

ՀՀ ԳԱԱ ԻՆՖՈՐՄԱՏԻԿԱՅԻ ԵՎ ԱՎՏՈՄԱՏԱՑՄԱՆ  
ՊՐՈՔԼԵՄՆԵՐԻ ԻՆՍՏԻՏՈՒՏ

---

Մխիթարյան Կարեն Կառլենի

**ՀԱՄԱՅՆՔՆԵՐԻ ՀԱՅՏՆԱԲԵՐՄԱՆ ԻՆՖՈՐՄԱՑԻՈՆ ՏԵՍԱԿԱՆ ՀԵՏԱԶՈՏՈՒՄ ԵՎ  
ԽՈՐՀՐԴԱՏՎԱԿԱՆ ՄԻՋԱՎԱՅՐԻ ՄՇԱԿՈՒՄ**

Ե. 13.05 «Մաթեմատիկական մոդելավորում, թվային մեթոդներ և ծրագրերի  
համալիրներ» մասնագիտությամբ  
տեխնիկական գիտությունների թեկնածուի գիտական աստիճանի հայցման  
ատենախոսության

ՍԵՂՄԱԳԻՐ

ԵՐԵՎԱՆ – 2019

---

---

ИНСТИТУТ ПРОБЛЕМ ИНФОРМАТИКИ И  
АВТОМАТИЗАЦИИ НАН РА

---

Мхитарян Карен Карленович

**ИНФОРМАЦИОННО ТЕОРЕТИЧЕСКОЕ ИССЛЕДОВАНИЕ ОБНАРУЖЕНИЯ  
СООБЩЕСТВ И РАЗРАБОТКА РЕКОМЕНДАТЕЛЬНОЙ СРЕДЫ**

АВТОРЕФЕРАТ

диссертации на соискание ученой степени  
кандидата технических наук  
по специальности

05.13.05 – “Математическое моделирование, численные методы и комплексы  
программ”

Ереван – 2019

Ատենախոսության թեման հաստատվել է ՀՀ ԳԱԱ Ինֆորմատիկայի և ավտոմատացման պրոբլեմների ինստիտուտում:

Գիտական ղեկավար՝	Ֆիզ. մաթ. գիտ. դոկտոր	Մ. Ե. Հարությունյան
Պաշտոնական ընդդիմախոսներ՝	Ֆիզ. մաթ. գիտ. դոկտոր PhD	Է. Մ. Պողոսյան Մ. Շեվալիե (Ֆրանսիա)
Առաջատար կազմակերպություն՝	Հայաստանի ազգային համալսարան	պոլիտեխնիկական

Պաշտպանությունը կայանալու է 2019թ. հունիսի 14-ին, ժամը՝ 16:00-ին ՀՀ ԳԱԱ Ինֆորմատիկայի և ավտոմատացման պրոբլեմների ինստիտուտում գործող 037 «Ինֆորմատիկա» մասնագիտական խորհրդի նիստում հետևյալ հասցեով՝ Երևան, 0014, Պ. Սևակի 1:

Ատենախոսությանը կարելի է ծանոթանալ ՀՀ ԳԱԱ ԻԱՊԻ գրադարանում:  
Սեղմագիրն առաքված է 2019թ. մայիսի 3-ին:

Մասնագիտական խորհրդի  
գիտական քարտուղար, ֆիզ. մաթ. գիտ. դոկտոր

Հ.Գ. Սարվխանյան

---

Тема диссертации утверждена в Институте проблем информатики и автоматизации НАН РА

Научный руководитель:	доктор физ. мат. наук	М.Е.Арутюнян
Официальные оппоненты:	доктор физ. мат. наук PhD	Э. М. Погосян М. Шевалье (Франция)

Ведущая организация: Национальный Политехнический Университет Армении

Защита состоится 14-ого июня 2019г. в 16:00 на заседании специализированного совета 037 “Информатика” в Институте проблем информатики и автоматизации НАН РА по адресу: 0014, г. Ереван, ул. П. Севака 1.

С диссертацией можно ознакомиться в библиотеке ИПИА НАН РА.

Автореферат разослан 3-ого мая 2019г.

Ученый секретарь специализированного совета,  
доктор физ. мат. наук

А. Г. Саруханян

### Թեմայի արդիականությունը

Վերջին տարիներին տեղեկատվական տեխնոլոգիաների բուռն զարգացումը մասնագետներին տրամադրել է զանգվածային տվյալներ և հաշվարկային բավարար ռեսուրսներ այդ տվյալների մշակման և վերլուծության համար: Զգալիորեն աճել են նաև իրական ցանցերի չափերը հասնելով միլիոնավոր կամ միլիարդավոր գագաթների և կողերի: Նման մեծ թվով միավորների հետ աշխատելու անհրաժեշտությունը խորապես փոխել է գրաֆների հետազոտման մոտեցումները:

Ցանցը գագաթների հավաքածու է, որոնք միմյանց հետ կապված են տարբեր հասկություններով օժտված կողերով: Բարդ ցանցերը գրաֆների կամ ցանցերի տեսակ են, որոնք օժտված են ոչ տրիվիալ տոպոլոգիական հատկանիշներով, որոնք ի տարբերություն պարզ գրաֆների, հիմնականում հանդիպում են իրական աշխարհի փոխազդեցությունները մոդելավորող ցանցերում: Բարդ ցանցերի օրինակ են սոցիալական, տեխնոլոգիական, տեղեկատվական և կենսաբանական ցանցերը, որոնցում էականորեն նկատվում է ոչ տրիվիալ տոպոլոգիական կառուցվածք՝ անհամասեռ գագաթային աստիճանների բաշխում, համայնքային և հիերարխիկ կառուցվածք, ինչպես նաև գագաթների միջև ո՛չ կատարելապես կանոնավոր, ո՛չ ամբողջովին պատահական կապեր: Բարդ ցանցերի ուսումնասիրումը երիտասարդ գիտական ուղղություն է, որի զարգացմանը խթանել են համակարգչային և սոցիալական ցանցերի ուսումնասիրումը:

Իրական աշխարհի բարդ ցանցերը պարունակում են գագաթների թաքնված խմբավորումներ՝ համայնքներ, որոնք իրենց համայնքի ներսում ունեն ավելի խիտ փոխազդեցություններ քան ցանցի մնացած այլ համայնքների հետ: Ցանցերում համայնքների ուսումնասիրումը ունի բազմաթիվ կիրառություններ տարբեր ոլորտներում, ինչպիսիք են համակարգչային գիտությունը, մեքենայական ուսուցումը, բժշկությունը, կենսաբանությունը և սոցիոլոգիան: Համայնքների հայտնաբերման նպատակն է ցանցերում գտնել խիտ փոխազդեցություններով օժտված գագաթների խմբավորումներ<sup>1</sup>:

Չնայած նրան, որ գրականության մեջ առկա են համայնքների հայտնաբերման բազմաթիվ մոտեցումներ և ալգորիթմներ, դրանց գնահատումը մնում է բարդ խնդիր, քանի որ տարբեր ցանցերի համար ստացվում են տարբեր արդյունքներ: Այդ պատճառով ոլորտում դեռ կան բազմաթիվ բաց խնդիրներ: Որոշ դեպքերում

---

<sup>1</sup> S. Fortunato and D. Hric, "Community detection in networks: A user guide," *Physics Reports*, vol. 659, no. 11, pp. 1-44, 2016.

համայնքների հայտնաբերման շատ մոտեցումներ կարող են հաջողել, մինչդեռ այլ դեպքերում ալգորիթմները կարող են ձախողվել հաշվարկային բարդության պատճառով: Այսպիսով, կարևոր խնդիր է ուսումնասիրել այն իրավիճակները, երբ համայնքների հայտնաբերումը կարող է լուծվել արդյունավետորեն կամ ինֆորմացիոն-տեսականորեն:

Համայնքների հայտնաբերման ալգորիթմների միջոցով ստացված համայնքային կառուցվածքների գնահատման համար կարևոր է դրանց ուսումնասիրումը տարբեր տեսակի ցանցերի վրա, օգտագործելով համայնքների և համայնքային կառուցվածքների գնահատման և համեմատման ներքին և արտաքին չափեր:

Հավանականային ցանցերի մոդելները կամ պատահական գրաֆները կարող են օգտագործվել իրական ցանցերը մոդելավորելու և խնդիրների բարդությունը գնահատելու համար, կամ ստեղծելու ցանցեր՝ որտեղ իրական համայնքային կառուցվածքը հայտնի է: Վերջինս հնարավորություն է տալիս հասկանալու, թե որքանով է մոդելը համապատասխանում իրական տվյալներին, և շատ կարևոր է համայնքների հայտնաբերման խնդիրներում, քանի որ կիրառության մեջ բավականաչափ քիչ են ցանցերը, որտեղ իրական համայնքային կառուցվածքը հայտնի է: Դա, մասնավորապես, հայտնի մարտահրավեր է մեծ տվյալների խնդիրներում, որտեղ հնարավոր չէ որոշել կլաստերների որակը մեխանիկորեն:

Համայնքների հայտնաբերման բարձր արդյունավետությամբ ալգորիթմներ որոնելիս, որոնք կկարողանան հայտնաբերել «ճիշտ» համայնքներ կամ համեմատել տրված համայնքային կառուցվածքը իրական համայնքային կառուցվածքի հետ, օգտագործվում են տարբեր համեմատման չափեր: Այս ոլորտում ինֆորմացիոն տեսական չափերը հիմնարար տեղ են զբաղեցնում և վերջին տարիներին մեծ հետաքրքրություն են առաջացրել իրենց ուժեղ մաթեմատիկական հիմքի և ոչ գծային նմանությունների հայտնաբերման ունակությունների շնորհիվ: Այնուամենայնիվ, ըստ տարբեր հեղինակների պնդումների<sup>2</sup>, հայտնի ինֆորմացիոն տեսական չափերը, ինչպիսիք են օրինակ՝ նորմավորված փոխադարձ ինֆորմացիան (NMI) և ինֆորմացիայի փոփոխությունը (VI), հուսալի վարք չեն դրսևորում<sup>3</sup>, քանի որ դրանք կախում ունեն ցանցում գազաթների և համայնքների քանակից: Այդ կոնտեքստում

---

<sup>2</sup> N. X. Vinh, J. Epps and J. Bailey, "Information Theoretic Measures for Clusterings Comparison: Variants, Properties, Normalization and Correction for Chance," *The Journal of Machine Learning Research*, vol. 11, pp. 2837-2854, 2010.

<sup>3</sup> A. Amelio and C. Pizzuti, "Is Normalized Mutual Information a Fair Measure for Comparing Community Detection Methods?," in *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, Paris, France, 2015.

կարևոր է փորձարկել դրանց կատարողականը, հայտնաբերելու այն դեպքերը, երբ ստացվում են սխալ արդյունքներ և առաջարկել առկա չափերի ձևափոխություններ կամ այլընտրանքներ, որոնք կերաշխավորեն ավելի ճշգրիտ համեմատություններ:

Համայնքների հայտնաբերման մոտեցումները ներմուծվել են բազմաթիվ այլ գիտական ուղղություններ, որոնցից մեկը խորհրդատվական համակարգերն են: Խորհրդատվական համակարգը ինֆորմացիայի որոնման համակարգի տեսակ է, որը կանխագուշակում է օգտատերերի նախընտրություններն առարկաների վերաբերյալ՝ նպատակ ունենալով ստեղծել անհատականացված խորհրդատվություններ: Խորհրդատվական համակարգերի օգտագործումը բազմաթիվ ընկերություններին հնարավորություն է տվել ավելացնել եկամուտները, օգտատերերի գոհունակությունը, ինչպես նաև ակտիվացնել անհատականացում և օգտատերերի համար նոր ինֆորմացիայի բացահայտում: Բովանդակության վրա հիմնված և համագործակցային ֆիլտրման մեթոդներն առավել հայտնի մոտեցումներն են խորհրդատվական համակարգերում, որոնք կանխագուշակում են օգտատերերի նախընտրությունները՝ հիմնվելով համակարգում առկա օգտատերերի և առարկաների միջև համագործակցային տվյալների վրա:

Սակայն, առկա խորհրդատվական մոտեցումներն ունեն որոշակի թերություններ, ինչպիսիք են նոսր տվյալների հետ աշխատելու դժվարությունները, մեծածավալ տվյալների դեպքում հաշվողական բարդությունը, որոնք սահմանափակում են դրանց օգտագործումը որոշ կիրառություններում: Այդպիսի դեպքերում համայնքների հայտնաբերման մեթոդների կիրառումը հեռանկարային է, քանի որ դրանց միջոցով հնարավոր է բարելավել խորհրդատվություններ իրականացնելու գործընթացի արդյունավետությունը և խորհրդատվությունների ճշգրտությունը:

### **Հետազոտության նպատակն ու խնդիրները**

Առենախոսության հիմնական նպատակն էր առաջարկել գիտականորեն հիմնավորված մոտեցում համայնքների հայտնաբերման ալգորիթմներին և գնահատման չափերին որոշ կիրառություններում:

Այդ նպատակով առաջին խնդիրն էր ուսումնասիրել համայնքների հայտնաբերման առկա մոտեցումները և ալգորիթմները, հետազոտել համայնքների և համայնքային կառուցվածքների գնահատման ներքին և արտաքին առկա չափերը՝ համեմատելով դրանք տարբեր տեսակի փոքր և մեծ իրական բարդ և արհեստականորեն ստացված ցանցերի օգտագործմամբ:

Հաջորդ խնդիրն էր առաջարկել նոր մոտեցումներ համայնքների հայտնաբերման ալգորիթմների համեմատման և օգտագործման համար որոշ կիրառություններում:

Վերջապես խնդիր էր դրվել ուսումնասիրել համայնքների հայտնաբերման և խորհրդատվական համակարգերի հնարավոր փոխկապակցվածությունը, նպատակ հետապնդելով առաջարկել ավելի արդյունավետ խորհրդատվական մոտեցումներ:

### **Հետազոտման օբյեկտը**

Ատենախոսության մեջ հետազոտվել են բարդ ցանցերը, պատահական գրաֆների մոդելները, համայնքների հայտնաբերման հիմնական ալգորիթմները, համայնքների և համայնքային կառուցվածքների գնահատման և համեմատման ներքին և արտաքին չափերը և ինֆորմացիոն տեսական չափերը: Ուսումնասիրվել է նաև համայնքների հայտնաբերման մոտեցումների ներառումը խորհրդատվական համակարգերի որոշ կիրառություններում:

### **Հետազոտման մեթոդները**

Ատենախոսության շրջանակներում կիրառվել են հավանականությունների տեսության, ինֆորմացիայի տեսության, ցանցերի տեսության և կլաստերային վերլուծության հիմնարար մեթոդները: Փորձարարական հետազոտությունների համար օգտագործվել են R և Python ծրագրավորման լեզուները և դրանց՝ ցանցերի հետազոտության և կլաստերային վերլուծության համար նախատեսված գրադարանները: Խոշոր ցանցերում հաշվողական մեծ բարդություն պահանջող հետազոտությունների համար օգտագործվել է «OSIRIM» գերհամակարգիչը (<http://osirim.irit.fr/site/>):

### **Հետազոտության գիտական նորույթը**

1. Առաջարկվել է  $\chi^2$  – տարամիտության վրա հիմնված նոր ինֆորմացիոն տեսական չափ համայնքային կառուցվածքների համեմատման համար, որը գրականության մեջ առկա չափերի համեմատ ունի առավելություն:
2. Առաջարկվել է համայնքների հայտնաբերման մեթոդների վրա հիմնված խորհրդատվական միջավայր, որը որոշ կիրառություններում ունակ է կատարել խորհրդատվություններ, երբ այլ խորհրդատվական մոտեցումները ձախողվում են:

### **Ստացված արդյունքների կիրառական նշանակությունը**

Կատարվել է համայնքների հայտնաբերման ալգորիթմների խորը հետազոտություն և գնահատում օգտագործելով իրական աշխարհի փոքր և մեծ բարդ և ստոխաստիկ բլրկ մոդելի միջոցով արհեստականորեն ստացված ցանցեր: Կիրառվել են համայնքների գնահատման՝ ներքին և համայնքային կառուցվածքների համեմատման՝ արտաքին չափեր: Ստացված արդյունքները շատ կարևոր են

հասկանալու, թե ինչպես են աշխատում ալգորիթմները, և որ ալգորիթմն է գերադասելի տրված դեպքերում:

Խորհրդատվական համակարգերի կոնտեքստում համայնքների հայտնաբերման մոտեցումների օգտագործմամբ առաջարկված և իրականացված նոր միջավայրը կարող է կիրառվել բազմաթիվ ոլորտներում, որտեղ այլ խորհրդատվական մոտեցումներն անօգուտ են: Միջավայրի իրականացումը կարող է օգտագործվել տարբեր կիրառություններում:

### **Ներդրումը**

Ստեղծված խորհրդատվական միջավայրը ներդրված է որպես գործիք Հայաստանի Ակադեմիական Գիտահետազոտական Կոմպյուտերային Ցանցի ([www.asnet.am](http://www.asnet.am)) ծառայությունների ցանկում:

### **Պաշտպանությանը ներկայացվող հիմնական դրույթները**

1. Առաջարկվել է համայնքային կառուցվածքների կամ կլաստերների համեմատման ինֆորմացիոն տեսական չափի հիմնված  $\chi^2$ - տարամիտության վրա:
2. Առաջարկվել է համայնքների հայտնաբերման մոտեցումների վրա հիմնված նոր խորհրդատվական միջավայր և ներկայացվել դրա իրականացումը:

### **Ապրոբացիա**

Ատենախոսության արդյունքները զեկուցվել են՝

- «Համակարգչային գիտություններ և տեղեկատվական տեխնոլոգիաներ (CSIT 2017)» 11-րդ միջազգային գիտաժողովում, Երևան, սեպտ. 25-29, 2017թ.,
- Էմիլ Արթինի 120 ամյակին նվիրված միջազգային գիտաժողովում, Երևան, մայ. 27 - հուն. 2, 2018թ.,
- «Collaborative Technologies and Data Science in Smart City Applications (CODASSCA)» միջազգային գիտաժողովում, Երևան, սեպտ. 12-15, 2018թ.
- «Science and Technology Convergence» ֆորումում, Երևան, հոկ. 12-13, 2018թ.,
- ՀՀ ԳԱԱ ԻԱՊԻ ընդհանուր սեմինարում, Երևան, 2019:

### **Հրատարակումները**

Ատենախոսության հիմնական արդյունքները հրապարակվել են 7 գիտական աշխատություններում, որոնք թվարկված են սեղմագրի վերջում [1-7]:

## Աշխատանքի կառուցվածքը և ծավալը

Ատենախոսությունը բաղկացած է ներածությունից, 4 գլուխներից՝ իրենց ամփոփումներով, եզրակացությունից, գրականության ցանկից՝ իր 104 հղումներով: Ատենախոսության ընդհանուր ծավալը 109 էջ՝ 19 աղյուսակով և 26 պատկերներով:

### ԱՇԽԱՏԱՆՔԻ ՀԻՄՆԱԿԱՆ ԲՈՎԱՆԴԱԿՈՒԹՅՈՒՆԸ

Աշխատանքի ներածության մեջ հիմնավորվել է ատենախոսության թեմայի արդիականությունը, ներկայացվել են հետազոտության նպատակը և խնդիրները, գիտական նորույթը, պաշտպանությանը ներկայացվող հիմնական դրույթները, հետազոտության տեսական և գործնական նշանակությունն:

Ատենախոսության ստաջին գլխում նկարագրվել են համայնքների հայտնաբերման հիմնական հասկացությունները և գրականության մեջ առկա արդյունքները: Հատուկ հետազոտվել [1] և ներկայացվել է ինֆորմացիայի տեսության դերը համայնքների հայտնաբերման խնդիրներում: Բերվել է խորհրդատվական համակարգերի ընդհանուր նկարագիրը և կապը համայնքների հայտնաբերման հետ:

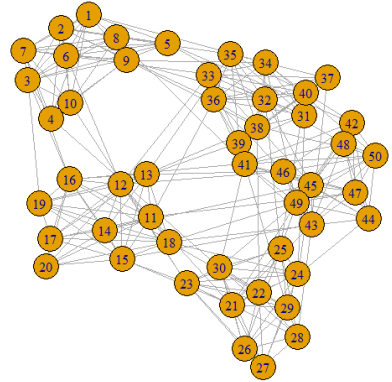
**Բաժին 1.1-ում** տրվել է բարդ ցանցերի համառոտ նկարագիրը, ձևակերպվել է համայնքների հայտնաբերման խնդրի կարևորությունը: Ներկայացվել է գրականության մեջ առկա արդյունքների հետազոտումը համայնքների հայտնաբերման ալգորիթմների ուսումնասիրման և համեմատման վերաբերյալ:

**Բաժին 1.2-ում** նկարագրվել են պատահական գրաֆների կառուցման մոդելները և դրանց կարևորությունը ցանցերի գիտության և համայնքների հայտնաբերման մեջ:

**Բաժին 1.3-ում** նկարագրվել է ստոխաստիկ բլոկ մոդելը, որն իրենից ներկայացնում է համայնքային կառուցվածքով օժտված ցանցերի գեներատիվ մոդել, հնարավորություն տալով ստեղծել իրական աշխարհի ցանցերին նման անսահմանափակ քանակությամբ ցանցեր՝ որպես մուտքային պարամետրեր տալով ցանցում գագաթների քանակը ( $N$ ), համայնքների քանակը ( $K$ ), համայնքներում գագաթների քանակը ( $N_{C_1}, \dots, N_{C_K}$ ) և համայնքներում ( $W_{ij}, i = j$ ) և նրանց միջև ( $W_{ij}, i \neq j$ ) գագաթները կապող կողերի հավանականությունների մատրիցը ( $W$ ): Նաև ներկայացված են ստոխաստիկ բլոկ մոդելում համայնքային կառուցվածքների ուժեղ, մասնակի և թույլ վերականգնման պահանջները: Նկար 1-ում պատկերված է ստոխաստիկ բլոկ մոդելի միջոցով ստեղծված ցանց (50 գագաթ, 5 համայնք), որտեղ  $W_{ij} = 0.95, i = j$  և  $W_{ij} = 0.05, i \neq j$ :



$W$	$C_1$	$C_2$	$C_3$	$C_4$	$C_5$
$C_1$	$W_{11}$	$W_{12}$	$W_{13}$	$W_{14}$	$W_{15}$
$C_2$	$W_{21}$	$W_{22}$	$W_{23}$	$W_{24}$	$W_{25}$
$C_3$	$W_{31}$	$W_{32}$	$W_{33}$	$W_{34}$	$W_{35}$
$C_4$	$W_{41}$	$W_{42}$	$W_{43}$	$W_{44}$	$W_{45}$
$C_5$	$W_{51}$	$W_{52}$	$W_{53}$	$W_{54}$	$W_{55}$



Նկար 1 – Ստոխաստիկ բլոկ մոդելի միջոցով ստեղծված ցանցի օրինակ:

**Բաժին 1.4-ում** ձևակերպվել են համայնքների հայտնաբերման կարևոր հասկացությունները, մասնավորապես, համայնքների գնահատման ֆունկցիաները, որոնք հնարավորություն են տալիս մոտավոր որոշելու համայնքների որակը:

**Բաժին 1.5-ում** նկարագրվել են ատենախոսության շրջանակներում օգտագործվող համայնքների հայտնաբերման հիմնական մոտեցումները, որոնք են հիերարխիկ կլաստերինգի ալգորիտի և բաժանարար ալգորիթները, ցանցի մոդուլյարության մաքսիմիզացիայի հիման վրա ստեղծված Fast greedy և Multilevel (Louvain), ինֆորմացիոն տեսական Infomap, պատահական քայլերի վրա հիմնված Walktrap, սպեկտրալ ալգորիթներից Leading eigenvector-ը, վիճակագրական մեխանիկայից Spinglass, գրաֆում կենտրոնական կողերի վրա հիմնված Edge betweenness և գագաթների պիտակավորման վրա հիմնված Label propagation ալգորիթները:

**Բաժին 1.6-ը** նվիրված է ինֆորմացիայի տեսության դերի հետազոտությանը համայնքների հայտնաբերման ոլորտում: Ներկայացված են ինֆորմացիայի տեսության հիմնական գաղափարները, չափերը, դրանց կիրառումը համայնքային կառուցվածքների համեմատման համար: Կարևորվել է համայնքների հայտնաբերման մեջ ինֆորմացիոն-տեսական չափերի բավարարումը մետրիկայի, նորմավորման և հաստատուն բազային<sup>4</sup> հատկություններին, ինչպես նաև ներկայացվել է այդ չափերի թերությունները:

<sup>4</sup> N. X. Vinh, J. Epps and J. Bailey, "Information theoretic measures for clusterings comparison: is a correction for chance necessary?," in *ICML '09 Proceedings of the 26th Annual International Conference on Machine Learning*, Montreal, Quebec, Canada, 2009.

**Բաժին 1.7-ում** ներկայացվել են համայնքային կառուցվածքների գնահատման և համեմատման հիմնական օգտագործվող ներքին և արտաքին չափերը, որոնք հիմնականում ներմուծվել են ինֆորմացիայի տեսության և կլաստերային վերլուծության ոլորտներից:

**Բաժին 1.8-ում** նկարագրվել են խորհրդատվական համակարգերի կազմակերպման հիմնական մոտեցումները:

Ատենախոսության **երկրորդ** գլխում ուսումնասիրվում, գնահատվում և համեմատվում են համայնքների հայտնաբերման ալգորիթմները, օգտագործելով բազմապիսի իրական աշխարհի փոքր և մեծ բարդ, ապա ստոխաստիկ բլոկ մոդելի միջոցով արհեստականորեն ստացված ցանցեր [2,3]:

**Բաժին 2.1-ը** նվիրված է վերը նշված համայնքների հայտնաբերման ալգորիթմների տարբեր տեսակի 1589 – 334,863 գազաթ պարունակող իրական աշխարհի փոքր և մեծ բարդ ցանցերի<sup>5</sup> վրա փորձարկումների նկարագրմանը:

Ալգորիթմների համեմատումը իրականացվել է գնահատելով ցանցերում դրանց հայտնաբերած համայնքային կառուցվածքների որակը, օգտագործելով ցանցի մոդուլյարությունը, conductance և triangle participation ratio համայնքների գնահատման ֆունկցիաները, արտաքին չափերից նորմավորված փոխադարձ ինֆորմացիան (NMI), ինֆորմացիայի փոփոխությանը (VI) և adjusted rand index-ը: Արդյունքում ստացվել է, որ ըստ մոդուլյարության ցուցանիշի Fast greedy և Louvain ալգորիթմները հայտնաբերել են լավ որակի համայնքներ բոլոր ցանցերի վրա: Սակայն տարբեր են դրանց հայտնաբերած համայնքների քանակները: Այսպիսով, կարելի է եզրակացնել, որ Fast greedy և Louvain ալգորիթմները հայտնաբերում են քիչ քանակությամբ, սակայն լավ որակի համայնքային կառուցվածքներ ի տարբերություն մնացած ալգորիթմների, որոնք հիմնականում հայտնաբերում են մեծ թվով համայնքներ պարունակող համայնքային կառուցվածքներ՝ փոքր մոդուլյարությամբ: Ըստ conductance և triangle participation ratio գնահատման ֆունկցիաների, որոնք ըստ տարբեր հեղինակների հանդիսանում են լավ որակի համայնքների գնահատման ֆունկցիաներ, Fast greedy, Louvain, Leading eigenvector և Label propagation ալգորիթմների հայտնաբերած համայնքները ունեն ցածր conductance և մեծ triangle participation ratio, այդպիսով՝ լավ որակ: NMI-ի միջոցով ստացված արդյունքները ցույց են տալիս, որ չորս ցանցերի վրա իրար նման համայնքային կառուցվածքներ են հայտնաբերել Infomap և Walktrap ալգորիթմները, իսկ չորս ալգորիթմների զույգեր եղել են ամենանմանը և՛ ըստ NMI-ի և՛ VI-ի:

---

<sup>5</sup> J. Leskovec and R. Soric, "SNAP: A General-Purpose Network Analysis and Graph-Mining Library," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 8, no. 1, 2016.

**Բաժին 2.2-ում** ալգորիթմները գնահատվել և համեմատվել են ստոխաստիկ բլոկ մոդելի միջոցով արհեստականորեն ստացված ցանցերի վրա:

Առաջին փուլում ստեղծվել են 200 գագաթ և 4 նույնաքանակ գագաթ պարունակող համայնքներով 2000 ցանցեր, որտեղ դիտարկվել է համայնքների ներսում գագաթների համապատասխանաբար նոսր ( $P_{in} = 0.25$ ) և խիտ ( $P_{in} = 0.75$ ) փոխազդեցություններ: Փոփոխելով համայնքների միջև կողերի հավանականությունը, ըստ մոդուլյարության ցուցանիշի գնահատվել է ալգորիթմների աշխատանքը ֆիքսված հավանականության համար միջինացնելով 1000 ցանցերի վրա ստացված արդյունքները: Արդյունքում ստացվել է, որ Spinglass, Louvain և Fast greedy ալգորիթմները համայնքների ներսում և՛ նոսր և՛ խիտ փոխազդեցությունների դեպքում, անկախ համայնքների միջև կողերի հավանականության արժեքից մնացած ալգորիթմների համեմատ հայտնաբերում են բարձր մոդուլյարությամբ համայնքային կառուցվածքներ:

Երկրորդ փուլում ալգորիթմները համեմատվել են ըստ հայտնաբերման ժամանակի (հաշվողական ժամանակ) տարբեր չափերի արհեստական ցանցերի օգտագործմամբ: Փորձարկումների արդյունքում պարզվել է, որ Spinglass և Edge betweenness ալգորիթմները մնացած ալգորիթմների համեմատ հարմարավետ չեն մեծ ցանցերում համայնքներ հայտնաբերելու համար:

Եվ վերջում ալգորիթմները համեմատվել են ըստ «ground-truth»-ի, այն է՝ ցանց, որի իրական համայնքային կառուցվածքը հայտնի է: Այդպիսի ցանցեր ստեղծվել են ստոխաստիկ բլոկ մոդելի միջոցով՝ ապահովելով համայնքների ներսում կողերի համեմատաբար ավելի մեծ հավանականություն քան համայնքների միջև: Օգտագործելով այդպիսի 1000 նմանօրինակ ցանցեր, դրանց վրա կիրառելով համայնքների հայտնաբերման ալգորիթմները, համեմատելով յուրաքանչյուրի հայտնաբերած համայնքային կառուցվածքը իրական կառուցվածքի հետ և ապա միջինացնելով բոլոր արդյունքները, ստացվել է, որ իրականին ամենամոտ համայնքային կառուցվածք հայտնաբերում է Walktrap ալգորիթմը:

Ամփոփելով ստացված արդյունքները և՛ իրական և՛ արհեստական ցանցերի վրա, աղյուսակ 1-ում ներկայացվում է ալգորիթմների առավելությունները և թերությունները ըստ տարբեր պարամետրերի:

Աղյուսակ 1 - Ալգորիթմների առավելությունները և թերությունները ըստ տարբեր պարամետրերի:

Ալգորիթմ	Բարդություն	Մոդուլյարություն	«Conducance»	«Triangle participation ratio»	Մեծ $P_{out}$ (մոդուլյարություն)	Մեծ $P_{out}$ (ծամանակ)	Իրական համայնքային կառուցվածքի հետ նմանություն
Fast greedy	+	++	++	+	+	++	+
Louvain	++	++	++	++	++	+	++
Infomap	+	++	--	-	--	--	--
Walktrap	-	++	-	+	--	+	++
Leading eigenvector	+	+	++	++	+	-	+
Label propagation	++	+	+	++	--	++	--
Spinglass	--	«չկա»	«չկա»	«չկա»	++	--	++
Edge betweenness	--	«չկա»	«չկա»	«չկա»	«չկա»	«չկա»	+

Ատենախոսության **կորոզ** գլուխը նվիրված է ինֆորմացիայի տեսության նոր մոտեցումների հետազոտմանը համայնքային կառուցվածքների համեմատման խնդրի լուծման նպատակով [4,5]:

**Բաժին 3.1-ում** դիտարկվում են  $f$ -տարամիտությունների հնարավոր կիրառությունը համայնքային կառուցվածքների համեմատման համար, որոնք ինֆորմացիայի տեսության և հավանականության տեսության մեջ ծառայում են որպես տարբեր հավանականային բաշխումների համեմատման չափեր:

**Բաժին 3.2-ում** նախորդ բաժնում ստացված հատկություններից ելնելով առաջարկվում է մոդիֆիկացված  $\chi^2$  –տարամիտություն:

**Սահմանում.** Երկու համայնքային կառուցվածքների՝  $X = (x_1, x_2, \dots, x_N)$ -ի և  $Y = (y_1, y_2, \dots, y_N)$ -ի միջև մոդիֆիկացված  $\chi^2$ -տարամիտությունը սահմանվում է

$$MD_{\chi^2}(X, Y) = 1 - \frac{\chi^2(P_{XY} \parallel P_X P_Y)}{K_{max}} = 1 - \frac{\sum_{x,y} \frac{p^2(x,y)}{p(x)p(y)} - 1}{K_{max} - 1},$$

որտեղ  $P_{XY}$ ,  $P_X$ , և  $P_Y$ -ը  $X$ -ի և  $Y$ -ի համապատասխանաբար համատեղ և եզրային հավանականային բաշխումներն են, իսկ  $K_{max} = \max\{K_X, K_Y\}$ -ը  $X$  և  $Y$  համայնքային կառուցվածքներում համայնքների քանակի մեծագույն արժեքը:

Ապացուցվել է **թեորեմ**, ըստ որի մոդիֆիկացված  $\chi^2$ -տարամիտությունը բավարարում է մետրիկայի բոլոր հատկություններին, բացառությամբ եռանկյան անհավասարմանը, և նորմավորված չափ է՝ ընդունելով արժեքներ  $[0,1]$  միջակայքում:

Փորձարկումների միջոցով ստուգվել է նաև դրա բավարարումը՝ հաստատուն բազային հատկությանը: Ստեղծելով բազմաթիվ արհեստական պատահական փոփոխականներ, որոնք կարելի է դիտարկել որպես համայնքային կառուցվածքներ, համեմատվել են տարբեր քանակությամբ գազաթներով և համայնքներով ավելի քան 100,000 ցանցեր: Ուսումնասիրությունները ցույց են տվել, որ ի տարբերություն լայնորեն կիրառվող նորմավորված ինֆորմացիայի փոփոխության և նորմավորված ինֆորմացիոն հեռավորության, մոդիֆիկացված  $\chi^2$ -տարամիտությունը չունի կախում ցանցում համայնքների քանակից և կարող է օգտագործվել այնպիսի դեպքերում, երբ ցանցերում գազաթների քանակը համեմատաբար փոքր է համայնքների քանակից:

Չափերը նաև համեմատվել են ստոխաստիկ բլոկ մոդելի միջոցով արհեստականորեն ստեղծված ցանցերում՝ ալգորիթմների հայտնաբերած համայնքային կառուցվածքները իրական համայնքային կառուցվածքների հետ համեմատելով:

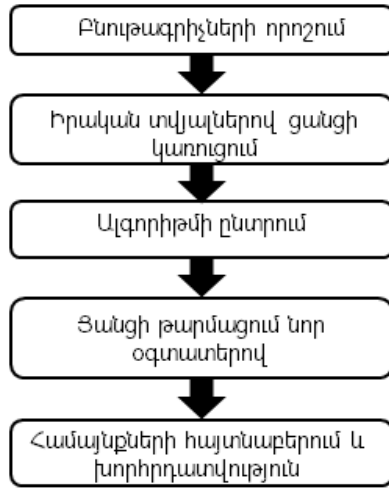
Ատենախոսության **չորրորդ** գլխում առաջարկվում է համայնքների հայտնաբերման մոտեցումների վրա հիմնված խորհրդատվական միջավայր, որը այլ մոտեցումների համեմատ ունի առավելություններ որոշ կիրառություններում [6,7]:

**Բաժին 4.1-ում** ներկայացված են խորհրդատվական համակարգերում առկա մոտեցումները, այն է՝ բովանդակության վրա հիմնված և համագործակցային ֆիլտրման մեթոդները, որոնք ժամանակակից խորհրդատվական համակարգերի հիմնական բաղադրիչ մասն են:

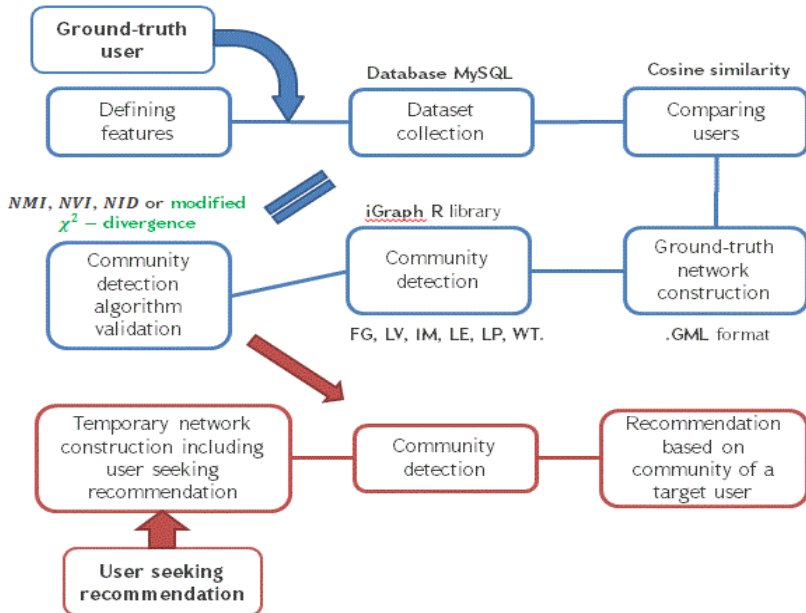
**Բաժին 4.2-ում** ներկայացվում է առաջարկվող միջավայրի իրականացման հինգ հիմնական քայլերը (տես. նկար 2):

**Բաժին 4.3-ում** նկարագրված է առաջարկված խորհրդատվական միջավայրի իրականացումը (տես. նկար 3):

**Բաժին 4.4-ում** նկարագրվում է միջավայրի փորձարկումը օրինակի վրա, մասնավորապես, հարցաշարի միջոցով դիմորդներին մասնագիտություններ առաջարկելու մեխանիզմը և արհեստական տվյալների օգտագործմամբ ցույց է տրվում դրա իրագործումը:



Սկար 2 - Միջավայրի իրականացման քայլերը:



Սկար 3 – Միջավայրի իրականացումը:

## Աշխատանքի հիմնական արդյունքները:

Ատենախոսության մեջ ստացվել են հետևյալ հիմնական արդյունքները.

1. Հետազոտվել և գիտականորեն հիմնավորվել են համայնքների հայտնաբերման խնդիրների և ինֆորմացիայի տեսության գործիքների միջև կապերը [1]:
2. Ուսումնասիրվել, գնահատվել և համեմատվել են համայնքների հայտնաբերման հիմնական ալգորիթմները ինչպես փոքր և մեծ իրական բարդ ցանցերի, այնպես էլ ստոխաստիկ բլոկ մոդելի միջոցով ստեղծված բազմաթիվ արհեստական ցանցերի վրա՝ օգտագործելով տարբեր ներքին և արտաքին չափման մեծություններ: Արդյունքները մանրամասն նկարագրում են տարբեր ցանցերում ալգորիթմների աշխատանքը, հնարավորություն տալով ընտրել համապատասխան ալգորիթմ տրված ցանցի դեպքում [2,3]:
3. Համայնքային կառուցվածքների համեմատման ինֆորմացիոն-տեսական չափերը փորձարկվել են երեք հիմնական հատկությունների հիման վրա. մետրիկա, նորմավորում և հաստատուն բազային հատկություն: Դիտարկվել են ինֆորմացիայի տեսությունից  $f$ -տարամիտությունների կիրառությունները համայնքների հայտնաբերման գնահատման խնդիրներում: Որպես արդյունք առաջարկվել է մոդիֆիկացված  $\chi^2$ -տարամիտությունը, որն առկա ինֆորմացիոն-տեսական չափերի համեմատ ունի առավելություն [4,5]:
4. Առաջարկվել է համայնքների հայտնաբերման վրա հիմնված նոր խորհրդատվական միջավայր: Միջավայրը հնարավորություն է տալիս կատարել խորհրդատվություններ որոշ կիրառություններում, որտեղ այլ խորհրդատվական մոտեցումները ձախողվում են: Ներկայացվել է նաև միջավայրի իրականացումը [6,7]:

## Հրատարակված աշխատությունների ցանկը:

- [1] M. Haroutunian, K. Mkhitarian, "Information-Theoretic Approach to Community Detection Problem", *Mathematical Problems of Computer Science*, vol. 47 pp. 50-60, 2017.
- [2] J. Mothe, K. Mkhitarian, M. Haroutunian, "Community detection: Comparison of state of the art algorithms", *Proc. of Intern. Conf. Computer science and information technologies*, pp. 252-256, 2017, Reprint in IEEE Revised selected papers, pp. 125-129, 2017.

- [3] J. Mothe, K. Mkhitarian and M. Haroutunian, "Detecting Communities from Networks: Comparison of Algorithms on Real and Synthetic Networks," *Journal of Information Theories and Applications*, pp. 53-73, 2019.
- [4] M. Haroutunian, K. Mkhitarian and J. Mothe, "Divergence Measures For Community Detection Evaluation," in *Proc. of Emil Artin International Conference*, pp. 71-73, Yerevan, Armenia, 2018.
- [5] M. Haroutunian, K. Mkhitarian and J. Mothe, " $f$ -Divergence Measures for Evaluation in Community Detection," in *Workshop on Collaborative Technologies and Data Science in Smart City Applications (CODASSCA)*, pp. 137-144, Yerevan, Armenia, 2018.
- [6] K. Mkhitarian, "Community Detection-Based Recommendation Framework," *Mathematical Problems of Computer Science*, vol. 50, pp. 61-66, 2018.
- [7] K. Mkhitarian, "Realization of Recommender Framework Based on Community Detection," *Mathematical Problems of Computer Science*, vol. 51, pp. 52-60, 2019.

## ИНФОРМАЦИОННО-ТЕОРЕТИЧЕСКОЕ ИССЛЕДОВАНИЕ ОБНАРУЖЕНИЯ СООБЩЕСТВ И РАЗРАБОТКА РЕКОМЕНДАТЕЛЬНОЙ СРЕДЫ

В последнее время компьютерная революция предоставила специалистам огромные данные и достаточные вычислительные ресурсы для обработки и анализа этих данных. Размер реальных сетей также значительно вырос, достигнув миллионов или даже миллиардов вершин и ребер. Изучение сложных сетей является молодой областью научных исследований, в значительной степени стимулируемой изучением сетей реального мира, таких как компьютерные сети и социальные сети.

Сложные сети реального мира обладают скрытой информацией, называемой сообществами или кластерами, которые состоят из узлов, тесно связанных внутри сообществ и слабо связанных между сообществами. Исследования сообществ имеют множество применений в различных науках, таких как информатика и машинное обучение, медицина и биология, экономика и социология. Цель обнаружения сообщества - разделить сеть на сообщества, чтобы извлечь из этой группы подгруппы плотно связанных узлов.

Основная цель диссертации состояла в том, чтобы предложить научно обоснованный подход к алгоритмам обнаружения сообщества и оценочным мерам в определенных приложениях. Для этой цели первая задача состояла в том, чтобы исследовать методы и алгоритмы, которые были разработаны для обнаружения сообщества, изучить существующие внутренние и внешние меры оценки обнаружения



сообщества, сравнить их на различных типах сетей, как в реальном мире сложных, так и синтетических. Следующая задача состояла в том, чтобы предложить прогресс в области сравнения алгоритмов обнаружения сообщества. Последняя задача состояла в том, чтобы изучить возможные взаимосвязи обнаружения сообществ и рекомендательных систем, преследуя цель предложить сравнительно лучшие подходы для рекомендаций.

Основные результаты работы:

- Исследованы и научно обоснованы связи между проблемами обнаружения сообщества и информационно теоретическими инструментами [1].
- Алгоритмы обнаружения сообщества были оценены и сравнены как на небольших, так и крупных реальных сложных сетях и синтетических сетях, сгенерированных из стохастической блочной модели с использованием различных внутренних и внешних метрик для оценки сообщества. Результаты показывают детальную картину того, как алгоритмы работают в сетях с различными свойствами, которые полезны для нахождения подходящего алгоритма для данной сети [2, 3].
- Информационно-теоретические меры для сравнения структур сообществ были протестированы на основе трех основных свойств; метрика, нормализация и постоянное базовое свойство. Кроме того, рассмотрено применение  $f$ -дивергенций из теории информации для оценки обнаружения сообществ. В результате предложена модифицированная  $\chi^2$ -дивергенция, которая по сравнению с существующими информационно-теоретическими мерами имеет преимущество [4, 5].
- Предложена новая рекомендательная среда, которая включает методы обнаружения сообщества в процесс принятия решений. Среда дает возможность делать рекомендации в определенных приложениях, где другие рекомендательные подходы терпят неудачу. Представлена также ее реализация [6, 7].

Представлено глубокое сравнение современных алгоритмов обнаружения сообщества, использующих как крупномасштабные сложные сети реального мира, так и искусственные сети, созданные на основе стохастической блочной модели. Применены меры внутренней оценки для оценки качества структуры сообщества и меры внешней

оценки для сравнения разделений. Эти результаты очень важны для понимания того, как работает конкретный алгоритм и какой алгоритм будет предпочтительным в определенном сценарии.

В контексте рекомендательных систем предложенная среда, разработанная с использованием методов обнаружения сообщества, может использоваться во многих реальных приложениях, где другие подходы рекомендаций бесполезны. Реализация платформы может быть использована для различных приложений.

## INFORMATION-THEORETIC INVESTIGATION OF COMMUNITY DETECTION AND RECOMMENDER FRAMEWORK DEVELOPMENT

In recent times, the computer revolution has provided specialists with massive data and sufficient computational resources to process and analyze these data. The size of real networks has also grown considerably, reaching millions or even billions of vertices and edges. The need to deal with such a large number of units has produced a deep change in the way graphs are approached. The study of complex networks is a young area of scientific research stimulated largely by the study of real world networks like computer networks and social networks.

Real world complex networks possess hidden information called communities or clusters, which are composed of nodes that are tightly connected within communities and weakly connected between communities. Investigation of communities proved to have countless applications in different sciences such as computer science and machine learning, medicine and biology, economics and sociology. The goal of community detection is to partition a network into communities to extract the subgroups of densely connected nodes from this network.

The main purpose of the dissertation was to suggest scientifically grounded approach for community detection algorithms and evaluation measures in certain applications. For this aim the first task was to investigate the methods and algorithms that have been developed for community detection, to study existing internal and external community detection evaluation measures, to compare them on different types of networks, both real world complex and synthetic. The next task was to propose an advance in the way of comparison of community detection algorithms. The final task was to study the possible interconnections of community detection and recommender systems pursuing an aim to suggest comparatively better recommendation approaches.

The main results of the work are the following:

- Connections between community detection problems and information theoretical tools are investigated and scientifically substantiated [1].
- Community detection algorithms were evaluated and compared on both small and large scale real world complex networks and synthetic networks generated from stochastic block model using various internal and external metrics for community evaluation. Results exhibit detailed picture on how algorithms work on networks with different properties that are useful to filter out an appropriate algorithm for a given network [2, 3].
- Information-theoretic measures for comparing network partitions or community structures were tested based on three main properties; metric, normalization and constant baseline properties. Additionally the application of  $f$ -divergences from information theory in community detection evaluation was observed. As a result, a modified  $\chi^2$ -divergence is suggested that compared with existing information-theoretic measures has an advantage [4, 5].
- New community detection based recommender framework is proposed that incorporates community detection techniques into decision making. The framework gives ability to make recommendations in certain applications where other recommender approaches fail. Its realization is also presented [6, 7].

A deep comparison of state-of-the-art community detection algorithms using both large scale real world complex networks and artificial networks generated from stochastic block model is reported. Internal evaluation measures to assess the quality of community structure and external evaluation measures to compare partitions are applied. These results are very important to understand how particular algorithm works and which algorithm would be preferable in certain scenario.

In the context of recommender systems, the newly proposed framework designed with the incorporation of community detection techniques, can be employed in many real life applications where other recommendation approaches are useless. The realization of the platform can be used for various applications.