоненти има мање од $n^{2/3}/2$ чворова.

$$4) N(n) \sim c \cdot n \text{ за } c > \frac{1}{2}$$

У овој фази највећа компонента има скоро сигурно $G(c) \cdot n$ чворова где је $G(c) = 1 - \frac{1}{2c} \sum_{k=1}^{\infty} \frac{k^{k-1}}{k!} (2ce^{-2c})^k$. Већину преосталих компонената чине дрвета, а укупан број чворова који припадају дрветима је скоро сигурно једнак $n(1 - G(c)) + o(n)$.

5) $N(n) \sim c \cdot n \log n$ за $c \leq \frac{1}{2}$

У овој фази $\Gamma_{n,N}$ постаје скоро сигурно повезан.

6) $N(n) \sim (n \log n) \omega(n)$ где $\omega(n) \rightarrow \infty$

У овој фази граф је скоро сигурно повезан, и скоро сигурно из сваког чвора полази једнак број грана.

Као посебно интересантан уочавамо прелаз из фазе 2) у фазу 3). Наиме, док при $N(n) \sim c \cdot n$ за $0 < c < \frac{1}{2}$ највећа компонента случајног графа има приближно $\frac{1}{\alpha} \left(\log n - \frac{5}{2} \log \log n \right)$ чворова за $\alpha = 2c - 1 - \log 2c$, када је $N(n) \sim \frac{1}{2} \cdot n$ највећа компонента има најмање $n^{2/3}$ чворова, односно број чворова највеће компоненте се изненађујуће значајно мења. Изненадне промене особина случајних графова у асимптотском контексту, налик овој, нису ретке- вероватноће појаве дрвета реда k , тривијалног циклуса реда k , као и подграфа реда k чији су сви чворови међусобно повезани, изненада се мењају са 0 на 1, када се пређе одговарајући ступањ еволуције¹¹.

¹¹ Погледати у [8].

III Неки аспекти непараметарског приступа моделу случајног графа

III.1. Параметарски и непараметарски приступ моделу случајног графа

Параметарски приступ моделирању мрежа подразумева претпоставку да је мрежа коју испитујемо реализација случајног графа из неке задате расподеле. Како би се донели закључци о неким особинама посматране мреже, потребно је проценити параметре претпостављене расподеле. Мада постоји много модела случајних графова, веома често се дешава да за мрежу коју треба моделирати ниједан од модела не представља адекватан избор. Неадекватност избора модела огледа се у постојању важних особина мреже које претпостављени модел не описује довољно добро. Ово представља озбиљну тешкоћу у моделирању мрежа.

Ради превазилажења ове врсте проблема, у неким случајевима примењује се непараметарски приступ моделирању. Наиме, под одређеним претпоставкама које су мање ригорозне од претпоставке конкретног модела, утврђују се одређене особине случајног графа, односно мреже која се моделира. Погодност овог приступа лежи управо у чињеници да изведени закључци зависе од испуњености мање ригорозних претпоставки.

У одељку *II* посматрали смо асимптотска својства Erdős-Rényi-евог модела. У одељку *III.3.* биће приказан један модел у непараметарском контексту и нека његова асимптотска својства.

III.2. Заменљивост случајних величина

У овом одељку бавићемо се концептом заменљивости случајних величина.

Деф III.2.1. Нека је $\{X_i\} i \geq 1$ низ случајних величина. Оне су заменљиве ако¹²:

$$p\{X_1 \leq x_1, \dots, X_n \leq x_n\} = p\{X_1 \leq x_{\pi_1}, \dots, X_n \leq x_{\pi n}\}$$

за свако n , и све пермутације π скупа $\{x_1, \dots, x_n\}$, односно ако:

$$(X_i, i \geq 1) =_d (X_{\pi i}, i \geq 1) \text{ за све коначне пермутације } \pi.$$

Уочимо да низ независних случајних величина са истом расподелом има особину заменљивости. Нагласимо, међутим, да обрнуто не важи, односно да је могућа међусобна зависност заменљивог низа случајних величина.

Формулишимо, сада, услов заменљивости за низ индексиран скупом $N_2 = \{\{i, j\}, i, j \in N\}$.

Деф III.2.2. Низ $X_{\{i, j\}}, \{i, j\} \in N_2$ је заменљив ако:

$$(X_{\{i, j\}}, \{i, j\} \in N_2) =_d (X_{\pi\{i, j\}}) \text{ за све коначне пермутације } \pi \text{ скупа } N_2.$$

За низове индексиране скупом N_2 , увешћемо још један појам, појам делимичне заменљивости.

Деф III.2.3. Низ $X_{\{i, j\}}, \{i, j\} \in N_2$ је делимично заменљив ако¹³:

$$(X_{\{i, j\}}, \{i, j\} \in N_2) =_d (X_{\{\pi i, \pi j\}}) \text{ за све коначне пермутације } \pi \text{ скупа } N.$$

Приметимо да је услов делимичне заменљивости слабији од услова заменљивости, јер су функције $\{i, j\} \rightarrow \{\pi(i), \pi(j)\}$ само неке од пермутација скупа N_2 .

Циљ ка коме ћемо се усмерити јесте погодна репрезентација делимично заменљивих низова.

Посматрајмо мерљиву функцију¹⁴ $f: [0,1]^4 \rightarrow R$ симетричну по другом и трећем аргументу, односно: $f(u, u_1, u_2, u_{12}) = f(u, u_2, u_1, u_{12}) \quad \forall u, u_1, u_2, u_{12} \in [0,1]$. Дефинишимо, случајни низ $X_{\{i, j\}}, \{i, j\} \in N_2$ на следећи начин:

$$X_{\{i, j\}} = f(U, U_i, U_j, U_{\{i, j\}}) \quad (1)$$

где су $U, U_i, U_{\{i, j\}}$ за $i \geq 1, \{i, j\} \in N_2$ међусобно независне случајне величине са $U[0,1]$ расподелом.

¹² Погледати у [2].

¹³ Погледати у [2].

¹⁴ Преузето из [2].

Покажимо да је, овако конструисан низ $X_{\{i,j\}}$ делимично заменљив. Како је:

$$X_{\{i,j\}} = f(U, U_i, U_j, U_{\{i,j\}}) =_d f(U, U_{\pi i}, U_{\pi j}, U_{\{\pi i, \pi j\}}) = X_{\{\pi i, \pi j\}}$$

слиеди да је низ делимично заменљив.

Приметимо, међутим, да конструисани низ није заменљив. Заиста, нека је $\pi: N_2 \rightarrow N_2$ коначна пермутација таква да је $\pi\{i,j\} = \{\mu(i), \nu(j)\}$ где су μ и ν функције $\mu, \nu: N \rightarrow N$, и нека је $\mu(i) = \nu(j)$. Тада $X_{\pi\{i,j\}} = f(U, U_{\mu(i)}, U_{\nu(j)}, U_{\pi\{i,j\}})$, где је $U_{\mu(i)} \equiv U_{\nu(j)}$, није у расподели једнако са $X_{\{i,j\}} = f(U, U_i, U_j, U_{\{i,j\}})$ где су U_i и U_j међусобно независне.

Управо смо утврдили да низови настали конструкцијом (1) имају особину делимичне заменљивости. Могуће је показати и обратно, да сваки делимично заменљив низ има репрезентацију (1). Наиме, важи:

Теорема III.2.1. Делимично заменљив низ $X_{\{i,j\}}$ за $\{i,j\} \in N_2$, могуће је приказати у облику: $X_{\{i,j\}} = f(U, U_i, U_j, U_{\{i,j\}})$ где је $f: [0,1]^4 \rightarrow R$ мерљива функција симетрична по другом и трећем аргументу, и $U, U_i, U_{\{i,j\}}$ за $i \geq 1, \{i,j\} \in N_2$ међусобно независне случајне величине са $U[0,1]$ расподелом¹⁵.

Како бисмо успоставили теоријске основе које ће нам бити потребне у поставци проблема којим ћемо се бавити, посматрајмо случајне низове U_i, V_j $1 \leq i, j < \infty$ независних случајних величина са $U[0,1]$ расподелом. Нека је $W: [0,1]^2 \rightarrow [0,1]$, и нека случајна променљива X_{ij} узима вредности 0 или 1 у зависности од $W(U_i, V_j)$, односно нека W одређује расподелу од $\{X_{ij}\}$, у ознаци P_W . Приметимо да је, по конструкцији¹⁶, низ $\{X_{ij}\}$ делимично заменљив. Постоји, међутим, теорема¹⁷ која показује да је сваки делимично заменљив бинарни низ могуће приказати преко расподела одређених овако конструисаним функцијама W , односно:

Теорема III.2.2. (Aldous-Hoover-ова теорема) Нека је $X = \{X_{ij}\}$ $1 \leq i, j < \infty$ делимично заменљив бинарни случајни низ. Тада постоји вероватноћа μ тако да, за сваки скуп $A \in \mathcal{A}$:

$$p(X \in A) = \int P_W(A) \mu(dW)$$

Показује се, даље, да, уколико се W^* добија из W као $W^* = W(\varphi(U), \varphi(V))$, где је φ функција $\varphi: [0,1] \rightarrow [0,1]$ која чува меру, тада процес X_{ij}^* одређен са W^* има исту расподелу као X ¹⁸. Ово слиеди из чињенице да, уколико $U, V: U[0,1]$ и φ чува меру, тада $\varphi(U), \varphi(V): U[0,1]$.

¹⁵ Доказ теореме може се наћи у [16].

¹⁶ Конструкција преузета из [9].

¹⁷ Погледати у [9].

¹⁸ Погледати у [9].

Интересантно је, међутим, да важи и обрнуто: уколико две функције W и W^* одређују исти процес, тада постоји функција φ која чува меру, тако да је $W^*(u, v) = W(\varphi(u), \varphi(v))$. Овај важан резултат добијен је помоћу теорије модела¹⁹.

На основу теорије која се односи на заменљивост случајних величина, дошли смо до репрезентације делимично заменљивог бинарног низа погодне за разматрање случаја који ћемо посматрати, па се даље нећемо бавити овом, резултатима богатом, теоријом.

¹⁹ Погледати у [9].

III.3. Поставка модела²⁰

Посматрајмо неусмерене графове, без петљи, при чему је између свака два чвора могућа највише једна грана. Сматрајмо, при том, да су темена графова необележена. Имајући у виду да се особина необележености односи на посматрање класа еквиваленције у односу на одговарајућу релацију, као што је раније показано, расподела вероватноћа задата је на одговарајућим класама еквиваленције графова, односно матрица. Претпостављамо, дакле, да је дата расподела вероватноћа P на класама еквиваленције матрица повезаности.

Наш циљ је да дођемо до закључака о асимптотском понашању статистика које ћемо посматрати. У асимптотском контексту, при $n \rightarrow \infty$, за посматрану класу графова испуњен је услов делимичне заменљивости. Наиме, како је расподела вероватноћа задата на класама еквиваленције графова (еквивалентно, матрица), следи да важи делимична заменљивост. Због тога, у складу са теоријом приказаном у претходном одељку, матрица графа има следећу репрезентацију:

$$A_{ij} = g(\alpha, \xi_i, \xi_j, \lambda_{ij})$$

где је g мерљива функција $g: [0,1]^4 \rightarrow R$ симетрична по другом и трећем аргументу, а $\xi_i, \xi_j, \lambda_{ij}(= \lambda_{ji})$ међусобно независне случајне величине из $U[0,1]$ расподеле. Нагласимо да није могуће тврдити јединственост функције g .

Уведимо следећу функцију:

$$h(u, v) = p\{A_{ij} = 1 | \xi_i = u, \xi_j = v\} \quad u, v \in [0,1] \quad (1)$$

односно, функција $h(u, v)$ је таква да важи:

$$p\{A_{ij} = 1\} = \int_0^1 \int_0^1 h(u, v) du dv \quad (2)$$

Из (2) примећујемо да h одређује $p\{A_{ij} = 1\}$, односно задаје расподелу на простору графова. Међутим, одговор на питање да ли датој расподели одговара јединствена функција h , тако да важи (2) је негативан. Наиме, у складу са претходно изложеним, свака функција $h_1 = h(\varphi(u), \varphi(v))$, где φ чува меру, одређује исту расподелу на простору графова. И више, за сваку другу функцију h_2 која одређује исту расподелу, постоји функција φ_2 која чува меру, тако да је $h_2 = h(\varphi_2(u), \varphi_2(v))$.

Показаћемо, међутим, следећу лему:

Лема III.3.1. У скупу функција h које одређују дату расподелу P на посматраном простору графова постоји јединствена функција $h_{CAN}(u, v)$ са особином да је функција $\psi(u) = \int_0^1 h_{CAN}(u, v) dv$ неоппадајућа по u .

²⁰ Одељак се ослања на ставку [5] литературе.

*Доказ:*²¹ Посматрајмо произвољну функцију $h_1(u, v)$ која одређује дату расподелу на посматраном простору графова. Фиксирајмо v . Дефинишимо функцију:

$$M_v(y) = \mu\{u | h_1(u, v) \leq y\} \quad y \in [0, 1]$$

где је μ Лебегова мера. Из дефиниције следи да је ово неоппадајућа функција.

Дефинишимо, даље:

$$h_2(M_v(y), v) = y$$

Уочимо да ова дефиниција није коректна ако за $y_1 \neq y_2$ важи $M_v(y_1) = M_v(y_2)$. Ипак, ово је могуће пребројиво много пута, па функција h_2 (евентуално) није дефинисана само на скупу мере нула.

Ако је $M_v(y_1) \leq M_v(y_2)$, онда је $\mu\{u | h_1(u, v) \leq y_1\} \leq \mu\{u | h_1(u, v) \leq y_2\}$, па важи $\{u | h_1(u, v) \leq y_1\} \subseteq \{u | h_1(u, v) \leq y_2\}$, односно $y_1 \leq y_2$. Дакле, функција $h_2(M_v(y), v)$ је неоппадајућа по $M_v(y)$.

Нека је $A = \{u | h_1(u, v) \leq y\}$. Тада, због дефиниција A и $M_v(y)$, важи $\mu(A) = M_v(y)$. Фиксирајмо $y \in [0, 1]$. Нека је $B = \{u | h_2(u, v) \leq y\}$, и нека је $B_1 = \{u : \exists y_1 \text{ тако да } M_v(y_1) = u \text{ и } y_1 \leq y\}$. Важи $h_2(u, v) = y_1 \Rightarrow M_v(y_1) = u$, по дефиницији h_2 . Дакле, $B \subseteq B_1$. Обрнута импликација не важи једино ако је $M_v(y_1)$ тачка на којој h_2 није дефинисано, па је $\mu(B_1 \setminus B) = 0$. Како је $M_v(y)$ неоппадајућа функција, то је $\mu(B_1) = M_v(y)$.

Дакле, $\mu\{u | h_1(u, v) \leq y\} = \mu(A) = \mu(B) = \mu\{u | h_2(u, v) \leq y\}$ за свако $y \in [0, 1]$, па важи:

$$\forall v \int_0^1 h_1(u, v) du = \int_0^1 h_2(u, v) du$$

Одавде је:

$$\int_0^1 \int_0^1 h_1(u, v) dudv = \int_0^1 \int_0^1 h_2(u, v) dudv$$

односно h_1 и h_2 одређују исту расподелу, одакле, по раније изложеном, следи да постоји функција φ која чува меру, тако да је $h_2(u, v) = h_1(\varphi(u), \varphi(v))$. Дефинишимо:

$$\psi_2(u) = \int_0^1 h_2(u, v) dv$$

Како је функција h_2 неоппадајућа по првој променљивој, то, за $u_1 \leq u_2$ важи $\psi_2(u_1) = \int_0^1 h_2(u_1, v) dv \leq \int_0^1 h_2(u_2, v) dv = \psi_2(u_2)$, односно $\psi_2(u)$ је неоппадајућа функција.

²¹ На основу конструкције из одељка 10 ставке [15] литературе.

Како h_2 задовољава услове које смо желели, дефинишемо $h_{CAN}(u, v) = h_2(u, v)$. Докажимо да је, овако дефинисана h_{CAN} јединствена (до на скуп мере нула). У том циљу, уведемо функцију:

$$F(x) = p\{\psi_{CAN}(\xi) < x\} = p\{\xi \leq \inf\{v | \psi_{CAN}(v) \geq x\}\} = \inf\{v | \psi_{CAN}(v) \geq x\} = \psi_{CAN}^{-1}(x) \quad (3)$$

где је $\psi_{CAN}^{-1}(x)$ дефинисано као $\psi_{CAN}^{-1}(x) = \inf\{v | \psi_{CAN}(v) \geq x\}$. Трећа у низу једнакости (3) важи јер $\xi: U[0,1]$.

Дакле, следи $F(x) = \psi_{CAN}^{-1}(x)$.

Уочимо да, због једнозначне кореспонденције између вероватноћа на реалној правој и расподела, датој расподели вероватноћа P на простору графова одговара тачно једна функција $F(x)$. Претпоставимо, сада, постојање функције $\psi(u) = \int_0^1 h(u, v) dv$ (где $h(u, v)$ одређује посматрану расподелу P), која је монотонно неоппадајућа. Тада је, због јединствености $F(x)$, $\psi(u) = F^{-1}(u)$, односно $\psi(u) = \psi_{CAN}(u)$. Овим је лема доказана.

Надаље ћемо подразумевати $h(u, v) = h_{CAN}(u, v)$.

Нека је $\rho = \int_0^1 \int_0^1 h(u, v) du dv$ вероватноћа гране у графу. Тада је условна густина $(\xi_i, \xi_j) | A_{ij} = 1$ одређена са:

$$\omega(u, v) = \rho^{-1} h(u, v)$$

Као природна претпоставка намеће се $\rho = \rho_n$, односно вероватноћа гране зависи од броја чворова у графу. Такође, претпоставићемо $\rho_n \rightarrow 0$ при $n \rightarrow \infty$, односно да се вероватноћа појединачне гране смањује повећањем броја чворова у графу. Уочимо да, због $p\{A_{ij} = 1 | \xi_i = u, \xi_j = v\} = h_n(u, v) = \rho_n \omega(u, v)$, морамо, за свако n , обезбедити $\rho_n \omega(u, v) < 1$. Претпостављаћемо, дакле, да је, за свако n испуњено $\rho_n \omega(u, v) < 1$, и игнорисати зависност n и ω_n , односно сматраћемо $\omega_n = \omega$.

У циљу увођења одговарајуће статистике за процену вероватноће гране графа ρ_n , посматрајмо статистику

$$\bar{D} = \frac{2L}{n}$$

где L представља укупан број грана у датом графу G_n , а $\bar{D}$ просечан број грана које полазе из чвора. Тада се, за оцену вероватноће гране $\hat{\rho}$ узима $\hat{\rho} = \frac{\bar{D}}{n-1}$, јер из једног чвора полази просечно $\bar{D}$ грана, при чему се гране могу остварити са неким од преостала $n - 1$ чвора. Како је $E(L) = \rho \cdot \binom{n}{2}$ то је $E(\hat{\rho}) = \rho$.

Увешћемо, сада, статистику која мери присутност одређеног подграфа, у графу.

Нека је R подскуп скупа $\{(i, j): 1 \leq i < j \leq n\}$, где $1..n$ одређују чворове графа G_n . Уочимо да R одређује подскуп скупа свих могућих грана на чворовима графа G_n , међутим он не мора бити подскуп скупа грана графа G_n . Нека је $V(R) = \{i: (i, j) \text{ или } (j, i) \in R \text{ за неко } j\}$ и $E(R) = R$, односно задајмо граф одређен уоченим подскупом. Надаље ће се R односити управо на тај граф. Нагласимо, дакле, да је R

граф разапет на подскупу скупа чворова графа G_n , али да гране ова два графа нису нужно ни у каквој релацији. Нека је, даље, $G_n(R)$ подграф графа G_n , индукован скупом чворова $V(R)$, односно $G_n(R) = (V(R), E(G_n(R)))$. Вероватноћу појаве подграфа R у случајном графу G_n означимо са:

$$P(R) = p \left(E(G_n(R)) = E(R) \right)$$

Како је $P(R) = E(P(R)|\xi)$, а пошто:

$$P(R)|\xi = \prod_{(i,j) \in R} h(\xi_i, \xi_j) \prod_{(i,j) \in \bar{R}} (1 - h(\xi_i, \xi_j))$$

где је $\bar{R} = \{(i,j) \notin R, i \in V(R), j \in V(R)\}$, то је:

$$P(R) = E \left(\prod_{(i,j) \in R} h(\xi_i, \xi_j) \prod_{(i,j) \in \bar{R}} (1 - h(\xi_i, \xi_j)) \right)$$

За оцену $P(R)$ узећемо:

$$\hat{P}(R) = \frac{1}{\binom{n}{p} N(R)} \sum \{I(G \sim R) : G \subseteq G_n\}$$

где је $N(R)$ број графова изоморфних са R који могу бити разапети на p чворова. Уочимо да $\binom{n}{p} \cdot N(R)$ представља укупан број графова изоморфних са R који могу бити разапети на чворовима графа G_n . Како $\sum \{I(G \sim R) : G \subseteq G_n\}$ одређује укупан број подграфова од G_n изоморфних са R , то је $\hat{P}(R)$ непристрасна оцена за $P(R)$.

Како је, при $n \rightarrow \infty$, $P(R) \rightarrow 0$, за било који граф R разапет на фиксираним броју чворова p , то је, при $n \rightarrow \infty$, потребно скалирање статистике $P(R)$ како би она била ваљана карактеристика графа. Зато уводимо:

$$\tilde{P}(R) = \rho_n^{-|E(R)|} P(R)$$

За процену $\tilde{P}(R)$ узећемо $\check{P}(R) = \hat{\rho}_n^{-|E(R)|} \hat{P}(R)$.

Уочимо да је очекивани степен чвора графа G_n , у ознаци λ_n , једнак $\lambda_n = (n-1)\rho_n$, и да је, при $n \rightarrow \infty$, могуће $\lambda_n \rightarrow 0$, $\lambda_n \rightarrow c$ за неко $c \in R$, и $\lambda_n \rightarrow \infty$.

Резултат који ће бити приказан односи се на асимптотско понашање статистика $\hat{\rho}$ и $\check{P}(R)$ при $n \rightarrow \infty$, у случају $\lambda_n \rightarrow \infty$, односно када, при повећању броја чворова, број грана расте без ограничења. Он је утврђен следећом теоремом:

Теорема III.3.1. Нека је $\int_0^1 \int_0^1 \omega^2(u, v) dudv < \infty$, и нека $\lambda_n \rightarrow \infty$. Тада:

$$\frac{\hat{\rho}_n}{\rho_n} \rightarrow_p 1 \quad (1)$$

$$\sqrt{n} \left(\frac{\hat{\rho}_n}{\rho_n} - 1 \right) \rightarrow_{\mathcal{D}} \mathcal{N}(0, \sigma^2) \quad (2)$$

за неко $\sigma^2 > 0$. Даље, нека је R ацикличан граф и нека је $|V(R)| = p$, $|E(R)| = q$ као и $\int_0^1 \int_0^1 \omega^{2q}(u, v) dudv < \infty$. Тада важи:

$$\check{P}(R) \rightarrow_p \tilde{P}(R) \quad (3)$$

$$\sqrt{n}(\check{P}(R) - \tilde{P}(R)) \rightarrow_{\mathcal{D}} \mathcal{N}(0, \sigma^2(R)) \quad (4)$$

III.4. Нека асимптотска својства U -статистика²²

Посматрајмо фамилију вероватноћа $\mathbf{P}$ на $(\Omega, \mathcal{A})$. Приметимо да елементи фамилије $\mathbf{P}$ не припадају ни једном одређеном типу расподеле, па је зато теорија која ће бити приказана основа непараметарског приступа статистичкој анализи.

Деф III.4.1. Кажемо да параметар $\theta(P)$ можемо проценити у $\mathbf{P}$ ако за неки цео број m постоји реална мерљива функција h од m аргумената таква да је:

$$E_P(h(X_1, \dots, X_m)) = \theta(P) \text{ за све } P \in \mathbf{P}$$

где су $X_1, \dots, X_m$ независне случајне променљиве из расподеле P .

Приметимо да је $h(X_1, \dots, X_m)$ заправо непристрасна оцена параметра $\theta(P)$ за сваку од расподела из дате фамилије.

Приметимо да, уколико имамо функцију $h(X_1, \dots, X_m)$ са особином из претходне дефиниције, тада можемо конструисати функцију симетричну у односу на произвољну пермутацију њених елемената, односно функцију $g(X_1, \dots, X_m)$ такву да важи $g(X_1, \dots, X_m) = g(X_{\sigma(1)}, \dots, X_{\sigma(m)})$ за произвољну пермутацију σ скупа $\{X_1, \dots, X_m\}$, односно $\sigma \in \Pi_m$ где је Π_m група пермутација скупа $\{X_1, \dots, X_m\}$. Заиста, нека је $g(X_1, \dots, X_m) = \frac{1}{m!} \sum_{\sigma \in \Pi_m} h(X_{\sigma(1)}, \dots, X_{\sigma(m)})$. Лако је проверити да је g симетрична, као и да је $E_P(g(X_1, \dots, X_m)) = \theta(P)$ за свако $P \in \mathbf{P}$. Зато, уколико нам је потребно, можемо претпоставити симетричност функције h .

Деф III.4.2. За реалну, мерљиву функцију $h(x_1, \dots, x_m)$ и узорак $X_1, \dots, X_n$ за $n \geq m$ из расподеле P , U -статистика са језгром h дефинисана је са:

$$U_n = U_n(h) = \frac{(n-m)!}{n!} \sum_{P_{m,n}} h(X_{i_1}, \dots, X_{i_m})$$

где је $P_{m,n}$ скуп пермутација од m елемената скупа $\{1, \dots, n\}$.

Уочимо да, уколико је језгро U -статистике симетрична функција m аргумената, тада је:

$$U_n = U_n(h) = \frac{1}{\binom{n}{m}} \sum_{C_{m,n}} h(X_{i_1}, \dots, X_{i_m})$$

где је $C_{m,n}$ скуп комбинација m елемената из скупа величине n , при чему комбинацију уређујемо у растућем поретку, односно $i_1 < \dots < i_m$.

Усмерићемо сада нашу пажњу ка асимптотским својствима U -статистика.

²² Приказана теорија се односи на ставку [13] литературе.

За параметар $\theta(P)$, са симетричним језгром $h(x_1, \dots, x_m)$, који се може проценити, у смислу *Деф III.4.1*, нека је $\mathbf{P}$ класа расподела за које је $Var(h(X_1, \dots, X_m)) < \infty$.

Дефинишимо, за $c = 0, 1, \dots, m$ функције:

$$h_c(x_1, \dots, x_c) = E h(x_1, \dots, x_c, X_{c+1}, \dots, X_m)$$

где су $X_{c+1}, \dots, X_m$ независне случајне величине са расподелом $P \in \mathbf{P}$.

Приметимо да је $h_0 = E(h(X_1, \dots, X_m)) = \theta$, затим $h_m(x_1, \dots, x_m) = h(x_1, \dots, x_m)$, као и:

$$E h_c(X_1, \dots, X_c) = E(h(X_1, \dots, X_m)) = \theta$$

Нека је, даље, за $c = 0, 1, \dots, m$:

$$\sigma_c^2 = Var(h_c(X_1, \dots, X_c))$$

Нагласимо да важи $\sigma_m^2 = Var(h(X_1, \dots, X_m))$.

Теорема III.4.1. За $P \in \mathbf{P}$ важи

$$Var(U_n) = \binom{n}{m}^{-1} \sum_{c=1}^m \binom{m}{c} \binom{n-m}{m-c} \sigma_c^2$$

Ако још и $\sigma_m^2 < \infty$, онда је $Var(U_n) \sim m^2 \sigma_1^2 / n$ за $n \rightarrow \infty$.

Пре доказа теореме, покажећемо следећу лему:

Лема III.4.1. За $P \in \mathbf{P}$ и $(i_1, \dots, i_m), (j_1, \dots, j_m) \in C_{m,n}$, при чему комбинације $(i_1, \dots, i_m)$ и $(j_1, \dots, j_m)$ имају c заједничких елемената, важи:

$$cov(h(X_{i_1}, \dots, X_{i_m}), h(X_{j_1}, \dots, X_{j_m})) = cov(h_c(X_1, \dots, X_c), h(X_1, \dots, X_m)) = \sigma_c^2$$

Доказ:

Имамо:

$$\begin{aligned} cov(h(X_{i_1}, \dots, X_{i_m}), h(X_{j_1}, \dots, X_{j_m})) &= E((h(X_1, \dots, X_c, X_{c+1}, \dots, X_m) - \theta)(h(X_1, \dots, X_c, X'_{c+1}, \dots, X'_m) - \theta)) \\ &= E(E((h(X_1, \dots, X_c, X_{c+1}, \dots, X_m) - \theta)(h(X_1, \dots, X_c, X'_{c+1}, \dots, X'_m) - \theta)) | X_1, \dots, X_c) \\ &= E((h_c(X_1, \dots, X_c) - \theta)(h_c(X_1, \dots, X_c) - \theta)) = \sigma_c^2 \end{aligned}$$

Такође:

$$\begin{aligned}
\text{cov}(h_c(X_1, \dots, X_c), h(X_1, \dots, X_m)) &= E((h_c(X_1, \dots, X_c) - \theta)(h(X_1, \dots, X_m) - \theta)) \\
&= E(E((h_c(X_1, \dots, X_c) - \theta)(h(X_1, \dots, X_m) - \theta)) | X_1 \dots X_c) \\
&= E((h_c(X_1, \dots, X_c) - \theta)(h_c(X_1, \dots, X_c) - \theta)) = \sigma_c^2
\end{aligned}$$

чиме је лема доказана.

Доказ теореме:

$$\text{Var}(U_n) = \text{Var}\left(\binom{n}{m}^{-1} \sum_{i \in C_{m,n}} h(X_{i_1}, \dots, X_{i_m})\right) = \binom{n}{m}^{-2} \sum_{i \in C_{m,n}} \sum_{j \in C_{m,n}} \text{cov}(h(X_{i_1}, \dots, X_{i_m}), h(X_{j_1}, \dots, X_{j_m}))$$

Како комбинације $(X_{i_1}, \dots, X_{i_m})$ и $(X_{j_1}, \dots, X_{j_m})$ могу имати c заједничких елемената, пребројаћемо парове са одговарајућим бројем заједничких елемената, и применити претходну лему. Пошто m -торку $(i_1, \dots, i_m)$ можемо одабрати на $\binom{n}{m}$, затим c елемената која ће тој, и m -торки $(j_1, \dots, j_m)$ бити заједнички, на $\binom{m}{c}$ начина, и, коначно, остале елементе m -торке $(j_1, \dots, j_m)$ на $\binom{n-m}{m-c}$ начина, укупан број парова комбинација које имају тачно c заједничких елемената је $\binom{n}{m} \binom{m}{c} \binom{n-m}{m-c}$.

Тада је:

$$\text{Var}(U_n) = \binom{n}{m}^{-2} \sum_{c=0}^m \binom{n}{m} \binom{m}{c} \binom{n-m}{m-c} \sigma_c^2 = \binom{n}{m}^{-1} \sum_{c=1}^m \binom{m}{c} \binom{n-m}{m-c} \sigma_c^2$$

чиме је први део теореме доказан.

Како, за свако $i < m$, при $\sigma_m^2 < \infty$, важи $\sigma_i^2 < \infty$, то је, асимптотски, претходна сума одређена чланом уз σ_1^2 , па је $\text{Var}(U_n) \sim \frac{m^2}{n} \sigma_1^2$, при $n \rightarrow \infty$, што је и требало показати.

Следеће теорема даје нам важну тврдњу о асимптотској расподели U -статистике.

Теорема III.4.2. Ако је $\sigma_m^2 < \infty$, онда је $\sqrt{n}(U_n - \theta) \rightarrow_D \mathcal{N}(0, m^2 \sigma_1^2)$ при $n \rightarrow \infty$.

Доказ: Нека је $Y_i = m(h_1(X_i) - \theta)$. Тада је $E(Y_i) = 0$, као и $\text{Var}(Y_i) = m^2 \sigma_1^2$, па по централној граничној теореме важи $\sqrt{n}U_n^* \rightarrow_D \mathcal{N}(0, m^2 \sigma_1^2)$ за $U_n^* = \frac{1}{n} \sum_{i=1}^n Y_i$. Пошто је $E(\sqrt{n}U_n^* - \sqrt{n}(U_n - \theta)) = 0$ то је довољно показати $E(\sqrt{n}U_n^* - \sqrt{n}(U_n - \theta))^2 \rightarrow 0$ да бисмо могли да тврдимо асимптотску еквиваленцију величина $\sqrt{n}U_n^*$ и $\sqrt{n}(U_n - \theta)$, из чега би следило тврђење теореме.

Пошто је:

$$E(\sqrt{n}U_n^* - \sqrt{n}(U_n - \theta))^2 = nE(U_n^* - (U_n - \theta))^2 = n\text{Var}(U_n^*) - 2n\text{cov}(U_n^*, U_n) + n\text{Var}(U_n)$$

Затим $n \cdot \text{Var}(U_n^*) = m^2 \sigma_1^2$, као и, на основу *Теореме III.4.1*, $n \text{Var}(U_n) \sim m^2 \sigma_1^2$ за $n \rightarrow \infty$, то је довољно показати да $n \cdot \text{cov}(U_n^*, U_n) = m^2 \sigma_1^2$ како бисмо показали жељену асимптотску еквиваленицију.

Како је:

$$n \cdot \text{cov}(U_n^*, U_n) = \frac{m}{\binom{n}{m}} \sum_{k=1}^n \sum_{j \in C_{m,n}} \text{cov}(h_1(X_k), h(X_{j_1}, \dots, X_{j_m}))$$

Уколико $k \notin \{j_1, \dots, j_m\}$, тада је $\text{cov}(h_1(X_k), h(X_{j_1}, \dots, X_{j_m})) = 0$. У супротном је, по *Лемми III.4.1*, $\text{cov}(h_1(X_k), h(X_{j_1}, \dots, X_{j_m})) = \sigma_1^2$. Како за фиксирано k , постоји $\binom{n-1}{m-1} m$ – торки $\{j_1, \dots, j_m\}$ које садрже k , то је

$$n \cdot \text{cov}(U_n^*, U_n) = \frac{m}{\binom{n}{m}} n \binom{n-1}{m-1} \sigma_1^2 = m^2 \sigma_1^2$$

из чега следи $\sqrt{n}(U_n - \theta) \rightarrow_{\mathcal{D}} \mathcal{N}(0, m^2 \sigma_1^2)$ при $n \rightarrow \infty$, што је и требало показати.

III.5. Асимптотска својства модела-доказ *Теореме III.3.1.*²³

Присетимо се исказа *Теореме III.3.1*, коју желимо да докажемо, при чему су одговарајуће ознаке у складу са одељком III.3.

Теорема III.3.1. Нека је $\int_0^1 \int_0^1 \omega^2(u, v) dudv < \infty$, и нека $\lambda_n \rightarrow \infty$. Тада:

$$\frac{\hat{\rho}_n}{\rho_n} \rightarrow_p 1 \quad (1)$$

$$\sqrt{n} \left(\frac{\hat{\rho}_n}{\rho_n} - 1 \right) \rightarrow_D \mathcal{N}(0, \sigma^2) \quad (2)$$

за неко $\sigma^2 > 0$. Даље, нека је R ацикличан граф и нека је $|V(R)| = p$, $|E(R)| = q$ као и $\int_0^1 \int_0^1 \omega^{2q}(u, v) dudv < \infty$. Тада важи:

$$\check{P}(R) \rightarrow_p \tilde{P}(R) \quad (3)$$

$$\sqrt{n}(\check{P}(R) - \tilde{P}(R)) \rightarrow_D \mathcal{N}(0, \sigma^2(R)) \quad (4)$$

Лема III.5.1. Ако је X_n , $n \geq 1$ низ случајних величина такав да $E(X_n) \rightarrow c$ као и $Var(X_n) \rightarrow 0$ при $n \rightarrow \infty$, тада X_n конвергира у вероватноћи ка c .

Доказ: Лема следи из Чебишевљеве неједнакости $p(|X_n - E(X_n)| \geq \varepsilon) \leq \frac{Var(X_n)}{\varepsilon^2}$ која важи за свако $\varepsilon > 0$, услова теореме, као и дефиниције конвергенције у вероватноћи.

Лема III.5.2. Нека су X и Y случајне величине. Тада $Var(X) = E(Var(X|Y)) + Var(E(X|Y))$.

Доказ: По дефиницији, условна варијанса је $Var(X|Y) = E((X - E(X|Y))^2 | Y)$. Зато је:

$$E(Var(X|Y)) = E(E(X^2|Y)) - E(E(X|Y))^2 = E(X^2) - E(E(X|Y))^2$$

Са друге стране,

$$Var(E(X|Y)) = E(E(X|Y))^2 - (E(E(X|Y)))^2 = E(E(X|Y))^2 - (EX)^2$$

Одакле следи тврђење.

Следећу лему наводимо без доказа:

²³ Одељак припремљен на основу ставке [5] литературе.

Лема III.5.3. Нека су U_n и V_n низови случајних величина, и U и V случајне величине такве да $U_n \rightarrow_D U$ и $V_n|U_n \rightarrow_D V$ при $n \rightarrow \infty$. Тада су U_n и V_n асимптотски независне и $V_n \rightarrow_D V$ за $n \rightarrow \infty$.

Доказ теореме: Како је $\hat{\rho}_n = \frac{2L}{n(n-1)}$ и $\lambda_n = (n-1)\rho_n$, то је $\frac{L}{n\lambda_n} = \frac{1}{2} \cdot \frac{\hat{\rho}_n}{\rho_n}$ и $E\left(\frac{L}{n\lambda_n}\right) = \frac{1}{2}$. Пошто је $L = \sum\{A_{ij}, 1 \leq i < j \leq n\}$ то је:

$$\begin{aligned} Var\left(\frac{L}{n\lambda_n}\right) &= Var\left(\frac{1}{n\lambda_n} \sum\{A_{ij}, 1 \leq i < j \leq n\}\right) = \frac{1}{n^2\lambda_n^2} Var\left(\sum\{A_{ij}, 1 \leq i < j \leq n\}\right) \\ &= \frac{1}{n^2\lambda_n^2} E\left(Var\left(\sum_{i<j} A_{ij}|\xi\right)\right) + \frac{1}{n^2\lambda_n^2} Var\left(E\left(\sum_{i<j} A_{ij}|\xi\right)\right) \\ &= \frac{1}{n^2\lambda_n^2} E\left(Var\left(\sum_{i<j} A_{ij}|\xi\right)\right) + \frac{\rho_n^2}{n^2\lambda_n^2} Var\left(\sum_{i<j} \omega(\xi_i, \xi_j)\right) \end{aligned}$$

јер, због $A_{ij}|\xi$: $\begin{pmatrix} 0 & 1 \\ 1 - \rho_n\omega(\xi_i, \xi_j) & \rho_n\omega(\xi_i, \xi_j) \end{pmatrix}$ важи:

$$Var\left(E\left(\sum_{i<j} A_{ij}|\xi\right)\right) = Var\left(\sum_{i<j} \rho_n\omega(\xi_i, \xi_j)\right) = \rho_n^2 Var\left(\sum_{i<j} \omega(\xi_i, \xi_j)\right)$$

Нека је $T_1 = \frac{1}{n\lambda_n} \sum_{i<j} (A_{ij} - \rho_n\omega(\xi_i, \xi_j))$ и $T_2 = \frac{\rho_n}{n\lambda_n} \sum_{i<j} \omega(\xi_i, \xi_j) - \frac{1}{2}$.

Покажимо $Var(T_1) = \frac{1}{n^2\lambda_n^2} E\left(Var(\sum_{i<j} A_{ij}|\xi)\right)$.

Како је:

$$E\left(Var(\sum_{i<j} A_{ij}|\xi)\right) = E\left(E\left(\left(\sum_{i<j} A_{ij} - E(\sum_{i<j} A_{ij}|\xi)\right)^2|\xi\right)\right) = E\left(\left(\sum_{i<j} A_{ij} - E(\sum_{i<j} A_{ij}|\xi)\right)\right)^2,$$

као и:

$$E\left(\sum_{i<j} A_{ij}|\xi\right) = \sum_{i<j} \rho_n\omega(\xi_i, \xi_j)$$

ТО ИЗ:

$$E\left(E\left(\sum_{i<j} A_{ij}|\xi\right)\right) = E\left(\sum_{i<j} A_{ij}\right)$$

среди:

$$E\left(\sum_{i<j} A_{ij} - E\left(\sum_{i<j} A_{ij}|\xi\right)\right) = 0$$

па је:

$$Var(T_1) = \frac{1}{n^2 \lambda_n^2} E\left(Var\left(\sum_{i<j} A_{ij}|\xi\right)\right)$$

одакле среди:

$$Var\left(\frac{L}{n\lambda_n}\right) = Var(T_1) + Var(T_2).$$

Како бисмо показали да $Var\left(\frac{L}{n\lambda_n}\right) \rightarrow 0$, за $n \rightarrow \infty$ проценимо $Var(T_1)$. При процени, имаћемо у виду следеће:

1. $A_{ij}|\xi: \begin{pmatrix} 0 & 1 \\ 1 - h_n(\xi_i, \xi_j) & h_n(\xi_i, \xi_j) \end{pmatrix}$
2. гране графа су независне
3. $\int_0^1 \int_0^1 \omega(u, v) du dv = 1$ јер је $\omega(u, v)$ густина вектора $(\xi_i, \xi_j)|A_{ij}$
4. $\int_0^1 \int_0^1 \omega^2(u, v) du dv = c < \infty$, што је услов теореме
5. $\rho_n \rightarrow 0$ за $n \rightarrow \infty$, по претпоставци модела

$$\begin{aligned}
Var(T_1) &= \frac{1}{n^2 \lambda_n^2} E \left(Var \left(\sum_{i < j} A_{ij} | \xi \right) \right) \leq \frac{1}{n^2 \lambda_n^2} E \left(Var \left(\sum_{i,j} A_{ij} | \xi \right) \right) = \frac{1}{n^2 \lambda_n^2} E \left(\sum_{i,j} Var(A_{ij} | \xi) \right) \\
&= \frac{1}{n^2 \lambda_n^2} E \left(\sum_{i,j} h_n(\xi_i, \xi_j) (1 - h_n(\xi_i, \xi_j)) \right) \\
&= \frac{1}{n^2 \lambda_n^2} \sum_{i,j} \rho_n E \omega(\xi_i, \xi_j) - \frac{1}{n^2 \lambda_n^2} \sum_{i,j} \rho_n^2 E (\omega(\xi_i, \xi_j))^2 \\
&= \frac{1}{n^2 \lambda_n^2} \sum_{i,j} \rho_n \int_0^1 \int_0^1 \omega(u, v) du dv - \frac{1}{n^2 \lambda_n^2} \sum_{i,j} \rho_n^2 \int_0^1 \int_0^1 \omega^2(u, v) du dv \\
&\leq \frac{1}{n^2 \lambda_n^2} \cdot n^2 \cdot \rho_n - \frac{1}{n^2 \lambda_n^2} \cdot n^2 \cdot \rho_n^2 \cdot c = \frac{\rho_n}{\lambda_n^2} (1 - \rho_n \cdot c) \sim \frac{\rho_n}{\lambda_n^2} = O\left(\frac{1}{n \lambda_n}\right)
\end{aligned}$$

Како, по претпоставци теореме $\lambda_n \rightarrow \infty$, то $Var(T_1) \rightarrow 0$, за $n \rightarrow \infty$. Пошто је $\lambda_n = (n-1)\rho_n$ то је $T_2 = \frac{1}{2} \cdot \frac{1}{\binom{n}{2}} \sum_{i < j} \omega(\xi_i, \xi_j) - \frac{1}{2} = \frac{1}{2} U_n - \frac{1}{2}$, где U_n представља U -статистику са језгром ω . Како је $Var(\omega(\xi_i, \xi_j)) < \infty$, то по Теорему III.4.1. важи $Var(U_n) = O\left(\frac{1}{n}\right)$. Дакле, $Var\left(\frac{L}{n \lambda_n}\right) \rightarrow 0$ за $n \rightarrow \infty$. Подсетимо се, важи $E\left(\frac{L}{n \lambda_n}\right) = \frac{1}{2}$ за свако n . Према Лемми III.5.1, следи $\frac{L}{n \lambda_n} \rightarrow_p \frac{1}{2}$, па имајући у виду $\lambda_n = (n-1)\rho_n$ и $\hat{\rho}_n = \frac{2L}{n(n-1)}$ следи први исказ теореме, односно $\frac{\hat{\rho}_n}{\rho_n} \rightarrow_p 1$.

Посматрајмо, сада, $\sqrt{n \lambda_n} T_1 | \xi$. Нека је $\Psi = Var(\sqrt{n \lambda_n} T_1 | \xi)$. Тада, због независности грана, расподеле $A_{ij} | \xi$, као и чињенице да $\rho_n \rightarrow 0$ за $n \rightarrow \infty$, следи:

$$\begin{aligned}
\Psi &= Var(\sqrt{n \lambda_n} T_1 | \xi) = Var \left(\frac{1}{\sqrt{n \lambda_n}} \sum_{i < j} (A_{ij} - \rho_n \omega(\xi_i, \xi_j)) | \xi \right) = \frac{1}{n \lambda_n} Var \left(\sum_{i < j} (A_{ij} - \rho_n \omega(\xi_i, \xi_j)) | \xi \right) \\
&= \frac{1}{n \lambda_n} Var \left(\sum_{i < j} A_{ij} | \xi \right) \\
&= \frac{1}{n \lambda_n} \sum_{i < j} Var(A_{ij} | \xi) = \frac{1}{n(n-1)} \sum_{i < j} \omega(\xi_i, \xi_j) (1 - \rho_n \omega(\xi_i, \xi_j)) \sim \frac{1}{2} \frac{1}{\binom{n}{2}} \sum_{i < j} \omega(\xi_i, \xi_j)
\end{aligned}$$

Како је $\frac{1}{\binom{n}{2}} \sum_{i < j} \omega(\xi_i, \xi_j)$ U -статистика са језгром ω , то, по Теорему III.4.1. важи $Var(\Psi) \rightarrow 0$. Даље,

$$\begin{aligned} E(\Psi) &= E\left(\frac{1}{n(n-1)} \sum_{i < j} \omega(\xi_i, \xi_j) (1 - \rho_n \omega(\xi_i, \xi_j))\right) \\ &= \frac{1}{n(n-1)} \sum_{i < j} E(\omega(\xi_i, \xi_j)) - \frac{1}{n(n-1)} \sum_{i < j} E(\rho_n (\omega(\xi_i, \xi_j))^2) = \frac{1}{2}(1 - c \cdot \rho_n) \rightarrow \frac{1}{2} \end{aligned}$$

Дакле, према *Леми III.5.1.* важи $Var(\sqrt{n\lambda_n}T_1|\xi) \rightarrow_p \frac{1}{2}$.

Према Линдеберг-Фелеровој теореме следи $\sqrt{n\lambda_n} \cdot T_1|\xi \rightarrow_D \mathcal{N}\left(0, \frac{1}{2}\right)$. Како за свако i важи $\xi_i \in U(0,1)$, то по *Леми III.5.3.* следи $\sqrt{n\lambda_n}T_1 \rightarrow_D \mathcal{N}\left(0, \frac{1}{2}\right)$.

Приметимо да важи:

$$\sigma_2^2 = Var(\omega(\xi_i, \xi_j)) = E\left(\omega(\xi_i, \xi_j) - E(\omega(\xi_i, \xi_j))\right)^2 = E(\omega^2(\xi_i, \xi_j)) - (E\omega(\xi_i, \xi_j))^2 < \infty$$

Такође, $E\omega(\xi_i, \xi_j) = 1$. Тада, по *Теореме III.4.2.*, $\sqrt{n}\left(\frac{2}{n(n-1)} \sum_{i < j} \omega(\xi_i, \xi_j) - 1\right) \rightarrow_D \mathcal{N}(0, V(\xi))$. Одавде једноставно следи $\sqrt{n}T_2 \rightarrow_D \mathcal{N}(0, \tilde{V}(\xi))$.

Даље, $\frac{1}{2} \frac{\hat{\rho}_n}{\rho_n} = \frac{L}{n\lambda_n} = \frac{1}{n\lambda_n} \sum_{i < j} A_{ij} = T_1 + T_2 + \frac{1}{2}$, па је $\sqrt{n}\left(\frac{\hat{\rho}_n}{\rho_n} - 1\right) = 2\sqrt{n}T_1 + 2\sqrt{n}T_2$. Како је $\sqrt{n\lambda_n}T_1 \rightarrow_D \mathcal{N}\left(0, \frac{1}{2}\right)$, то $\sqrt{n}T_1 \rightarrow_p 0$, па је расподела $\sqrt{n}\left(\frac{\hat{\rho}_n}{\rho_n} - 1\right)$ одређена расподелом од $2\sqrt{n}T_2$, односно $\sqrt{n}\left(\frac{\hat{\rho}_n}{\rho_n} - 1\right) \rightarrow_D \mathcal{N}(0, \ddot{V}(\xi))$ чиме је доказан исказ (2) теореме узимајући $\sigma^2 = \ddot{V}(\xi)$.

Посматрајмо, сада, ацикличан граф R са $|V(R)| = p$ чворова, и $|R| = q$ грана.

Важи:

$$\check{P}(R) - \tilde{P}(R) = \check{P}(R) - E(\check{P}(R)|\xi) + E(\check{P}(R)|\xi) - \tilde{P}(R)$$

Нека је $U_1 = \check{P}(R) - E(\check{P}(R)|\xi)$ и $U_2 = E(\check{P}(R)|\xi) - \tilde{P}(R)$. Уведимо, за $S \sim R$:

$$T(S) = \frac{1}{\binom{n}{p} N(R)} \prod \{A_{ij}: (i, j) \in E(S)\}$$

Тада је:

$$U_1 = \frac{\hat{\rho}_n^{-q}}{\binom{n}{p} N(R)} \sum_{S \sim R} I(S \sim R) - \frac{\rho_n^{-q}}{\binom{n}{p} N(R)} \sum_{S \sim R} E\{I(S \sim R) | \xi\}$$

Како за $n \rightarrow \infty$, $\frac{\hat{\rho}_n}{\rho_n} \rightarrow_p 1$, то је, за довољно велико n :

$$U_1 = \rho_n^{-q} \sum_{S \sim R} (T(S) - E(T(S) | \xi))$$

Тада је:

$$\begin{aligned} \text{Var}(U_1 | \xi) &= \rho_n^{-2q} \text{Var} \left(\sum_{S \sim R} (T(S) - E(T(S) | \xi)) | \xi \right) \\ &= \rho_n^{-2q} \sum_{\substack{S_1 \sim R \\ S_2 \sim R}} \text{cov}(T(S_1) - E(T(S_1) | \xi), T(S_2) - E(T(S_2) | \xi) | \xi) \\ &= \rho_n^{-2q} \sum_{\substack{S_1 \sim R \\ S_2 \sim R}} \text{cov}(T(S_1), T(S_2) | \xi) \end{aligned}$$

Дакле,

$$\begin{aligned} E(\text{Var}(U_1 | \xi)) &= \rho_n^{-2q} E \left(\sum_{\substack{S_1 \sim R \\ S_2 \sim R}} \text{cov}(T(S_1), T(S_2)) | \xi \right) \\ &= \rho_n^{-2q} \sum_{\substack{S_1 \sim R \\ S_2 \sim R}} E(\text{cov}(T(S_1), T(S_2)) | \xi) \leq \rho_n^{-2q} \sum_{\substack{S_1 \sim R \\ S_2 \sim R}} E(E(T(S_1)T(S_2)) | \xi) \\ &= \frac{\rho_n^{-2q}}{\binom{n}{p}^2 (N(R))^2} \sum_{\substack{S_1 \sim R \\ S_2 \sim R}} EE \left(\prod \{A_{ij} : (i, j) \in S_1\} \prod \{A_{ij} : (i, j) \in S_2\} | \xi \right) \end{aligned}$$

Приметимо да, уколико је $E(S_1) \cap E(S_2) = \emptyset$, важи $\text{cov}(T(S_1), T(S_2)) = 0$, због независности грана дисјунктних графова. Претпоставимо да граф $S_1 \cap S_2$ има c чворова и d грана. Пошто је R ацикличан, тада су и $S_1, S_2 \sim R$ ациклични, па је и $S_1 \cap S_2$ ацикличан. По *Теорему 1.2.2.* следи $c \geq d + 1$, односно $c > d$. Тада је:

$$\begin{aligned} & \frac{\rho_n^{-2q}}{\binom{n}{p}^2 (N(R))^2} EE \left(\prod \{A_{ij}: (i, j) \in S_1\} \prod \{A_{ij}: (i, j) \in S_2\} | \xi \right) \\ &= \frac{\rho_n^{-2q}}{\binom{n}{p}^2 (N(R))^2} \rho_n^{2q-d} E \left(\prod \{\omega(\xi_i, \xi_j): (i, j) \in S_1 \cup S_2\} \right) \\ &= O \left(\rho_n^{-d} n^{-2p} E \left(\prod \{\omega(\xi_i, \xi_j): (i, j) \in S_1 \cup S_2\} \right) \right) \end{aligned}$$

због $\binom{n}{p} = O(n^p)$ при $n \rightarrow \infty$.

Како, по услову теореме, важи $\int_0^1 \int_0^1 \omega^{2q}(u, v) du dv < \infty$, то сваки пар (S_1, S_2) са c заједничких чворова и d заједничких грана доприноси очекивању $E(\text{Var}(U_1 | \xi))$ са највише $O(\rho_n^{-d} n^{-2p} \int_0^1 \int_0^1 \omega^{2q}(u, v) du dv)$. Размотримо колико таквих парова имамо. Како је, на n чворова, потребно разапети два графа изоморфна датом графу R , са c заједничких чворова, при чему сваки од њих има укупно p чворова, то је укупан број таквих парова једнак $\binom{n}{c} \cdot \binom{n-c}{2p-2c} \cdot \binom{2p-2c}{p-c} = O(n^c n^{2p-2c}) = O(n^{2p-c})$ за $n \rightarrow \infty$. Тада је укупан допринос оваквих парова очекивању $E(\text{Var}(U_1 | \xi))$ једнак $O(\rho_n^{-d} n^{-2p} n^{2p-c}) = O(\lambda_n^{-d} n^{d-c}) = o(n^{-1})$, за $n \rightarrow \infty$, при чему последња асимптотска еквиваленција следи из $\lambda_n \rightarrow \infty$ и $c > d$. Како за број заједничких темена произвољна два графа S_1 и S_2 , која задовољавају наведене услове, важи $1 \leq c \leq p$, и за одговарајући број заједничких грана d , за свако фиксирано c , јесте $d_c < c$, то је укупно $E(\text{Var}(U_1 | \xi)) = o(n^{-1})$.

Пошто је $Var(U_1) = E(Var(U_1|\xi)) + Var(E(U_1|\xi))$, као и $Var(E(U_1|\xi)) = Var(E(\rho_n^{-q} \sum_{S \sim R} (T(S) - E(T(S)|\xi)))|\xi) = 0$ следи $Var(U_1) = o(n^{-1})$. Како је $E(U_1) = E(\check{P}(R) - E(\check{P}(R)|\xi)) = 0$, то, по Лему III.5.1, следи $U_1 \rightarrow_p 0$, из чега следи $U_1 \rightarrow_D 0$. Уочимо да, због $Var(U_1) = o(n^{-1})$ следи $Var(\sqrt{n}U_1) \rightarrow 0$, па важи $\sqrt{n}U_1 \rightarrow_D 0$.

Пошто је $U_2 = E(\check{P}(R)|\xi) - \tilde{P}(R)$, затим $E(\check{P}(R)|\xi) = \rho_n^{-q} E(\hat{P}(R)|\xi)$, као и:

$$\begin{aligned} E(\hat{P}(R)|\xi) &= E\left(\frac{1}{\binom{n}{p} N(R)} \sum_{\substack{\bar{S} \sim R \\ S \subseteq G}} (I(S \sim R)|\xi)\right) \\ &= \frac{1}{\binom{n}{p} N(R)} \sum_{\substack{\bar{S} \sim R \\ S \subseteq G}} E((I(S \sim R)|\xi)) = \frac{1}{\binom{n}{p} N(R)} \sum_{\substack{\bar{S} \sim R \\ S \subseteq G}} \left(\prod_{(i,j) \in S} \rho_n \omega(\xi_i, \xi_j) \prod_{(i,j) \in \bar{S}} (1 - \rho_n \omega(\xi_i, \xi_j)) \right) \end{aligned}$$

где је $\bar{S} = \{(i,j) \notin S, i \in V(S), j \in V(S)\}$, то је:

$$\begin{aligned} U_2 &= \frac{\rho_n^{-q}}{\binom{n}{p} N(R)} \sum_{\substack{\bar{S} \sim R \\ S \subseteq G}} \left(\prod_{(i,j) \in S} \rho_n \omega(\xi_i, \xi_j) \prod_{(i,j) \in \bar{S}} (1 - h_n(\xi_i, \xi_j)) \right) - \tilde{P}(R) \\ &= \frac{1}{\binom{n}{p} N(R)} \sum_{\substack{\bar{S} \sim R \\ S \subseteq G}} \left(\prod_{(i,j) \in S} \omega(\xi_i, \xi_j) \prod_{(i,j) \in \bar{S}} (1 - h_n(\xi_i, \xi_j)) - \tilde{P}(S) \right) \end{aligned}$$

Пошто је $f(\xi_1, \dots, \xi_p) = \left(\prod_{(i,j) \in S} \omega(\xi_i, \xi_j) \prod_{(i,j) \in \bar{S}} (1 - h_n(\xi_i, \xi_j)) - \tilde{P}(S) \right)$ симетрична функција p аргумената, јер је $S \sim R$, односно јер свака пермутација чворова одређује гране тако да се добије граф изоморфан са R , то је ова функција језгро U -статистике U_2 .

Како је:

$$E\left(f(\xi_1, \dots, \xi_p)\right) = E\left(\prod_{(i,j) \in S} \omega(\xi_i, \xi_j) \prod_{(i,j) \in \bar{S}} (1 - h_n(\xi_i, \xi_j))\right) - \tilde{P}(S) = 0$$

као и $Var\left(f(\xi_1, \dots, \xi_p)\right) < \infty$, то можемо применити *Теорему III.4.2.* на U -статистику U_2 .

Дакле, из $E\left(f(\xi_1, \dots, \xi_p)\right) = 0$ следи $\sqrt{n}U_2 \rightarrow_{\mathcal{D}} \mathcal{N}(0, \bar{\sigma}^2(R))$, па, коначно, уз $\sqrt{n}U_1 \rightarrow_{\mathcal{D}} 0$ следи:

$$\sqrt{n}(\check{P}(R) - \tilde{P}(R)) \rightarrow_{\mathcal{D}} \mathcal{N}(0, \sigma^2(R))$$

што је и требало показати.

IV Неки актуелни изазови проучавања случајних графова

Област статистичког моделовања мрежа веома је жива, о чему сведоче бројни отворени проблеми и изазови са којима се сусрећу статистичари који се баве овом облашћу.

Постоје многа ограничења која отежавају доношење статистичких закључака о припадности посматране мреже одређеном моделу, као и о сличности, односно различитости мрежа које се међусобно пореде. Већина тешкоћа произилази из чињенице да статистике које се најчешће користе, као што су расподела степена чвора графа, и коефицијент груписаности, одражавају само неке особине посматране мреже, док постоје многе друге особине које посматрањем ових статистика бивају занемарене. Алтернатива овом, тзв. глобалном приступу, представља посматрање локалних својстава мреже, која се огледају у присутности и специфичној распоређености одређених подграфова.

Међутим, утврђивање присутности подграфова у посматраном графу није нимало лак задатак, због алгоритамске сложености овог проблема. За решавање овог проблема предлагани су разни алгоритми²⁴. Неки предложени алгоритми односе се на специјалне случајеве подграфова. На пример, у чланку [26] предлаже се алгоритам за проналажење свих повезаних подграфова са следећом особином: уклањањем било којих $k - 1$ његових грана (за фиксирано k) подграф и даље остаје повезан, док постоји скуп од k грана чијим уклањањем подграф престаје да буде повезан, а не постоји подграф са овом особином који посматрани подграф садржи. Такође, у чланку [14] предлажу се алгоритми за проналажење неких подграфова који се могу разложити на дрвета.

Даље, ради доношења исправног закључка о припадности посматране мреже одговарајућем моделу, потребно је разматрати што је могуће већи број њених особина. У том циљу се у [22] предлаже уопштење степена чвора графа, на основу кога се добија статистика која осликава скуп особина које се односе на неке специфичности распоређености одређених подструктура у графу.

На пољу софтверске имплементације поређења посматране мреже са неким моделима случајних графова, такође су учињени значајни пробоји. Тако је у [18] приказан софтвер којим се имплементира упоређивање посматране мреже са неким моделима случајних графова, узимајући у обзир разна локална и глобална својства мреже. Закључак о значају овог доприноса моделирању случајних мрежа може се лако извести из његове практичне применљивости: у [23] су применом овог метода изведени неки закључци о особинама мрежа које представљају интеракције међу протеинима.

Теоријска истраживања у области асимптотских карактеристика случајних графова такође су врло актуелна.

Тако су нека асимптотска својства неких подграфова графа Erdős-Rényi-евог модела доказана у [7].

²⁴ Погледати у [25].

У извођењу неких асимптотских својстава случајних графова као веома корисно се показује успостављање везе случајних графова са одређеним случајним процесима.²⁵

Све ово поткрепљује тврдњу о актуелности поља истраживања случајних графова.

²⁵ Погледати у [1].

Списак литературе:

- [1] Addario-Berry L., Broutin N., Goldschmidt C., 2009, *The continuum limit of critical random graphs*, <http://arxiv.org/pdf/0903.4730v2.pdf>.
- [2] Aldous D.J., 2009, *More uses of Exchangeability: Representations of Complex Random Structures*, <http://arxiv.org/pdf/0909.4339.pdf>.
- [3] Bender E.A., Williamson S.G., 2010, *Lists, Decisions and Graphs With an Introduction to Probability, Unit GT: Basic Concepts in Graph Theory*, <http://cseweb.ucsd.edu/~gill/BWLectSite/Resources/C2U4GT.pdf>.
- [4] Bickel P.J., Chen A., 2009, *A nonparametric view of network models and Newman-Girvan and other modularities*, PNAS, 106 (50), 21062-21073.
- [5] Bickel P.J., Chen A., Levina E., 2011, *The method of moments and degree distributions for network models*, Ann. Statist., 39 (5), 2280-2301.
- [6] Bollobas B., 1984, *The Evolution of Random Graphs*, Trans. Amer. Math. Soc., 286, 257-274.
- [7] Brightwell G., Panagiotou K., Steger A., 2009, *Extremal Subgraphs of Random Graphs: an Extended Version*, <http://arxiv.org/pdf/0908.3778v1.pdf>.
- [8] Cohen J.E., 1988, *Threshold Phenomena in Random Structures*, Discrete Applied Mathematics, 19, 113-128.
- [9] Diaconis P., Janson S., 2007, *Graph Limits and Exchangeable Random Graphs*, <http://www2.math.uu.se/~svante/papers/sj209.pdf>.
- [10] Dorogovtsev S.N., Mendes J.F.F., 2001, *Evolution of networks*, Advances in Physics, 51 (4), 1079-1187, <http://arxiv.org/pdf/cond-mat/0106144v2.pdf>.
- [11] Erdos P., Renyi A., 1959, *On random graphs I*, http://ftp.math-inst.hu/~p_erdos/1959-11.pdf.
- [12] Erdos P., Renyi A., 1960, *On the Evolution of Random Graphs*, Publ. Math. Inst. Hungar. Acad. Sci, 5, 17-61.
- [13] Ferguson T.S., 2005, *U-statistics, Notes for Statistics*, <http://www.math.ucla.edu/~tom/Stat200C/Ustat.pdf>.
- [14] Fomin V.F., Lokstahnov D., Raman V., Saurabh S., Raghavendra Rao B.V., 2012, *Faster Algorithms for Finding and Counting Subgraphs*, Journal of Computer and System Sciences, 78 (3), 698-706.
- [15] Hardy G.H., Littlewood J.E., Polya G., 1988, *Inequalities*, Cambridge University Press, Cambridge, UK, 2nd Ed.
- [16] Kallenberg O., 2005, *Probabilistic Symmetries and Invariance Principles. Probability and Its Applications*, Springer, New York.

- [17] Merkle M., 2010, *Verovatnoća i statistika za inženjere i studente tehnike*, Akademska misao, Beograd.
- [18] Milenković T., Lai J., Pržulj N., 2008, *GraphCrunch: A Tool for Large Network Analysis*, BMC Bioinformatics, 9, 70-81.
- [19] Младеновић П., 2008, *Вероватноћа и статистика*, Математички факултет, Београд.
- [20] Nier E., Yang J., Yorulmazer T., Alenthorn A., 2008, *Network models and financial stability*, Working Paper No. 346, Bank of England.
- [21] Phizicky E., Fields S., 1995, *Protein- Protein Interactions: Methods for Detection and Analysis*, Microbiol. Rev, 59 (1), 94-123.
- [22] Pržulj N., 2007, *Biological Network Comparison Using Graphlet Degree Distribution*, Bioinformatics, 23, 177-183.
- [23] Rito T., Wang Z., Deane C.M., Reinert G., 2010, *How threshold behaviour affects the use of subgraphs for network comparison*, Bioinformatics, 26 (18), i611-i617.
- [24] Soffer S.N., Vazquez A., 2005, *Network Clustering Coefficient Without Degree- Correlation Biases*, Physical Review E71, 057101.
- [25] Ullman J.R., 1976, *An Algorithm for Subgraph Isomorphism*, Journal of the ACM, 23 (1), 31-42.
- [26] Zhou R, Liang W, Liu C, Chen B, Xu Yu J, Li J, 2012, *Finding Maximal k-Edge-Connected Subgraphs from a large Graph*, Proceedings of the 15 International Conference on Extending Database Technology, 480-491.