ICWR2022
May 11-12, 2022; Tehran, Iran
8th International Conference on Web Research

A New Pruning Method for Identifying Influential Nodes in Social Networks

Parinaz Soltanzadeh[1], Alireza Rezvanian*[2]

[1] Department of Computer Engineering, University of Science and Culture, Tehran, Iran
parisoltan77@gmail.com

[2] Assistant professor, Department of Computer Engineering, University of Science and Culture, Tehran, Iran
rezvanian@usc.ac.ir

Abstract

Identifying influential nodes to influence maximization plays an important role in social networks. Social networks are a type of graph data in which each node represents one person, and each edge represents a relationship between two people. Based on the relationship and interaction among people in social networks, they are influenced by each other, and different users in social networks propagate a large amount of information daily. Thus, it is essential to identify the influential people in spreading information. Identifying influential users is modeled as a problem of influence maximization. This paper proposes a pruning method to identify influential nodes effectively in a smaller graph. Therefore, after extracting the social network data, the edges of the graph are weighted by the edge betweenness centrality measure, and then the edges that weigh less than the average weight are pruned, and the influential nodes in the pruned graph are selected using any centrality measure. To evaluate the proposed pruning method, using the LTM diffusion model, the running time and the average number of activated nodes based on a set of initial active nodes compared to the baseline algorithms based on the centrality algorithm have been reported. The simulation results show a relative improvement in the results obtained.

Keywords: Social Network Analysis, Centrality Measures, Graph Pruning, Influence Maximization, Influential Nodes.

JCWR2022
May 11-12, 2022; Tehran, Iran
8th International Conference on Web Research

یک روش هرس جدید برای شناسایی گره‌های موثر در شبکه‌های اجتماعی

پری‌ناز سلطان‌زاده[۱]، علیرضا رضوانیان*[۲]

[۱] گروه مهندسی کامپیوتر، دانشگاه علم و فرهنگ، تهران، ایران
parisoltan77@gmail.com

[۲] استادیار، گروه مهندسی کامپیوتر، دانشگاه علم و فرهنگ، تهران، ایران
rezvanian@usc.ac.ir

چکیده

شناسایی گره‌های موثر در شبکه‌های اجتماعی به منظور بیشینه سازی انتشار از اهمیت بالایی برخوردار است. شبکه‌های اجتماعی، نوعی از داده‌های گرافی هستند که در آنها هر گره معرف یک فرد و هر یال معرف ارتباط میان دو فرد است. با توجه به آنکه افراد مرتبط در شبکه‌های اجتماعی، تحت تاثیر یکدیگر هستند و حجم اطلاعات وسیع روزانه توسط کاربران مختلف در شبکه‌های اجتماعی ارسال می‌شود، تعیین افراد تاثیر گذار در تولید و انتشار اطلاعات بسیار مهم است. استخراج کاربران موثر، به به صورت مسئله بیشینه سازی انتشار مدل می‌شود. در این مقاله، یک روش هرس پیشنهاد شده است تا بتوان در گراف کوچکتری به طور کاراتری گره‌های موثر را شناسایی کرد. بنابراین، پس از استخراج داده‌های شبکه‌های اجتماعی، یالهای گراف توسط معیار مرکزیت بینابینی یال وزن دار شده و سپس یالهایی که وزنی کمتر از میانگین دارند هرس شده و با استفاده از معیارهای مرکزیت گره‌های موثر در گراف هرس شده انتخاب می‌شوند. به منظور ارزیابی روش پیشنهادی، با استفاده از مدل انتشار آستانه خطی، زمان اجرا و متوسط تعداد گره‌های فعال شده براساس مجموعه‌ای از گره‌های فعال اولیه در مقایسه با الگوریتم‌های پایه مبتنی بر مرکزیت گزارش شده است. نتایج شبیه سازی حاکی از بهبود نسبی در نتایج بدست آمده است.

کلمات کلیدی

تحلیل شبکه‌های اجتماعی، معیارهای مرکزیت، هرس گراف، بیشینه سازی انتشار، گره‌های موثر.

۱- مقدمه

امروزه، شبکه‌های اجتماعی در زمینه‌ی گسترش اطلاعات، نظرات، ایده‌ها، نوآوری‌ها و شایعات در مقیاس وسیعی نقش مهمی دارند [۱]. شبکه‌های اجتماعی به بستری ضروری برای سازمان‌ها برای گسترش سیاست‌های خود، تبلیغات شرکت‌ها برای تولید محصولاتشان و تبلیغ عقاید مردم تبدیل شده‌اند [۲]. در همین راستا، موضوع به بیشینه‌سازی تاثیر، مساله‌ی مهمی است. به این معنی که k گره تاثیرگذار (هر گره نشان دهنده‌ی یک کاربر شبکه اجتماعی است.) در شبکه های اجتماعی، معروف به seeds، گره‌های دیگر را متعاقد به اتخاذ یک محصول (یا یک سرویس، یک ایده، یا یک نظر سیاسی) می‌کنند و با استفاده از روش دهان به دهان انتشار اطلاعات، سعی در فعال کردن سایر گره‌ها در شبکه‌های اجتماعی و استفاده از آنها می‌کنند. k گره‌ی مدنظر در شبکه‌های اجتماعی باید به گونه‌ای انتخاب شوند که تعداد گره‌های فعال شده (تحت تاثیر) در شبکه‌های اجتماعی به حداکثر برسند [۲].

با توجه به کاربردهای بالقوه‌ی این مساله در موضوعات مختلف، مانند توصیه شخصی، رتبه بندی خوراک، تبلیغات هدفمند، انتخاب توییت‌های تاثیرگذار و انتخاب وبلاگ‌های آموزنده در سال‌های اخیر، شاهد توجه بسیار زیادی در زمینه‌ی مطالعه‌ی انتشار و به حداکثر رساندن تاثیر در شبکه‌های

اجتماعی آنلاین هستیم [۱۷]. یکی از کاربردهای معروف این حوزه، بازاریابی ویروسی است. جایی که ممکن است شرکتی بخواهد از طریق پیوندهای اجتماعی میان کاربران، محصول جدید خود را در میان برخی اتخاذکنندگان اولیه منتشر کند [۳].

مسئله‌ی یافتن تاثیرگذارترین گره‌ها به منظور بیشینه‌سازی انتشار به دلیل چالش‌های متفاوتی که در زمینه‌ی به حداکثر رساندن تاثیر دارد، مورد توجه است. یکی از چالش‌ها نحوه‌ی مدلسازی فرایند انتشار اطلاعات در شبکه اجتماعی است. دومین چالش این است که موضوع حداکثرسازی تاثیر، از لحاظ تئوری مسئله‌ای پیچیده است. پیدا کردن راه حلی بهینه براساس مدل‌های انتشار، مسئله ای دشوار است و به همین دلیل از آن به عنوان یک مسئله‌ی NP-hard یاد می‌شود. همچنین ارزیابی گسترش تاثیر در هر مجموعه seed جداگانه از لحاظ محاسباتی کاری دشوار است [۳].

۱-۱- بیان مسئله

بیشینه‌سازی انتشار اولین بار توسط Kempe و همکارانش در سال ۲۰۰۳ ، به عنوان یک مساله ی الگوریتمی مطرح شد [۴]. در این مسئله شبکه‌ی اجتماعی در قالب یک گراف G=(V, E) نمایش داده می شود که در آن V مجموعه‌ی گره‌ها (کاربران) و E مجموعه‌ی یال ها (ارتباطات میان کاربران) است. هدف مسئله‌ی بیشینه‌سازی انتشار پیدا کردن یک مجموعه k عضوی از کاربران با حداکثر تاثیر (نفوذ) در گراف G است.

به طور خاص، دو مدل: انتشار مستقل (IC) و آستانه خطی (LT) برای مطالعه‌ی روند انتشار وجود دارد. در مدل آستانه خطی (LT) مقدار آستانه نشان دهنده‌ی تمایل یک گره برای گرفتن یک ایده‌ی جدید از دیگران است. یک مقدار آستانه بزرگتر به این معنی است که گره احتمال تغییر موقعیت آن کمتر است. همچنین در این مدل رفتار جمعی عوامل جمع آوری می‌شود [۱].

در این مقاله، یک روش هرس گراف پیشنهاد شده است تا بتوان در گراف کوچکتری به طور کاراتری گره‌های موثر را شناسایی کرد. بنابراین، پس از استخراج داده‌های شبکه‌های اجتماعی، یال‌های گراف توسط معیار مرکزیت بینابینی یال وزن دار شده و سپس یال‌هایی که وزنی کمتر از میانگین دارند هرس شده و با استفاده از معیارهای مرکزیت گره‌های موثر در گراف هرس شده انتخاب می‌شوند. به منظور ارزیابی روش پیشنهادی، شبیه‌سازی‌هایی بر روی دادگان استاندارد شبکه‌های اجتماعی انجام شده است.

در ادامه در این مقاله، در بخش دوم، پیشینه پژوهش مورد بررسی قرار گرفته است. در بخش سوم الگوریتم پیشنهادی شرح داده شده است، نتایج شبیه‌سازی در بخش چهارم گزارش شده است و بخش پنجم به صورت نتیجه به این مقاله پایان می‌دهد.

۲- پیشینه پژوهش

در سال‌های اخیر روش‌های مختلفی در زمینه‌ی یافتن رئوس تاثیرگذار در شبکه‌های اجتماعی مطرح شده‌اند، اما به طور معمول و به نسبت روش‌های پرهزینه و زمانبری بودند. در این مقاله برخی از روش های مطرح شده در دو دهه ی اخیر را مطرح می کنیم.

ابتدا Domingos و همکارانش مسئله‌ی حداکثرسازی تاثیر را پیشنهاد دادند [۶]. آنها با استفاده از میدان تصادفی مارکوف، مسئله را مدل سازی کرده و راه‌حل‌های ابتکاری ارائه دادند [۷]. سپس در سال ۲۰۰۳، Kempe و همکارانش دو مدل انتشار IC و LT را پیشنهاد دادند [۴]. آنها ثابت کردند که مسئله‌ی حداکثر سازی تاثیر (نفوذ)، یک مسئله‌ی NP-hard است [۵] و از یک روش حریصانه استفاده کردند که اگر چه دارای تقریب خوبی برای گسترش تاثیر داشت، اما از زمان اجرایی خیلی بالایی برخوردار بود و عملا برای شبکه‌های بزرگ مقیاس پذیر قابل استفاده نیست [۴].

در سال ۲۰۰۷ ، Leskovec و همکارانش الگوریتم CELF را پیشنهاد دادند[۸]. آنها تشخیص دادند که لازم نیست همه‌ی گره‌های باقیمانده در هر دور ارزیابی شوند [۷]. این الگوریتم به منظور بهبود کارایی الگوریتم حریصانه پیشنهاد داده شد [۵] . Goyal و همکارانش الگوریتم CELF++ را پیشنهاد دادند [۹]. این الگوریتم از ۳۵ به ۵۵ درصد سریعتر از CELF است [۷]. درمقابل، Tang و همکارانش الگوریتم حریصانه‌ای را براساس محدودیت hop پیشنهاد دادند [۱۰]. اما بازده‌ی الگوریتم حریصانه هنوز ناکافی بود [۵]. در سال ۲۰۱۰ ، Wang و همکارانش، الگوریتمی حریصانه مبتنی بر جوامع به منظور انتخاب k گره ی تاثیرگذار ارائه دادند. نام الگوریتم CGA است. الگوریتم پیشنهادی شامل دو مولفه است : ۱) الگوریتمی برای شناسایی جوامع در یک شبکه ی اجتماعی با در نظر گرفتن انتشار اطلاعات ۲) یک الگوریتم برنامه نویسی پویا برای انتخاب جوامع به منظور یافتن گره‌های تاثیرگذار. الگوریتم پیشنهاد شده‌ی آنها نشان می‌دهد که در یک شبکه‌ی اجتماعی بزرگ موبایل، الگوریتم ذکر شده سریعتر از الگوریتم حریصانه‌ی پیشرفته به منظور یافتن k گره ی تاثیرگذار است و خطای تقریبی الگوریتم کوچک است [۱۱].

در سال‌های ۲۰۰۹ و ۲۰۱۰ ، Chen و همکارانش به حداکثر رساندن تاثیری کارآمد از دو جهت مکمل را مطالعه کردند. ۱) پیشنهاد یک الگوریتم حریصانه جدید برای بهبود الگوریتم حریصانه‌ی اصلی ۲) ارائه پیشنهادی هیوریستیک (اکتشافی) با هدف درجه تخفیفی که گسترش تاثیر را بهبود می‌بخشد. آنها چندین الگوریتم اکتشافی را برای محاسبه ی گسترش نفوذ ابداع کردند [۷]. در سال ۲۰۱۴ ، Chen و همکارانش چارچوب جدیدی را به نام حداکثر سازی تاثیر مبتنی بر جامعه (CIM) برای مقابله با مسئله‌ی حداکثرسازی تاثیر با تاکید بر مسئله کارایی زمان پیشنهاد دادند [۷,۱۲]. این چارچوب پیشنهادی شامل سه مرحله است: ۱) شناسایی جامعه ۲) تولید کاندیداها ۳) انتخاب مجموعه اولیه. فاز اول ساختار جامعه شبکه را کشف می‌کند. فاز دوم، از اطلاعات جوامع برای محدود کردن کاندیداهای احتمالی مجموعه اولیه استفاده می‌کند. فاز سوم، گره‌های مجموعه اولیه را از مجموعه کاندید، نهایی می‌کند. نتایج تجربی نشان می‌دهند که الگوریتم پیشنهادی CIM به طور قابل توجهی از نظر کارایی و مقیاس پذیری از الگوریتم‌های پیشرفته بهتر عمل می‌کند [۱۲].

در سال ۲۰۱۶ ، Gong و همکارانش، یک الگوریتم حافظه‌ای برای حداکثر نفوذ (تاثیر) مبتنی بر جامعه در شبکه‌های اجتماعی پیشنهاد دادند. الگوریتم پیشنهادی شامل سه فاز است: ۱) کشف ساختار جامعه و یافتن اجتماع ۲) استفاده از اطلاعات جوامع به منظور کاهش تعداد نامزدهای احتمالی اولیه (مرحله ی تولید نامزدها) ۳) انتخاب گره‌های اولیه از مجموعه

کاندیدا (انتخاب مجموعه ی اولیه). با استفاده از ویژگی‌های ساختارهای جوامع، می‌توان از اطلاعات همپوشانی جلوگیری کرد. آزمایش‌هایی که بر روی سه مجموعه داده واقعی انجام شده است، نشان می‌دهد که این الگوریتم از سایر الگوریتم‌ها از لحاظ گسترش تاثیر، زمان اجرا و مقیاس پذیری دارای عملکرد بهتری است [۱۳]. در سال ۲۰۱۹، Huang و همکارانش ، الگوریتمی به اسم CIMA برای حل مسئله بیشینه سازی انتشار مبتنی بر جوامع ارائه دادند. بیشتر روش‌هایی که قبل از این مطرح شده بودند، از مشکلات هزینه محاسباتی بالا و عدم تضمین عملکرد رنج می‌بردند. زیرا این روش‌ها به شدت به تجزیه و تحلیل و بهره برداری از ساختار شبکه بستگی داشتند. اما Huang و همکارانش، الگوریتمی مبتنی بر جامعه پیشنهاد دادند که شامل سه مرحله است: ۱) تشخیص جامعه ۲) تولید جامعه کاندید ۳) انتخاب مجموعه اولیه. ما ابتدا فرآیند تولید جامعه کاندیداها را پیشنهاد می‌دهیم، که از اطلاعات ساختار جامعه و همچنین ویژگی گره برای محدود کردن نامزدهای احتمالی جامعه استفاده می‌کند. سپس ما یک مدل برای پیش بینی قدرت تأثیر بین گره‌ها در شبکه منتسب ارائه می‌دهیم، که از شباهت ساختار توپولوژی و شباهت ویژگی بین گره‌ها علاوه بر قدرت تعامل اجتماعی بهره می‌برد، بنابراین دقت پیشبینی را در مقایسه با روش‌های موجود به طور قابل توجهی بهبود می‌بخشد. در این الگوریتم برای فاز اول و شناسایی و استخراج جوامع از الگوریتم لووین استفاده شده است. سپس برای فاز دوم و تولید جوامع کاندید، جوامع استخراج شده هرس می‌شوند. برای فاز سوم و انتخاب مجموعه‌ی اولیه گره‌ی تاثیرگذار براساس قانون حداکثر سود حاشیه‌ای تاثیرگذاری برای هر دور انتخاب می‌شود. در جدول استفاده شده، هر رکورد جدول در برگیرنده‌ی اطلاعات یک گره است که شامل گره، سود حاشیه‌ای تاثیرگذاری گره و دفعات تکرار است. در هربار تکرار جدول براساس سود حاشیه‌ای تاثیرگذاری گره مرتب شده و گره با بیشترین سود انتخاب می‌شود. در قیاس الگوریتم CIMA با سایر الگوریتم‌های موجود، از نظر گسترش تاثیر و زمان اجرا دارای عملکرد بهتری است [۱۴].

۳- الگوریتم پیشنهادی برای هرس گراف

همان طور که در بخش ۲ توضیح داده شد، در بسیاری از الگوریتم‌هایی که در حوزه ی بیشینه‌سازی انتشار وجود دارند، مشکلاتی برای زمان اجرا و کارایی وجود دارد. ما در الگوریتم پیشنهادی خود، سعی داریم که از نظر گسترش تاثیر و زمان اجرا عملکرد مطلوب‌تری را برای الگوریتم‌های موجود به وجود آوریم. الگوریتم پیشنهادی برای هرس گراف به منظور بیشینه سازی انتشار در شبکه های اجتماعی دارای ۵ فاز است:

۱) استخراج داده‌های گرافی یک شبکه اجتماعی،
۲) وزن دار کردن یال‌های گراف براساس مرکزیت بینابینی،
۳) هرس گراف با حذف یال‌هایی که دارای وزن کمتری نسبت به میانگین وزن یال‌های گراف دارند،
۴) شناسایی k گره موثر براساس معیارهای مرکزیت (درجه، بینابینی، نزدیکی و رتبه‌بندی پیج) بر روی گراف هرس شده، و
۵) اعمال مدل انتشار LTM بر روی گراف اصلی براساس گره‌های فعال بدست آمده.

معیارهای مرکزیت مورد استفاده در فاز چهارم به شرح زیر است:

۱) مرکزیت درجه: این معیار میزان پیوندهایی که فرد با دیگر افراد در شبکه دارد را محاسبه می‌کند. هر گره‌ای که مرکزیت درجه آن بیشتر باشد، موثرتر است. معیار مرکزیت درجه با استفاده از رابطه‌ی (۱) محاسبه می‌شود

$$C_D(v_i) = d_i = \sum_j A_{ij} \quad (۱)$$

۲) مرکزیت بینابینی[۱]: این معیار نسبت تعداد دفعاتی که یک گره یا یک یال بر روی کوتاهترین مسیر میان گره‌های مختلف یک گراف قرار می‌گیرد را محاسبه می‌کند. هر چه بینابینی یک گره بیشتر باشد، یعنی آن گره در مکان استراتژیکتری قرار گرفته است. در واقع نمایشی برای میزان قابلیت هر گره برای کمک به دسترسی سایرین به اطلاعات و یا گسترش تاثیر در یک شبکه است. پس با استفاده از این معیار می توان مشخص کرد که کدام گره بهترین واسطه برای ارتباط برقرار کردن بین سایر جفت گره‌ها است. این معیار با استفاده از رابطه‌ی (۲) و (۳) محاسبه می شود

$$\delta_{st}(u) = \frac{\sigma_{st}(u)}{\sigma_{st}} \quad (۲)$$

در رابطه‌ی (۲) احتمال این که u در ارتباط بین s و t نقش داشته باشد، بررسی می شود

$$C_B(u) = \sum s\#u \in V \sum t\#u \in V \delta_{st}(u) \quad (۳)$$

در رابطه‌ی (۳) مجموع تمام کوتاهترین مسیرها بین هر جفت گره تعریف می‌شود. بنابراین معیار بینابینی را می‌توان این گونه محاسبه کرد.

۳) مرکزیت نزدیکی[۲]: این معیار عکس متوسط فاصله ی یک گره تا گره های دیگر گراف را محاسبه می کند. گره ای که دارای بیشترین مقدار نزدیکی باشد ، سرعت دسترسی بیشتری به سایر گره ها دارد و می تواند در مدت زمان کمی به همه ی گره ها اطلاعات ارسال کند یا از آن ها اطلاعات بگیرد. . پس از محاسبه ی نزدیکی هر گره ، به کمک تابع sort() می توان گره ها را از بیشترین نزدیکی به کمترین در یک لیست مرتب کرد. این معیار با استفاده از رابطه ی (۴) و (۵) محاسبه می شود

$$D_{avg}(v_i) = \frac{1}{(n-1)} \sum_{j\#i}^{n} g(v_i, v_j) \quad (۴)$$

$$C_c(v_i) = \frac{n-1}{\sum_{j\#i}^{n} g(v_i, v_j)} \quad (۵)$$

٤) پیج‌رنک[۳]: برای این معیار در ابتدا فرض می‌شود که همه‌ی گره ها دارای ارزش یکسانی هستند. هر گره ارزش خودش را به نسبت مساوی میان اتصال‌های خروجی‌اش تقسیم می‌کند. ارزش جدید هر گره عبارت است از مجموع ارزشی که گره‌های ارجاع دهنده به او می‌دهند. این فرآیند تا زمانیکه

مقادیر گره‌ها همگرا شوند ادامه پیدا می‌کند. معیار رتبه پیج با استفاده از رابطه‌ی (۶) محاسبه می‌شود:

$$PR(u) = \sum_{v \in B_u} \frac{PR(v)}{L(v)} \quad (۶)$$

۵) **انتخاب تصادفی**: توسط این معیار، به صورت تصادفی گره‌ها مرتب می‌شوند.

با استفاده از معیار مرکزیت بینابینی یال، در فاز دوم، گراف وزن‌دار می‌شود و در فاز سوم، به منظر هرس گراف، یال‌هایی که وزنی کمتر از میانگین مقدار بینابینی کل یال‌های گراف دارند، حذف می‌شود تا گراف کوچکتری بدست آید، ایده اصلی حفظ یال‌های مهتر براساس معیار مرکزیت بینابینی یال است چون این یال‌ها نقش واسطه مهم‌تری در انتقال و انتشار اطلاعات به عهده دارند و از طرفی فرآیند سنگین محاسبه برای پیدا کردن k گره موثر بر روی گراف کوچتری اعمال می‌شود.

در فاز چهارم با استفاده از هریک از ۵ معیار درجه، نزدیکی، بینابینی، رتبه پیج و تصادفی می‌توان k گره‌ی موثر را انتخاب کرد.

در فاز پنجم، برای انتشار از مدل آستانه خطی (LTM) استفاده می‌شود، این مدل توسط Granovetter پیشنهاد شده است. در این مدل هر گره دو وضعیت فعال و غیر فعال دارد. این مدل عمدتا بر آستانه رفتاری در انتشار نفوذ تمرکز دارد. مقدار آستانه نشان دهنده تمایل یک گره برای گرفتن یک ایده جدید از دیگران است. یک مقدار آستانه بزرگتر به این معنی است که گره احتمال تغییر موقعیت آن کمتر است. روال این روش به این صورت است که یک گره v توسط هر همسایه‌ای مانند w مطابق با وزن، $b_{v,w}$ تحت تاثیر قرار می‌گیرد. هر گره‌ی v زمانی فعال می‌شود که جمع وزن همسایگان فعالش حداقل Θ_v باشد. مقدار Θ_v یک عدد ثابت تصادفی بین ۰ و ۱ است.

٤- نتایج شبیه سازی

برای ارزیابی الگوریتم پیشنهادی برای پیدا کردن گره های موثر به منظور بیشینه‌سازی انتشار، از کتابخانه های NetworkX، pandas و NDlib در پایتون استفاده شده است. گراف‌های مورد استفاده بدون وزن در نظر گرفته شده است. در کلیه آزمایش‌ها، از روش شبیه‌سازی مونت کارلو با تعداد اجرای ۱۰۰۰ مرتبه و مدل انتشار آستانه خطی (LTM) استفاده شده است به طوریکه به هر گره مقداری تصادفی آستانه در بازه [0, 1] به صورت وزنی برابر با یک تقسیم بر درجه هر گره خصاص داده شده است.

نتایج بدست آمده از نظر گره‌های فعال شده به صورت میانگین گیری شده با استفاده از ۵ معیار درجه، بینابینی، نزدیکی، تصادفی و پیج‌رنک بعد از هرس گراف در آزمایش اول و زمان اجرا در آزمایش دوم مورد مقایسه قرار گرفته است.

٤-١- مجموعه دادگان

آزمایش‌ها بر روی ۵ مجموعه داده کاراته، دولفین، فوتبال، جاز و ایمیل انجام شده است. در زیر جزئیات پنج مجموعه داده مورد استفاده، ذکر شده است [۱۵]

- کاراته (Zachary Karate Club): مجموعه داده باشگاه کاراته زاکاری یک مجموعه داده مشهور و پرکاربرد است که در سال ۱۹۷۷ توسط شخصی به نام وین زاکاری از اعضای یک باشگاه کاراته دانشگاهی جمع‌آوری شد. هرگره در گراف این مجموعه داده نماینده یک عضو از این باشگاه و هر یال نشان دهنده یک ارتباط بین دو عضو این باشگاه است. این مجموعه داده با ۳۴ گره کوچکترین گراف در مجموعه داده‌های انتخاب شده است.
- دولفین (Dolphins): مجموعه داده دولفین بین سال‌های ۱۹۹۴-۲۰۰۱ از نژاد خاصی از دولفین‌ها که در نیوزلند زندگی می‌کنند، جمع‌آوری شده است.
- جاز (Jazz Musicians): مجموعه داده جاز از همکاری بین نوازندگان جاز در سال ۲۰۰۳ جمع‌آوری شده است. در گراف این مجموعه داده، هر گره نماینده یک نوازنده جاز و هر یال نشان دهنده دو نوازنده‌ای است که در یک گروه با هم می‌نوازند.
- ایمیل (Email): این مجموعه داده یک شبکه ارتباطی از طریق ایمیل متعلق به دانشگاهی در اسپانیا است که گره‌ها نشان دهنده کاربران و یال‌ها نشان دهنده حداقل یک ایمیل ارسال شده است.

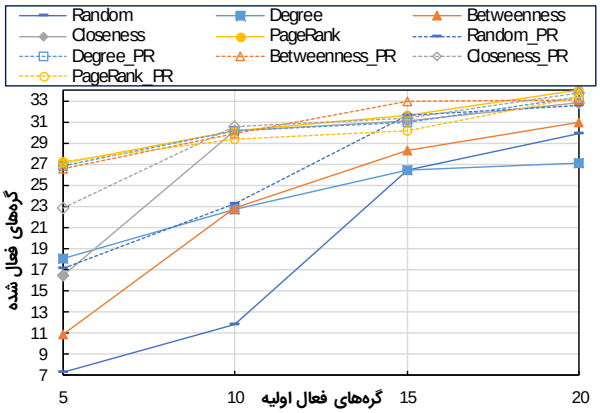
٤-٢- آزمایش اول: بررسی نتایج تعداد گره‌های فعال شده

برای گزارش تعداد کل گره های فعال شده، از متوسط تعداد کل گره‌های فعال شده در ۳۰ بار اجرای مستقل استفاده شده است. در شکل‌های (۱) تا (۵) به ترتیب برای کاراته، دولفین، فوتبال، جاز و ایمیل نتایج مربوط به میانگین گره‌های فعال شده به ازای مجموعه اولیه از ۵ گره اولیه تا ۲۰ گره اولیه براساس مدل انتشار LTM نمایش داده شده است.

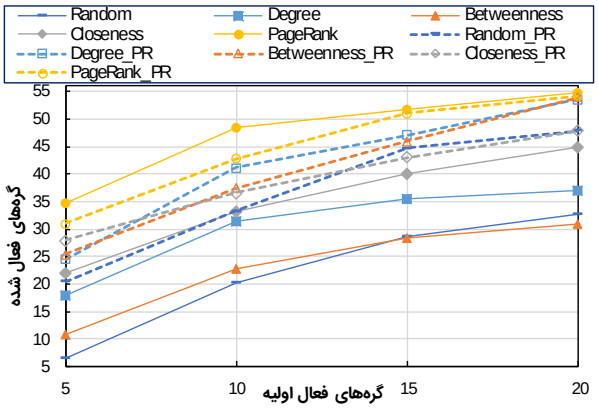
برای مقایسه از نتایج بدست آمده از روش تصادفی (Random)، درجه (Degree)، بینابینی (Betweenness)، نزدیکی (Closeness)، پیج‌رنک (PageRank) [۱، ۲، ۱۷] در مقایسه با نتایج بدست آمده این روش‌ها براساس الگوریتم هرس پیشنهادی با نام های Random_PR, Degree_PR, Closeness_PR, Betweenness_PR و PageRank_PR استفاده شده است. براساس نتایج گزارش شده، بهبود نسبی در روش هرس مشاهده می‌شود، اما در برخی موارد برای روش رتبه پیج خیلی موثر نیست.

٤-٣- آزمایش دوم: بررسی نتایج زمان اجرا

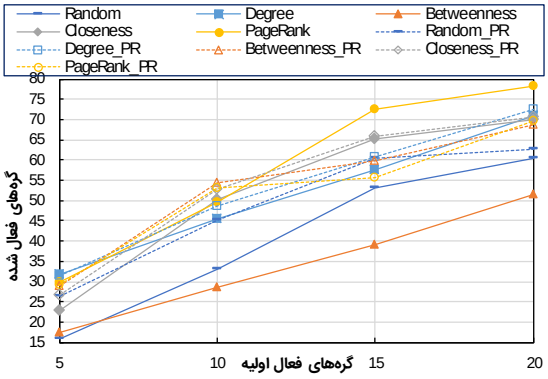
در این آزمایش متوسط زمان اجرای هر الگوریتم به ازای ۳۰ بار اجرای مستقل گزارش شده است. در شکل‌های (۶) تا ۱۰ متوسط زمان اجرای الگوریتم‌ها به ازای تعداد گره‌های اولیه مختلف از ۵ گره تا ۲۰ گره به صورت نمودار میله‌ای نمایش داده شده است. چنانچه از نتایج مشخص است، روش هرس گراف بخشی از زمان اجرا را به خود اختصاص می‌دهد ولی زمان اضافه شده با توجه به بهبود حاصل شده در بخش انتشار ناچیز است، بنابراین به نظر می‌رسد، روش پیشنهادی می‌تواند در گراف‌های خیلی بزرگ نتیجه کارایی داشته باشد.



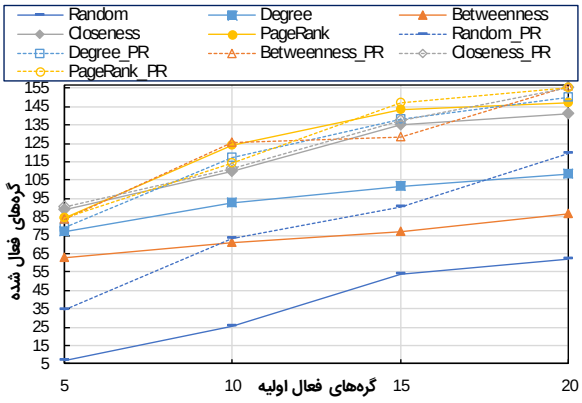
شکل (۱): نتایج گره‌های فعال شده برای کاراته



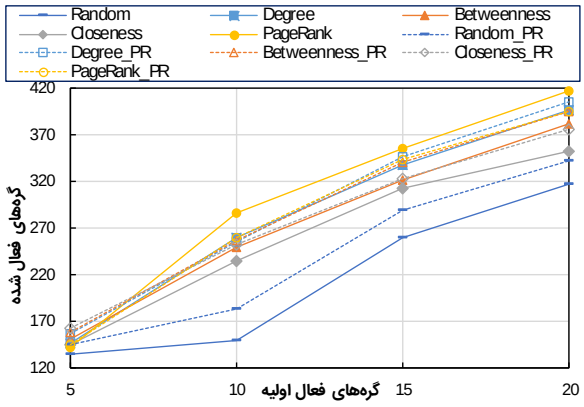
شکل (۲): نتایج گره‌های فعال شده برای دولفین



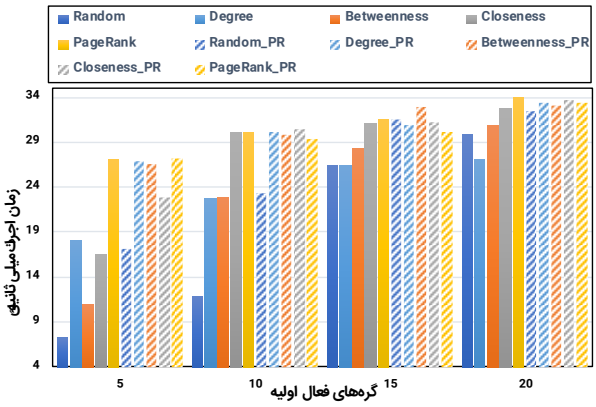
شکل (۳): نتایج گره‌های فعال شده برای فوتبال



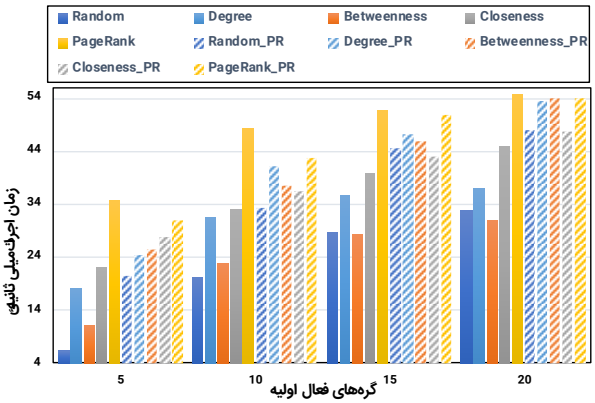
شکل (٤): نتایج گره‌های فعال شده برای جاز



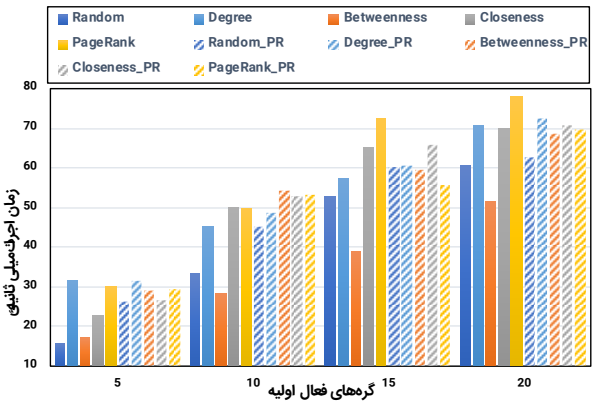
شکل (۵): نتایج گره‌های فعال شده برای ایمیل



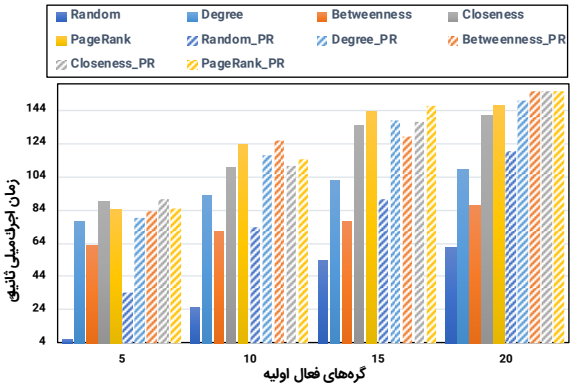
شکل (٦): نتایج زمان اجرا برای کاراته



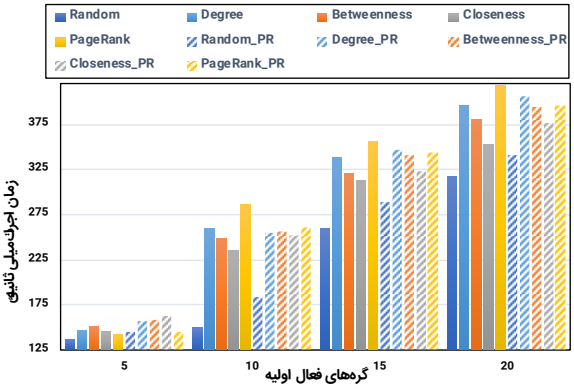
شکل (۷): نتایج زمان اجرا برای دولفین



شکل (۸): نتایج زمان اجرا برای فوتبال



شکل (۹): نتایج زمان اجرا برای جاز



شکل (۱۰): نتایج زمان اجرا برای ایمیل

با توجه به نتایج بدست آمده براساس گسترش انتشار و زمان اجرا و مقیاس پذیری الگوریتم، روش پیشنهادی می‌تواند به کاهش زمان اجرای الگوریتم در گراف‌های بزرگ با کاهش سایز گراف کمک کند. از طرفی ممکن است در گراف‌های کوچک اثر زمان و نتایج بدست آمده خیلی معنادار نباشد ولی در عین حال تاثیر بخش‌های کوچک و کم وزن نیز بر شبکه کاهش یافته است. در میان نتایج بدست آمده، روش پیج‌رنک از موفقیت بیشتری نسبت به سایر روش‌ها برخوردار است.

۵- نتیجه

در سال‌های اخیر، شبکه‌های اجتماعی در زمینه‌ی گسترش اطلاعات، نظرات، ایده‌ها، نوآوری‌ها و شایعات در مقیاس وسیعی نقش مهمی را ایفا میکنند به طوریکه این شبکه‌ها به بستری موثر برای سازمان‌ها برای گسترش سیاست‌های خود، تبلیغات شرکت‌ها برای تولید محصولاتشان و تبلیغ عقاید مردم تبدیل شده‌اند و به طور کلی یافتن گره‌های موثر برای این اهداف مورد توجه است. در این مقاله یک الگوریتم هرس گراف به صورت وزندهی براساس معیار بینابینی پیشنهاد شد تا برای هدف یافتن تاثیرگذارترین کاربران در شبکه اجتماعی مورد استفاده قرار گیرد به طوریکه در یک گراف هرس شده کوچکتر به دنبال محاسبه بر روی بخش‌های مهم گراف بود. برای ارزیابی کارایی الگوریتم هرس گراف پیشنهادی، آزمایش‌هایی بر روی چند مجموعه داده استاندارد شبکه‌های اجتماعی انجام شد که نتایج شبیه‌سازی حاکی از مطلوبیت نسبی نتایج الگوریتم هرس پیشنهادی دارد.

مراجع

[1] M. Azaouzi, W. Mnasri, L. B. Romdhane. "New trends in influence maximization models." Computer Science Review 40 (2021): 100393.
[2] M. M. D. Khomami, A. Rezvanian, M. R. Meybodi, “CFIN: A community-based algorithm for finding influential nodes in complex social networks.” The Journal of Supercomputing 77.3 (2021): 2207-2236.
[3] J. Li, T. Cai, K. Deng, X. Wang, T. Sellis, F. Xia, “Community-diversified influence maximization in social networks,” Information Systems, vol. 92, 2020, 101522,
[4] Y. Li, J. Fan, Y. Wang, K. Tan, “Influence Maximization on Social Graphs: A Survey,” in IEEE Transactions on Knowledge and Data Engineering, vol. 30, no. 10, pp. 1852-1872, 1 2018.
[5] D. Kempe, J. Kleinberg, É. Tardos, “Maximizing the Spread of Influence through a Social Network,” Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2003, 137-146.
[6] W. Li, Z. Li, A. M. Luvembe, C. Yang, “Influence maximization algorithm based on Gaussian propagation model,” Inf. Sci., vol. 568, 2021, pp. 386-402,
[7] P. Domingos, M. Richardson, “Mining the Network Value of Customers,” Proceedings of the Seventh International Conference on Knowledge Discovery and Data Mining, 2001 , pp. 57–66
[8] W. Zhu, Z. Zhenfang, f. Liu, “Influence Maximization in Social Network Considering Memory Effect and Social Reinforcement Effect,” Future Internet. V. 11, pp. 95, 2019.
[9] J. Leskovec, A. Krause, C. Guestrin, C. Faloutsos, J. Vanbriesen, N. Glance, “Cost-effective outbreak detection in networks,” Proceedings of the ACM SIGKDD Int. Conf. on Knowl Discov Data Min. pp. 420-429, 2007.
[10] A. Goyal, W. Lu, L. Lakshmanan, “CELF++: Optimizing the greedy algorithm for influence maximization in social networks,” Proceedings of the 20th Int. Conf. Companion on WWW 2011, pp. 47-48, 2011.
[11] J. Tang, X. Tang, J. Yuan, “An efficient and effective hop-based approach for influence maximization in social networks,” Social Network Analysis and Mining, 8, 2018.
[12] Y. Wang, G. Cong, G. Song, K. Xie, “Community-based Greedy algorithm for mining top-K influential nodes in mobile social networks,” KDD, 2010.
[13] S. Singh, K. Singh, A. Kumar, B. Biswas, “CoIM: Community-Based Influence Maximization in Social Networks.”, Int. Conf. on Adv. Inform. for Comput Res., Singapore 2019.
[14] M. Gong, C. Song, C. Duan, L. Ma, B. Shen, “An Efficient Memetic Algorithm for Influence Maximization in Social Networks,” in IEEE Computational Intelligence Magazine, vol. 11, no. 3, pp. 22-33, 2016.
[15] H. Huang, H. Shen, Z. Meng, “Community-based influence maximization in attributed networks,” Appl Intell 50, 354–364, 2020.
[16] J. Leskovec and A. Krevl, SNAP Datasets: Stanford Large Network Dataset Collection, http://snap.stanford.edu/data, Jun 2014.
[17] Z. Sajdeh, A. Rezvanian, Infulence Maximization in Social Networks based on User Similarity, 2nd National

Conference of Informatics Society of Iran (NIC99), IPM, Tehran, December 2020.

زیرنویس‌ها

[1] Betweenness
[2] Closeness
[3] PageRank