JOURNAL OF SOUTHERN COMMUNICATION ENGINEERING
ISLAMIC AZAD UNIVERSITY BUSHEHR BRANCH

E-ISSN: 2980-9231
http://jce.iaubushehr.ac.ir

https://doi.org/10.30495/jce.2022.686992

Vol. 11/ No. 43/Spring 2022

Research Article

Detect Stable Associations in Dynamic Social Networks Using Influential Nodes

Taleb Khafaie, MSc [1] | Alireza Tavakoli Taraghi, Associate Professor [2] | Mehdi Hosseinzadeh, Associate Professor[3,4] | Ali Rezaee, Assistant Professor [5]

[1]Department of Computer Engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran, khafaietaleb@gmail.com

[2]Computer Science Group of Mathematics Department, Shahid Beheshti University, Tehran, Iran, alireza.tavakoli@gmail.com

[3]Mental Health Research Center, Psychosocial Health Research Institute, Iran University of Medical Sciences, Tehran, Iran,

[4]Computer Science, University of Human Development, Sulaymaniyeh, Irag, mehdi.hoza@gmail.com

[5]Department of Computer Engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran, alirezaee.uni@gmail.com

Correspondence
Alireza Tavakoli Taraghi, Computer Science Group of Mathematics Department, Shahid Beheshti University, Tehran, Iran,
Mail: alireza.tavakoli@gmail.com

Received: 18 May 2021
Revised: 29 October 2021
Accepted: 14 December 2021

Abstract

Today, social networks have found many applications in human daily life, so that identifying the behavior of members of this type of networks and the associations within them is of particular importance. Due to the structure and communication between members of social networks, some members within this type of network have more important roles than other members. In this study, a method for identifying more important associations was discussed. For this purpose, new features were introduced using network centralization features and then the importance of this type of features was investigated by Rough set theory. The experimental results showed that with increasing the number of popular nodes among an association introduced in this study and at the same time decreasing the amount of density, intermediate and proximity characteristics, the effect of the number of popular nodes on the association will remain more evident.

Keywords: Social Networks, Forums, Important Nodes, Sentrality Criteria, Sustainable forums

Highlights

- The extraction and detection of highly stable communities during the life of social networks.
- Introduced the leader nodes and popular nodes in this study.
- Predict popular communities using the rules derived from the rough theory.

Citation: T. khafaei, A. Tavakoli Taraghi, M. Hoseyn Zadeh, and A. Rezaee, "Detect Stable Community in Dynamic Social Networks Using Influential Nodes," *Journal of Southern Communication Engineering*, vol. 11, no. 43, pp. 39–48, 2022, doi: 10.30495/jce.2022.686992 (in Persian).

مقاله پژوهشی

تشخیص انجمن‌های پایدار در شبکه‌های اجتماعی پویا با استفاده از گره‌های بانفوذ

طالب خفـائی [۱] | علیرضا توکلی طرقی [۲*] | مهدی حسین زاده [۳و۴] | علی رضـائی [۵]

[۱]گروه مهندسی کامپیوتر، دانشکده فنی و مهندسی، دانشگاه آزاد اسلامی واحد علوم و تحقیقات، تهران، ایران، khafaietaleb@gmail.com

[۲]گروه علوم کامپیوتر، دانشکده ریاضیات، دانشگاه شهید بهشتی، تهران، ایران، alireza.tavakoli@gmail.com

[۳]مرکز تحقیقات سلامت روان، پژوهشکده سلامت روان اجتماعی، دانشگاه علوم پزشکی ایران، تهران، ایران

[۴]علوم کامپیوتر، دانشگاه توسعه انسانی، سلیمانیه، عراق، mehdi.hoza@gmail.com

[۵]گروه مهندسی کامپیوتر، دانشکده فنی و مهندسی، دانشگاه آزاد اسلامی واحد علوم و تحقیقات، تهران، ایران، alirezaee.uni@gmail.com

نویسنده مسئول

*علیرضا توکلی طرقی ، دانشیار، گروه علوم کامپیوتر، دانشکده ریاضیات، دانشگاه شهید بهشتی، تهران، ایران، alireza.tavakoli@gmail.com

تاریخ دریافت: ۲۸ اردیبهشت ۱۴۰۰
تاریخ بازنگری: ۷ آبان ۱۴۰۰
تاریخ پذیرش: ۲۳ آذر ۱۴۰۰

چکیده:

شبکه‌های اجتماعی امروزه کاربردهای بسیاری در زندگی روزمره انسان‌ها پیداکرده است به‌نحوی‌که شناسایی رفتار اعضای این نوع شبکه‌ها و انجمن‌های درون آن‌ها از اهمیت ویژه‌ای برخوردار شده است. با توجه به ساختار و نحوه ارتباط بین اعضای شبکه‌های اجتماعی برخی اعضای درون این نوع شبکه‌ها نقش‌های مهم‌تری نسبت به دیگر اعضا دارند. در این مطالعه روشی جهت تشخیص انجمن‌های بااهمیت بیشتر پرداخته شد. برای این منظور با استفاده از ویژگی‌های مرکزیت شبکه به معرفی ویژگی‌های جدیدی پرداخته شد و سپس اهمیت این نوع ویژگی‌ها توسط تئوری مجموعه‌های راف مورد بررسی قرار گرفت. نتایج آزمایش نشان داد که با افزایش تعداد گره‌های محبوب در بین یک انجمن که در این مطالعه معرفی شد و درعین‌حال کاهش مقدار ویژگی‌های تراکم، بینابینی و نزدیکی میزان تأثیر ویژگی تعداد گره‌های محبوب بر محبوب ماندن انجمن بیشتر مشهود خواهد بود.

کلید واژه‌ها: شبکه‌های اجتماعی، انجمن‌ها، گره‌های بااهمیت، معیار مرکزیت، انجمن‌های پایدار

https://doi.org/10.30495/jce.2022.686992

۱-مقدمه

امروزه شبکه‌های اجتماعی در دنیای مجازی بستری بسیار مناسب و گسترده‌ای برای ارتباط انسان‌ها را فراهم آورده است. ساختار ارتباطی بین افراد در شبکه‌های اجتماعی را می‌توان با ساختار گراف‌ها نمایش داد به شکلی که گره‌های درون گراف نمایانگر افراد و یال‌های بین گره‌ها نشان‌دهنده ارتباط بین آن‌ها در شبکه‌های اجتماعی است.

با گسترش کاربردهای شبکه‌های اجتماعی کاربران شبکه‌های اجتماعی به‌سرعت رو به افزایش است و مردم می‌توانند زندگی روزمره و ارتباطات اجتماعی خود را با دوست‌های خود در این بستر داشته باشند. به دلیل سهولت و در دسترس‌تر بودن این نوع شبکه‌های اجتماعی آنلاین جایگاه ویژه‌ای در زندگی مردم برای خود بازکرده است. همچنین این شبکه‌های اجتماعی دارای جذابیت و کاربرد بسیار زیادی شده‌اند به همین دلیل آنالیز شبکه‌های اجتماعی[1] کاربردهای بسیاری در زمینه‌های مختلف پیداکرده است. شبکه‌های اجتماعی آنلاین پویا ازجمله شبکه‌های اجتماعی است که در طول زمان ارتباطات جدیدی بین افراد ایجاد و یا شکسته می‌شود درواقع در بازه‌های زمانی متفاوت ارتباط بین افراد می‌تواند ایجاد و یا شکسته شود و منجر به تغییر ساختار شبکه‌های اجتماعی گردد. این تعاملات و ارتباطات بین کاربران باعث می‌شود در بخش‌هایی از شبکه‌های اجتماعی نسبت

[1] SNA

به سایر بخش‌ها تراکم ارتباطی بیشتری ایجاد گردد که هر یک از این بخش‌ها انجمن[1] نامیده می‌شود. ازآن‌رو که ارتباط بین افراد یک انجمن و یا سایر انجمن‌ها در طول زمان تغییر می‌کند می‌تواند منجر به تغییر ساختار انجمن‌ها گردد که این حالت را رخداد انجمن می‌نامیم. با توجه به اهمیت انجمن‌ها در شبکه‌های اجتماعی آنلاین در آنالیز شبکه‌های اجتماعی پویا به استخراج، دنبال کردن و شناسایی رخدادهای انجمن‌ها به‌منظور شناسایی رفتار این نوع شبکه‌ها پرداخته می‌شود. درواقع با مدل‌سازی شبکه‌های پویا می‌توان ساختار شبکه‌های اجتماعی آنلاین را در طول زمان موردبررسی قرارداد.

همان‌طور که می‌دانیم ارتباط بین کاربران در شبکه‌های اجتماعی آنلاین پویا بر اساس تعامل بین آن‌ها است. و تغییر ساختار در شبکه‌های اجتماعی آنلاین ناشی از تغییر لینک بین کاربران است که خود منجر به تغییر ساختار انجمن‌ها در این شبکه در طول زمان می‌گردد. ازاین‌رو، موضوع پیش‌بینی لینک در شبکه اجتماعی آنلاین کاربرد بسیار زیادی پیداکرده است و مطالعات موفق بسیار خوبی در این زمینه صورت گرفته است[۲,۱].

علاوه بر موضوع پیش‌بینی لینک مطالعات بسیاری در زمینه شناخت ساختار شبکه‌های اجتماعی آنلاین پویا و تجزیه و تحلیل آن‌ها صورت پذیرفته است. در همین راستا ازآنجاکه شناخت رفتار شبکه‌های اجتماعی و پیش‌بینی ساختار آینده این نوع شبکه‌ها کاربردهای بسیاری دارد، بسیاری از پژوهشگران به شناخت نوع رخدادها روی انجمن‌ها در شبکه‌های اجتماعی آنلاین پرداخته‌اند. در این دسته از مطالعات برخی از پژوهشگران انجمن‌های درون شبکه‌های اجتماعی را ایستا فرض کرده‌اند که منجر به نادیده گرفته شدن تأثیر ارتباط‌های موقت و ویژگی‌های پویا در تجزیه و تحلیل این نوع شبکه‌ها می‌شود. و اما دسته‌ی دیگری از پژوهشگران تغییر ساختار در شبکه‌های اجتماعی پویا به دنباله‌ای شامل بازه‌هایی از شبکه‌های ایستا نگاشت می‌شود به شکلی که انجمن‌ها در هر بازه زمانی مستقل از بازه‌های زمانی قبل و بعد از خود استخراج می‌شوند[۳, ۴]. که این نوع آنالیز به ما اجازه می‌دهد تغییر رفتار انجمن‌ها را با مقایسه ساختار آن‌ها در دو بازه زمانی پشت سر هم مورد ارزیابی قرار دهیم.

۲-روش‌های پیشین

تجزیه و تحلیل شبکه‌های اجتماعی امروزه با توجه به کاربردهای متنوع آن موردتوجه بسیاری از پژوهشگران قرارگرفته است. ارتباط افراد مختلف در شبکه‌های اجتماعی منجر به تراکم ارتباطی در برخی جاهای شبکه می‌شود که معمولاً برای سهولت تحلیل شبکه‌های اجتماعی آن‌ها را به بخش‌های مختلف که بانام انجمن شناخته می‌شوند تقسیم می‌کنند که مطالعات بسیاری در این زمینه به‌منظور ارائه الگوریتم‌های تشخیص انجمن در شبکه‌های اجتماعی وجود دارد. بدیهی است که اهمیت نقش افراد ازنظر ارتباط درون انجمن‌ها و مابین انجمن‌های مختلف از اهمیت بالایی در تجزیه و تحلیل شبکه‌های اجتماعی برخوردار است. به شکلی که زمینه‌های مطالعاتی مختلفی همچون شایعه پراکنی و نفوذ پذیری در شبکه‌های اجتماعی به وجود آمده است.

برخی مطالعات که نقش افراد ازنظر نفوذپذیری را موردبحث قرار داده‌اند توجه ویژه‌ای به ساختار انجمن‌ها در شبکه‌های اجتماعی نیز داشته‌اند. در [۵] روشی جهت شناسایی انجمن‌ها مبتنی بر گره‌های رهبر در شبکه‌های اجتماعی[2] را معرفی نموده است. در این مطالعه گره‌هایی که دارای بیشترین مقدار بردار ویژه[3] در ماتریس مجاورت هستند به‌عنوان گره‌های رهبر معرفی‌شده‌اند. پس از شناسایی گره‌های رهبر، به کمک تابع شباهت همسایه‌های با بیشترین درجه شباهت را شناسایی و گره‌های مشابه بر اساس یک مقدار آستانه معین به گره‌های رهبر شناسایی‌شده نگاشت داده می‌شوند تا انجمن‌ها را تشکیل دهند. در مطالعه دیگری تشخیص انجمن‌های محلی را بر اساس رتبه‌بندی گره‌های بااهمیت بالا[4] را پیشنهاد داده است. این الگوریتم پیشنهادی، یک شاخص جدید برای محاسبه اهمیت گره ارائه داده است. با توجه به محل شبکه، شاخص پیشنهادی می‌تواند به‌طور کامل اهمیت گره در میان تمام گره‌های در شبکه را منعکس کند. روش LCDR در ابتدا گره‌های مهم را برای گسترش جوامع اولیه بر

[1] Community
[2] Leader-Community Detection Approach-LCDA
[3] Eigenvalue
[4] local community detection based on high importance nodes Ranking -LCDR

اساس معیار شباهت محلی انتخاب می‌کند تا زمانی که همه گره‌ها عضو یکی از جوامع شوند. درنهایت، جوامع کشف‌شده را ادغام می‌کند تا ساختارهای جامعه نهایی را تشکیل دهد[۶].

تشخیص گره‌های تأثیرگذار بر انتشار اطلاعات در شبکه‌های اجتماعی از اهمیت ویژه‌ای برخوردار است. در [۷] نویسندگان، الگوریتم IEBoCI[1] را مبتنی بر نفوذپذیری انجمن ارائه کردند. فرض اصلی در این پژوهش این بود که هرچه توانایی انجمن‌های متصل شده توسط گره‌ها در انتشار اطلاعات بیشتر باشد و نفوذ گره‌ها در انجمن بیشتر باشد ، اهمیت گره‌ها نیز بیشتر است. این الگوریتم بدین گونه عمل می‌کند که ابتدا، احتمال فعال کردن گره‌های قابل‌دسترس به هر گره بر اساس مدل انتشار اطلاعات را محاسبه می‌کند. ثانیاً ، شبکه بر اساس الگوریتم LPA[2] به جوامع تقسیم می‌شود. ثالثاً ، تأثیر هر جامعه و میزان تأثیر گره‌ها در جوامع متصل را محاسبه می‌نماید. سرانجام اهمیت گره با ترکیب تأثیر جامعه و تأثیر گره در جامعه محاسبه می‌شود.

در سال‌های اخیر برای تجزیه و تحلیل نفوذپذیری اجتماعی در کاربردهای مختلف مسئله نفوذپذیری مورد مطالعه قرارگرفته است. هدف اصلی پیدا کردن حداقل گره‌ها در شبکه است که بیشترین تعداد گره‌ها در شبکه را فعال کند. اخیراً برخی پژوهشگران یک گام جلوتر حرکت نموده‌اند و میزان تأثیر گره‌ها بر جوامع متنوع را نیز در نظر گرفته‌اند. در [۸] پژوهشگران مشکل حداکثر سازی نفوذ[3] را برای یافتن مؤثر k گره بررسی نمودند. به‌طوری‌که اگر پیامی توسط آن k گره شروع و منتشر شود، تعداد و همچنین تنوع جامعه گره‌های فعال‌شده در پایان فرآیند انتشار به حداکثر برسد. همچنین در [۹] یک روش حداکثر سازی نفوذپذیری که می‌تواند گره‌های بانفوذ را از طریق ساختار جامعه شناسایی کند، پیشنهاد داده است. روش پیشنهادی عمدتاً از دو قسمت تشکیل‌شده است: تشخیص انجمن مبتنی بر تقسیم شبکه به انجمن‌های باکیفیت ارتباطی بالا و حداکثر نفوذ مبتنی بر ساختار انجمن.

مطالعات زیادی درزمینه به دست آوردن معیارها و ویژگی‌های مهم نفوذپذیری در انجمن‌ها موردبررسی قرارگرفته است. اخیراً پژوهش‌هایی جهت ترکیب نمودن معیارهای مختلف برای بهینه نمودن این نوع معیارها صورت پذیرفته است. در [۱۰] با در نظر گرفتن ویژگی‌های پخش[4] و هابنس[5] که هر یک نشان‌دهنده یک ویژگی مهم اما متفاوت از هر گره است. (مقدار هابنس قدرت هر رأس را در محیط اطراف خود منعکس می‌کند و مقدار پخش نشان‌دهنده پتانسیل گسترش است.) ویژگی IVI[6] را معرفی نمودند. بدیهی است که هرچه حاصل‌ضرب نمرات پخش و هابنس بیشتر باشد، آن گره در کل شبکه تأثیرگذارتر است. بدین ترتیب، به‌منظور ادغام امتیازهای پخش و هابنس و محاسبه اثر هم‌افزایی آن‌ها، از تابع ضرب استفاده شد و IVI تولید شد. به‌عبارت‌دیگر، IVI محصول هم‌افزایی مهم‌ترین اقدامات مرکزی، نیمه محلی و جهانی است به‌گونه‌ای که به‌طور هم‌زمان سوگیری‌های موقعیتی را حذف می‌کند. در مطالعه دیگری پژوهشگران موفق شدند روشی دقیق برای تعیین تأثیرگذارترین گره‌ها در شبکه‌های تصادفی با نگاشت مسئله بر روی نفوذ بهینه و حل مسئله بهینه‌سازی با الگوریتمی که آن را مجموعه تأثیرگذاری[7] نامیدند، ابداع کنند. آن‌ها دریافتند که تعداد گره‌های بانفوذ بهینه بسیار کمتر است و گره‌های درجه پایین می‌توانند نقش بسیار مهم‌تری نسبت به آنچه قبلاً تصور می‌شد در شبکه ایفا کنند[۱۱].

در [۱۲] نویسندگان نشان دادند که برخلاف باور عمومی، شرایط قابل قبولی وجود دارد که در آن بهترین پخش‌کننده‌ها با افراد بسیار مرتبط یا مرکزی‌ترین افراد مطابقت ندارند. در عوض، ما متوجه شدیم که کارآمدترین پخش‌کننده‌ها آن‌هایی هستند که در هسته شبکه قرار دارند که توسط تجزیه و تحلیل k-shell قطعه قطعه شده‌اند[۱۳-۱۵]. همچنین هنگامی‌که چندین پخش‌کننده به‌طور هم‌زمان در نظر گرفته شوند، فاصله بین آن‌ها به پارامتر مهمی تبدیل می‌شود که میزان گسترش را تعیین

[1] Importance Evaluation Algorithm Based on Community Influence
[2] Label Propagation Algorithm
[3] Community-diversified Influence Maximization -CDIM
[4] Spreading
[5] Hubness
[6] Integrated Value of Influence
[7] Collective Influence

می‌کند. با بررسی این کمیت در تعدادی از شبکه‌های واقعی توانستند بهترین پخش‌کننده‌های منفرد در شبکه را زمانی که انتشار از یک گره منشأ می‌گیرد، شناسایی نمایند. در مورد یک فرآیند پخش[1] که در بسیاری از گره‌ها به‌طور هم‌زمان منشأ می‌گیرد، نشان دادند که با در نظر گرفتن مبدأهای پخش واقع در فاصله مشخصی از یکدیگر، کارایی را بیشتر بهبود دهند.

در مطالعات پیشین توجه کمتری به انجمن‌های با پایداری بالا در طول زمان شده است. درصورتی‌که انجمن‌های با پایداری بالا که توانسته‌اند در طول زمان حیات شبکه به فعالیت خود ادامه دهند و ساختار آن‌ها تجزیه نشود می‌توانند نقش مهمی در شبکه ایفا نمایند. همان‌طور که در مطالعات مختلف گره‌های بانفوذ و نفوذپذیر موردبررسی قرارگرفته‌اند این نوع انجمن‌ها نیز با توجه به ساختار محکم آن‌ها که منجر به ادامه حیات آن‌ها می‌شود مطمئناً شامل گره‌های با اهمیت بالا برای شبکه هستند ازاین‌رو می‌توانند نقش بسزایی در انتشار اطلاعات در شبکه ایفا نمایند. به همین دلیل در این مطالعه در گام اول: به استخراج و تشخیص انجمن‌های با پایداری بالا در طول حیات شبکه‌های اجتماعی مورد بررسی قرار گرفت که در همین راستا به تعریف انجمن‌های محبوب پرداختیم. در گام دوم: جهت بررسی ساختار انجمن‌های محبوب و وجه تمایز آن‌ها از دیگر انجمن‌ها به معرفی چند ویژگی ساختاری در کنار دیگر ویژگی‌های مرکزیت پرداخته شد که درنهایت گره‌های رهبر و گره‌های محبوب در این مطالعه معرفی شدند. در گام سوم: با استفاده از تئوری مجموعه‌های راف به بررسی اهمیت ویژگی‌های ساختاری معرفی‌شده در کنار دیگر ویژگی‌های مرکزیت جهت پیش‌بینی انجمن‌های محبوب با استفاده از قوانین مستخرج از این تئوری پرداختیم.

۳-روش پیشنهادی

جهت تحلیل رفتار انجمن‌ها در شبکه‌های اجتماعی آنلاین ، فعالیت شبکه را به بازه‌های زمانی با طول یکسان تقسیم می‌کنیم و انجمن‌های بازه‌ی زمانی i را به شکل $C_1\,t[i]\,,\,C_2\,t[i]\,,\ldots,\,C_n\,t[i]$ درحالی‌که n تعداد انجمن‌ها در بازه $t[i]$ و انجمن‌های بازه‌ی زمانی $i+1$ را به شکل $C_1\,t[i+1]\,,\,C_2\,t[i+1]\,,\ldots,\,C_m\,t[i+1]$ درحالی‌که m تعداد انجمن‌ها در بازه $t[i+1]$ است، نمایش می‌دهیم. انجمنی از بازه $t[i+1]$ را که حداقل ۴۵٪ از اعضای آن با یکی از انجمن‌های بازه زمانی قبلی ($t[i]$) مشترک باشد به‌عنوان انجمن محبوب در آن شبکه‌ی اجتماعی در نظر می‌گیریم. . شناسایی و پیش‌بینی انجمن‌های محبوب در تحلیل شبکه‌های اجتماعی از اهمیت ویژه‌ای برخوردار است چراکه می‌تواند کاربردهای بسیار مهمی در زمینه‌های مختلف ازجمله سیستم‌های توصیه‌گر داشته باشد. درواقع بدیهی است که چنین انجمن‌هایی به دلیل ارتباط قوی و منسجم بین اعضای آن‌ها به حیات خود در بازه زمانی بعدی ادامه داده‌اند. جهت پیش‌بینی انجمن‌های محبوب در شبکه‌های اجتماعی آنلاین نیاز به بررسی ویژگی‌های مبتنی بر انسجام و نزدیکی انجمن‌ها و پیدا کردن ارتباط بین این ویژگی‌ها با نوع انجمن‌ها(ازنظر محبوب بودن یا نبودن) خواهیم داشت. به همین منظور برای یادگیری مدل رفتاری انجمن‌ها سعی نمودیم از ویژگی‌هایی استفاده کنیم که عمدتاً معیارهای مبتنی بر انسجام بین اعضای انجمن باشند. درواقع ابتدا انجمن‌ها در بازه $t[i]$ را استخراج و ویژگی‌های ذکرشده برای هر یک محاسبه می‌گردد، سپس با مقایسه انجمن‌های استخراج‌شده در این بازه با بازه زمانی بعدی ($t[i+1]$) به شناسایی نوع انجمن ازنظر محبوب بودن یا نبودن پرداخته می‌شود. با تکرار این کار روی همه جفت بازه‌های زمانی متوالی دنباله‌ای شامل ویژگی‌های هر انجمن در آن بازه زمانی و نوع انجمن ایجاد می‌شود که به کمک تئوری مجموعه‌های راف، به استخراج قوانین روی آن دنباله‌ها پرداخته می‌شود. سپس از این قوانین جهت یادگیری مدل و پیش‌بینی نوع انجمن‌ها در شبکه مورداستفاده قرار می‌گیرد.

۳-۱-ویژگی‌های مرکزیت و انسجام انجمن

ازآنجاکه در این مطالعه به دنبال شناخت رفتار انجمن‌های محبوب و درنهایت پیش‌بینی این نوع انجمن‌ها در بازه‌های زمانی بعدی هستیم. بهتر آن است که به بررسی دلیل محبوبت انجمن‌ها(ادامه حیات حداقل ۴۵٪ از اعضای انجمن در بازه زمانی بعدی)

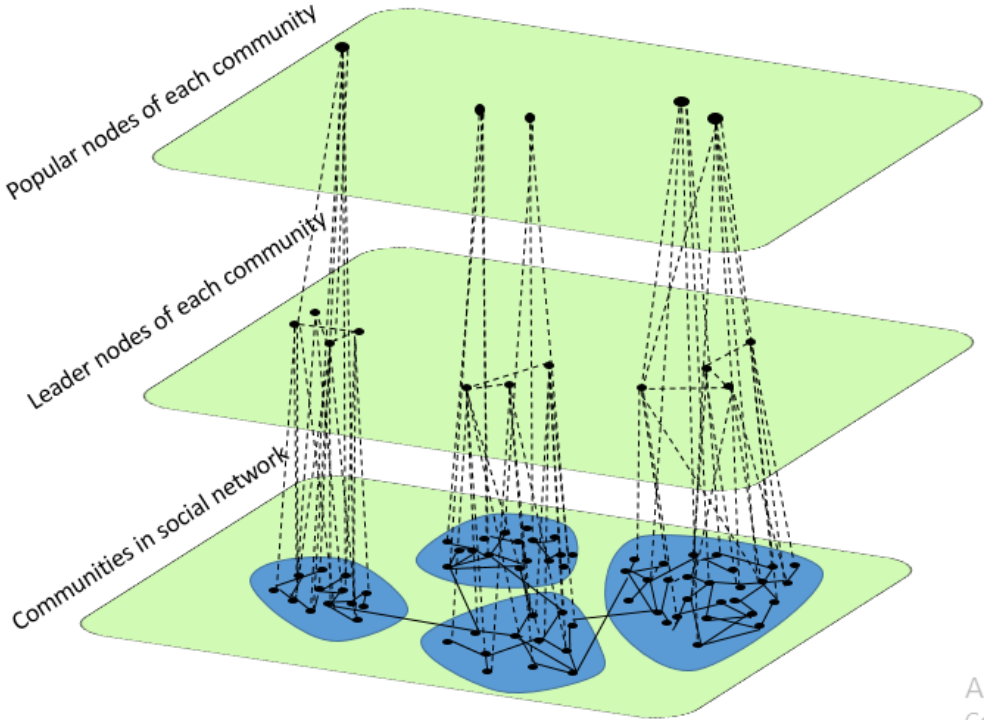
[1] Spreading Process

بپردازیم که یکی از راهکارها بررسی ویژگی‌های مرکزیت و انسجام انجمن‌ها است. ازاین‌رو در این مطالعه از ویژگی‌های گره‌های رهبر، گره‌های محبوب و معیارهای نزدیکی، بینابینی و مرکزیت استفاده‌شده است. همان‌طور که مشاهده می‌کنید علاوه بر ویژگی‌های مرکزیت، نزدیکی و بینابینی از یک ویژگی کمی که در این مطالعه معرفی‌شده است و نشان‌دهنده تعداد گره‌های محبوب درون یک انجمن است، نیز استفاده‌شده است. در ادامه تعریف گره‌های محبوب آورده شده است.

۳-۲-گره‌های محبوب

انجمن‌های درون شبکه‌های اجتماعی متشکل از اعضای مختلفی هستند که با توجه به نحوه اتصال آن‌ها با دیگر اعضای آن انجمن از ویژگی‌های مرکزیت متفاوتی برخوردار هستند. با توجه به این موضوع اعضای یک انجمن می‌توانند اهمیت متفاوتی داشته باشند و نقش کم‌رنگ‌تر و یا پررنگ‌تری در یک انجمن ایفا نمایند. ازاین‌رو در این مطالعه با استفاده از ویژگی‌های بردار ویژه که نشان‌دهنده میزان ارتباط یک عضو با اعضای مهم دیگر آن انجمن و بینابینی که نشان‌دهنده میزان بینابینی یک عضو در میان اعضای دیگر(قرار گرفتن در کوتاه‌ترین مسیر بین اعضای انجمن) است، عضوهای رهبر و محبوب در یک انجمن را تعریف نمودیم.

اعضایی از آن انجمن را که دارای مقدار بردار ویژه بزرگ‌تر از میانه آن انجمن بوده‌اند را به‌عنوان اعضای رهبر فرض نمودیم. همچنین از میان اعضای رهبر آن دسته از اعضایی که دارای مقدار بینابینی بالاتر از میانگین باشند را به‌عنوان اعضای محبوب فرض نمودیم. شکل ۱ شمایی انتزاعی از شبکه اجتماعی را نشان می‌دهد، که در شکل ۱ در یک سطح بالاتر از دیگر اعضا (لایه ۲) اعضای رهبر و در لایه بالاتر (لایه ۳) اعضای محبوب نشان داده‌شده است. در این مطالعه تعداد اعضای محبوب هر انجمن به‌عنوان یک ویژگی کمی از انجمن‌ها معرفی‌شده است.



شکل ۱: شمایی از شبکه اجتماعی در سه لایه انتزاعی

۳-۳-آزمایش و پیاده‌سازی

در این مطالعه، ما کارایی مدل پیشنهادی خود را بر روی یک دیتاست واقعی از شبکه‌های اجتماعی آنلاین مورد آزمایش قراردادیم. دیتاست FB که حاوی یک لیست از تمام پست‌های روی دیوار کاربران فیس بوک نیو اورلئال است و در این مطالعه از مدت زمان ۲۳ هفته از تاریخ ۳-۱۵-۲۰۰۷ تا ۸-۲۲-۲۰۰۷ استفاده‌شده است، که درمجموع شامل ۱۴۰۵۹۶ یال موقت در این فاصله زمانی است. در این مقاله دیتاست جهت آنالیز، به بازه‌های زمانی به طول ۱۲ روز تقسیم شد و سپس به ازای هر انجمن در انتقال از بازه‌ی $t[i]$ به $t[i+1]$ دنباله‌ای شامل ویژگی‌های آن انجمن در بازه‌ی $t[i]$ و برچسب‌های باینری نوع آن انجمن(محبوب بودن یا نبودن) در این انتقال، تشکیل‌شده است.

در این مطالعه از تئوری مجموعه راف برای محاسبه بی زائده داده‌ها و استخراج قوانین استفاده‌شده است. بدین منظور از ابزار Rosetta برای تولید بی زائده از مجموعه ویژگی‌های مرکزیت، نزدیکی و بینابینی استفاده شد. در محیط Rosetta از الگوریتم ژنتیک برای محاسبه بی زائده‌ها استفاده شد.

۳-۴-آزمایش روی دیتاست فیس بوک

همان‌طور که در جدول ۲ مشاهده می‌کنید ویژگی‌ها قبل از استفاده در یادگیری مدل به بازه‌های گسسته نگاشت شده‌اند، بدین ترتیب که pop node به دو گروه H و M (به ترتیب بازه با مقادیر بزرگ‌تر و کوچک‌تر روی میانه) و ویژگی‌ها بینابینی، نزدیکی و تراکم به سه بازه H، L وM با مقادیر مشخص‌شده در جدول نگاشت شده‌اند. ویژگی Com Centralization را بر اساس توزیع آنتروپی روی مقادیر آن گسسته نمودیم که بالای ۷۷٪ از داده‌ها در بازه (۰.۹۸۸۷۶, *] و تقریباً ۱۶٪ از داده‌ها در بازه [۰.۹۹۵۲۳, ۰.۹۸۸۷۶) نگاشت شدند. ۱۶٪ باقی‌مانده داده‌ها نیز در ۱۲ بازه متفاوت دیگر نگاشت شدند. آزمایش بر روی دیتاست FB صورت گرفت. در این آزمایش، ۵۲۳ انجمن از بازه‌های زمانی به طول ۱۲ روز استخراج و مورد ارزیابی قرار گرفت.

جدول ۲، تعداد ۱۲ قانون با LHS بیشتر از میان قوانین مستخرج از بی زائده‌ها روی دیتاست FB به کمک تئوری مجموعه‌های راف را نشان می‌دهد. با توجه به ماهیت تئوری مجموعه‌های راف و همان‌طور که در جدول۳ مشاهده می‌کنید قوانین به دودسته قطعی و غیرقطعی تقسیم می‌شوند. قوانین شماره ۷ و ۱۲ که دارای دقت RHS با مقدار ۱ هستند از دسته قوانین قطعی و دیگر قوانین غیرقطعی هستند. اگر قوانین با مقدار دقت RHS بالاتر از ۹۰٪ را در گروه قوانین قطعی فرض بگیریم، قوانین ۱، ۶ و ۱۰ نیز در این گروه قرار خواهند گرفت. از بین قوانین ۱،۶،۷،۱۰و ۱۲ همان‌طور که مشاهده می‌کنید، قوانین ۱،۷و۱۲ نشان‌دهنده روابط بین ویژگی‌های با مقادیری است که منجر به تشخیص انجمن در دسته انجمن‌های معمولی شده است. همچنین قوانین ۶ و ۱۰ نشان‌دهنده روابط بین ویژگی‌ها با مقادیری است که منجر به تشخیص انجمن در دسته انجمن‌های بااهمیت شده است.

با مقایسه قوانین ۶ و ۱۰ می‌توان ویژگی‌های این قوانین را به دو زیرمجموعه: (۱) ویژگی‌های Pop_Node (H) ، Com_Cenralization و Com_Betweenness که در مقادیر شباهت دارند و (۲) ویژگی‌های COM_Closeness و Com_Density که در مقادیر تفاوت دارند، تقسیم کرد. از بین ویژگی‌های (۱)، ویژگی‌های Pop_Node و Com_Cenralization بیشترین مقادیر و ویژگی Com_Betweenness کمترین مقدار در میان بازه‌های تعریف‌شده برای هر یک از این ویژگی‌ها را به خود اختصاص داده‌اند. و از بین ویژگی‌های (۲)، مقدار Com_Closeness در قانون ۶ بالاتر از قانون ۱۰ و بالعکس مقدار Com_Densityدر قانون ۶ کمتر از قانون ۱۰ است و نشان‌دهنده این است که برای اینکه یک انجمن در کلاس انجمن محبوب قرار گیرد باید علاوه بر داشتن شرایط (۱) درصورتی‌که مقدار ویژگی Com_Density کم باشد نباید مقدار ویژگی Com_Closeness زیاد باشد و نشان‌دهنده این امر است که برای اینکه در بازه زمانی بعدی بیش از ۴۵٪ اعضای آن انجمن بتوانند ادامه حیات داشته باشند و در انجمنی از بازه زمانی بعدی، حضور پیدا کنند علاوه بر کم بودن میزان تراکم انجمن باید مقدار معیار نزدیکی در بین اعضای آن انجمن نیز کم باشد.

جدول ۱: ویژگی‌های استفاده‌شده در تجزیه و تحلیل شبکه اجتماعی

Pope node 15			H 304 58.126194 L 219 41.873806
Com closeness			
L	M	H	H 142 27.151052 M 381 72.848946
0.096467	0.192934	0.2894	
Com Centralization			[*, 0.98876) 405 77.437859 [0.98876, 0.99523) 81 15.487572 [0.99523, 0.99592) 7 1.338432 [0.99592, 0.99604) 1 0.191205 [0.99604, 0.99650) 6 1.147228 [0.99650, 0.99662) 1 0.191205 [0.99662, 0.99677) 2 0.382409 [0.99677, 0.99695) 1 0.191205 [0.99695, 0.99794) 7 1.338432 [0.99794, 0.99827) 1 0.191205 [0.99827, 0.99886) 4 0.764818 [0.99886, 0.99929) 2 0.382409 [0.99929, 1.00766) 3 0.573614 [1.00766, *) 2 0.382409
Com Betweenness			
L	M	H	H 90 17.208412 L 68 13.001912 M 365 69.789673
0.044933	0.089866	0.1348	
Com Density			
L	M	H	H 23 4.397706 L 188 35.946465 M 312 59.65583
0.0598	0.1196	0.1794	
Pop Community			0 384 73.422562 1 139 26.577438

با مقایسه قوانین (۱) و (۲) که تنها در مقدار ویژگی Pop_Node تفاوت دارند به‌راحتی قابل‌مشاهده است که با افزایش مقدار این ویژگی از میزان قطعیت قانون برای کلاس ۰ کاسته و تمایل به کلاس ۱ افزایش می‌یابد. همچنین با مقایسه قوانین (۲) و (۳) که تنها در مقدار ویژگی Com_Density تفاوت دارند. مشاهده می‌شود که درصورتی‌که مقدار Com_Density کمتر شود ولی همچنان تعداد گره‌های محبوب زیاد باشد بازهم از میزان قطعیت قانون برای کلاس ۰ کاسته و تمایل قانون به کلاس۱ افزایش می‌یابد. نشان‌دهنده این است که اگر مقدار ویژگی تراکم انجمن کاهش یابد ولی تعداد گره‌های محبوب همچنان زیاد باشد تمایل انجمن به‌قرار گرفتن در کلاس انجمن‌های محبوب افزایش می‌یابد. اگر این مقایسه را ادامه دهیم و این بار قانون (۳) و (۱۱) که تنها در مقدار ویژگیCom_Betweenness با یکدیگر تفاوت دارند را در نظر بگیریم. مشاهده می‌کنیم که با کمتر شدن مقدار این ویژگی (در مقایسه قوانین (۳)و(۶) همچنان از میزان قطعیت قانون برای کلاس ۰ کاسته و تمایل به کلاس ۱ افزایش می‌یابد، به‌طوری‌که حتی در قانون ۶ میزان قطعیت این قانون برای کلاس ۱ بیشتر از کلاس ۰ شده است (حدوداً ۹۵٪ برای کلاس ۱ و ۵٪ برای کلاس ۰). برای درک بهتر این موضوع با توجه به شکل۲ با تغییر مقادیر ویژگی‌های Pop_Node (از قانون ۱ به ۲) ، Density (از قانون ۲ به ۳) و Betweenness (از قانون ۳ به ۶) به همین ترتیب به میزان قطعیت قوانین کلاس ۱ افزوده می‌شود.

برای درک بهتر تأثیر ویژگی‌های نزدیکی و مرکزیت در تعیین میزان قطعیت قوانین برای کلاس‌های ۰و۱ به مقایسه قوانین۶ با۱۱ که در مقدار ویژگی نزدیکی متفاوت هستند و ۱ با ۱۲ که در مقدار ویژگی مرکزیت متفاوت هستند می‌پردازیم. با افزایش مقدار ویژگی نزدیکی از مقدار M در قانون ۶ به مقدار H در قانون ۱۱ ، زمانی که ویژگی Pop_Node دارای مقدار H باشد، میزان قطعیت قانون برای کلاس ۱ کاهش می‌یابد، به‌طوری‌که میزان قطعیت قانون ۱۱ برای کلاس ۱ از ۹۶٪ به ۸۳٪ کاهش‌یافته است. با مقایسه قوانین ۱ با ۱۲ متوجه می‌شویم که در حالتی که مقدار Pop_Node کم باشد ولی مقدار مرکزیت بیشتر گردد

به میزان قطعیت قانون برای کلاس ۰ افزوده می‌شود. این امر نشان‌دهنده این موضوع است که هرچند که میزان مرکزیت یک انجمن افزایش یابد ولی اگر دارایpop_nodeهای کمی باشد نمی‌تواند به‌عنوان یک انجمن بااهمیت شناخته شود و در بازه‌ی زمانی بعدی کمتر از ۴۵٪ افراد آن انجمن (با توجه به تعریف انجمن بااهمیت) حضور خواهند داشت.

جدول ۲: برخی قوانین مستخرج از تئوری راف

1	Pop_Node(L) AND COM_Closeness(M) AND COM_Centralization([*, 0.98876)) AND COM_Betweenness(M) AND COM_DENSITY(M) => Pop_COM(0) OR Pop_COM(1)	94	87, 7	0.925532, 0.074468
2	Pop_Node(H) AND COM_Closeness(M) AND COM_Centralization([*, 0.98876)) AND COM_Betweenness(M) AND COM_DENSITY(M) => Pop_COM(0) OR Pop_COM(1)	79	63, 16	0.797468, 0.202532
3	Pop_Node(H) AND COM_Closeness(M) AND COM_Centralization([*, 0.98876)) AND COM_Betweenness(M) AND COM_DENSITY(L) => Pop_COM(0) OR Pop_COM(1)	72	47, 25	0.652778, 0.347222
4	Pop_Node(L) AND COM_Closeness(H) AND COM_Centralization([*, 0.98876)) AND COM_Betweenness(M) AND COM_DENSITY(M) => Pop_COM(1) OR Pop_COM(0)	38	8, 30	0.210526, 0.789474
5	Pop_Node(H) AND COM_Closeness(H) AND COM_Centralization([*, 0.98876)) AND COM_Betweenness(M) AND COM_DENSITY(M) => Pop_COM(1) OR Pop_COM(0)	23	10, 13	0.434783, 0.565217
6	Pop_Node(H) AND COM_Closeness(M) AND COM_Centralization([*, 0.98876)) AND COM_Betweenness(L) AND COM_DENSITY(L) => Pop_COM(1) OR Pop_COM(0)	22	21, 1	0.954545, 0.045455
7	Pop_Node(L) AND COM_Closeness(M) AND COM_Centralization([*, 0.98876)) AND COM_Betweenness(M) AND COM_DENSITY(L) => Pop_COM(0)	21	21	1
8	Pop_Node(H) AND COM_Closeness(M) AND COM_Centralization([0.98876, 0.99523)) AND COM_Betweenness(M) AND COM_DENSITY(L) => Pop_COM(1) OR Pop_COM(0)	18	4, 14	0.222222, 0.777778
9	Pop_Node(H) AND COM_Closeness(M) AND COM_Centralization([0.98876, 0.99523)) AND COM_Betweenness(M) AND COM_DENSITY(M) => Pop_COM(0) OR Pop_COM(1)	17	15, 2	0.882353, 0.117647
10	Pop_Node(H) AND COM_Closeness(H) AND COM_Centralization([*, 0.98876)) AND COM_Betweenness(L) AND COM_DENSITY(M) => Pop_COM(1) OR Pop_COM(0)	12	11, 1	0.916667, 0.083333
11	Pop_Node(H) AND COM_Closeness(H) AND COM_Centralization([*, 0.98876)) AND COM_Betweenness(L) AND COM_DENSITY(L) => Pop_COM(0) OR Pop_COM(1)	12	2, 10	0.166667, 0.833333
12	Pop_Node(L) AND COM_Closeness(M) AND COM_Centralization([0.98876, 0.99523)) AND COM_Betweenness(M) AND COM_DENSITY(M) => Pop_COM(0)	10	10	1

۴-بحث و نتیجه‌گیری

در این مطالعه روشی جهت تشخیص انجمن‌های با پایداری بالا در طول حیات شبکه‌های اجتماعی به نام انجمن‌های محبوب معرفی شد سپس جهت بررسی ساختار انجمن‌های محبوب و وجه تمایز آن‌ها از دیگر انجمن‌ها به معرفی چند ویژگی ساختاری در کنار دیگر ویژگی‌های مرکزیت پرداخته شد که درنهایت گره‌های رهبر و گره‌های محبوب در این مطالعه معرفی شدند. با استفاده از تئوری مجموعه‌های راف به بررسی اهمیت ویژگی‌های ساختاری در کنار دیگر ویژگی‌های مرکزیت جهت پیش‌بینی انجمن‌های محبوب با استفاده از قوانین مستخرج از این تئوری پرداخته شد.

نتایج حاصل از آزمایش نشان داد که با افزایش تعداد گره‌های محبوب و درعین‌حال کاهش مقدار ویژگی‌های تراکم، بینابینی و نزدیکی میزان تأثیر ویژگی تعداد گره‌های محبوب بر محبوب ماندن انجمن بیشتر نمایان می‌شود و نشان‌دهنده این موضوع است که درصورتی‌که انجمن بتواند دارای تعداد گره‌های محبوب قابل‌توجهی باشد این گره‌ها می‌توانند باعث محبوب ماندن انجمن گردند.

پیشنهاد می‌شود که برای کارهای آینده ویژگی‌های جدیدی از انجمن‌های بااهمیت برای شناسایی انجمن‌های محبوب، همچنین میزان نفوذپذیری این نوع انجمن‌ها در شبکه‌های اجتماعی جهت انتشار اطلاعات موردبررسی قرار گیرند.

مراجع

[1] V. Martínez, F. Berzal, and J.-C. Cubero, "A Survey of Link Prediction in Complex Networks," *ACM Computing Surveys (CSUR)*, vol. 49, no. 4, pp. 69, 2016, doi:10.1145/3012704.

[2] P. Symeonidis, E. Tiakas, and Y. Manolopoulos, "Transitive node similarity for link prediction in social networks with positive and negative links," *Proceedings of the ACM Conference on Recommender Systems*, 2010, pp. 183-190, ,doi:10.1145/1864708.1864744.

[3] B. Gliwa, P. Bródka, A. Zygmunt, S. Saganowski, P. Kazienko and J. Koźlak, "Different approaches to community evolution prediction in blogosphere," *IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 2013, pp. 1291-1298, doi: 10.1145/2492517.2500231.

[4] K. S. Xu, and A .O. Hero, "Dynamic stochastic blockmodels for time-evolving social networks," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 4, pp. 552-562, 2014, doi: 10.1109/JSTSP.2014.2310294.

[5] S. Ahajjam, M. El Haddad, and H. Badir, "A new scalable leader-community detection approach for community detection in social networks," *Social Networks*, vol. 54, pp. 41-49, 2018, doi:10.1016/j.socnet.2017.11.004.

[6] S. Aghaalizadeh, S. T. Afshord, A. Bouyer, and B. Anari, "A three-stage algorithm for local community detection based on the high node importance ranking in social networks," *Physica A: Statistical Mechanics and its Applications*, vol. 563, pp. 125420, 2021, doi:10.1016/j.physa.2020.125420.

[7] G. He, J. Luo, and M. Yin, "An Evaluation Algorithm of the Importance of Network Node Based on Community Influence." *International Conference DMBD*, 2020, pp. 57-70, doi:10.1007/978-981-15-7205-0_6.

[8] J. Li, T. Cai, K .Deng, X. Wang, T. Sellis, and F. Xia, "Community-diversified influence maximization in social networks," *Information Systems*, vol. 92, pp. 101522, 2020, doi:10.1016/j.is.2020.101522.

[9] Z. Zhang, X. Li, and C. Gan, "Identifying influential nodes in social networks via community structure and influence distribution difference," *Digital Communications and Networks*, vol. 7, no. 1, pp. 131-139, 2021, doi:10.1016/j.dcan.2020.04.011.

[10] A. Salavaty, M. Ramialison, and P. D. Currie, "Integrated value of influence: an integrative method for the identification of the most influential nodes within networks," *Patterns*, vol. 1, no. 5, pp. 100052, 2020, doi:10.1016/j.patter.2020.100052.

[11] F. Morone, B. Min, L. Bo, R. Mari, and H. A. Makse, "Collective influence algorithm to find influencers via optimal percolation in massively large social media," *Scientific reports*, vol. 6, no. 1, pp. 1-11, 2016, doi: 10.1038/srep30062.

[12] M. Kitsak, L. K. Gallos, S. Havlin, F. Liljeros, L. Muchnik, H. E. Stanley, and H. A. Makse, "Identification of influential spreaders in complex networks," *Nature physics*, vol. 6, no. 11, pp. 888-893, 2010, doi:10.1038/nphys1746.

[13] B. Bollobás, Graph Theory and Combinatorics*: Proceedings of the Cambridge Combinatorial Conference in Honour of Paul Erdös,[Trinity College, Cambridge, 21-25 March 1983]*: Academic Press, 1984.

[14] S. B. Seidman, "Network structure and minimum degree," *Social networks*, vol. 5, no. 3, pp. 269-287, 1983, doi:10.1016/0378-8733(83)90028-X.

[15] S. Carmi, S. Havlin, S. Kirkpatrick, Y. Shavitt, and E. Shir, "A model of Internet topology using k-shell decomposition," *Proceedings of the National Academy of Sciences*, vol. 104, no. 27, pp. 1-1150-1154, 2007, doi:10.1073/pnas.0701175104.