

یک الگوریتم چندهدفه برای شناسایی گره‌های پرنفوذ در شبکه‌های اجتماعی

چیمن صلواتی^۱، کارشناس ارشد؛ علیرضا عبدالله‌پوری^۲، استادیار؛ ژاله منبری^۳، کارشناس ارشد

۱- دانشکده مهندسی - گروه مهندسی کامپیوتر - دانشگاه کردستان - سنندج - ایران - Chiman.salavati@uok.ac.ir

۲- دانشکده مهندسی - گروه مهندسی کامپیوتر - دانشگاه کردستان - سنندج - ایران - Abdollahpour@uok.ac.ir

۳- دانشکده مهندسی - گروه مهندسی کامپیوتر - دانشگاه کردستان - سنندج - ایران - Zh.manbari@uok.ac.ir

چکیده: با گسترش شبکه‌های اجتماعی، روابط بین افراد شکل تازه‌ای به خود گرفته است. یکی از مسائل مطرح در شبکه‌های اجتماعی، مسئله نفوذ اجتماعی است. پژوهش‌های انجام‌شده در مورد نفوذ اجتماعی و چگونگی انتشار اطلاعات در شبکه‌های اجتماعی، بیان‌گر این است که پذیرش یا رد یک الگوی جدید توسط یک فرد، به پذیرش یا رد دوستان آن فرد بستگی دارد. زیرا افراد به دوستان خود بیشتر از تبلیغات سایر منابع اعتماد دارند. در نتیجه، بسیاری از شرکت‌ها به سمت این روش که بازاریابی ویروسی نامیده می‌شود، متمایل شده‌اند. با وجود تعداد بسیار زیاد کاربران شبکه‌های اجتماعی، انتخاب ارزشمندترین کاربران به‌عنوان کاربران هدف که بتوان از طریق آن‌ها به بیش‌ترین میزان گسترش در شبکه با کم‌ترین هزینه دست یافت، از اهمیت زیادی برخوردار است. در این مقاله، یک روش جدید برای شناسایی گره‌های پرنفوذ در شبکه‌های اجتماعی به نام الگوریتم چندهدفه مبتنی بر اطلاعات ساختاری (MOSI) شده است. عملکرد روش پیشنهادی بر مبنای دو هدف «بیشینه‌سازی سود» و «کمینه‌سازی شباهت میان کاربران انتخابی» است. ارزیابی بر روی مجموعه داده‌های واقعی، نشان می‌دهد که روش پیشنهادی دارای قدرت گسترش بیشتری در مقایسه با روش‌های دیگر است.

واژه‌های کلیدی: شبکه اجتماعی، افراد پرنفوذ، بهینه‌سازی چندهدفه، جبهه پرتو، الگوریتم ژنتیک، مدل SIR.

A Multi-objective Algorithm for Identifying Influential Nodes in Social Networks

Ch. Salavati¹, MSc; A. Abdollahpouri², Assistant professor; Zh. Manbari³, MSc

1- Department of Computer Engineering, University of Kurdistan, Sanandaj, Iran, Email: Chiman.salavati@uok.ac.ir

2- Department of Computer Engineering, University of Kurdistan, Sanandaj, Iran, Email: Abdollahpour@uok.ac.ir

3- Department of Computer Engineering, University of Kurdistan, Sanandaj, Iran, Email: Zh.manbari@uok.ac.ir

Abstract: With the expansion of social networks, relationship between people has taken a new form. One of the important issues in social networks is social influence. Research on social influences and how information is disseminated in social networks, indicates that accepting or rejecting a new pattern by a person depends on the acceptance or rejection of the friends of that person. That is, because the people usually trust their friends more than other sources of advertising. As a result, many companies are focused on this type of advertisement which is called viral marketing. Given a large number of users in a social network, selecting the most influential users as target users, through which a company can reach the highest expansion in the network with the lowest cost, is of great importance. In this paper, a new method for identifying the influential nodes in social networks is proposed which is called MOSI (Multi-Objective algorithm based on Structured Information). The proposed method has two goals: "maximize profit" and "minimize similarity among selected users". The evaluation of the proposed method on real datasets indicates that our method has a greater expansion power in comparison with other similar methods.

Keywords: Social networks, influential nodes, multi-objective optimization, Pareto front, genetic algorithm, SIR model.

تاریخ ارسال مقاله: ۱۳۹۷/۰۳/۱۹

تاریخ اصلاح مقاله: ۱۳۹۷/۰۷/۰۳ و ۱۳۹۷/۰۹/۰۷

تاریخ پذیرش مقاله: ۱۳۹۷/۰۹/۲۶

نام نویسنده مسئول: علیرضا عبدالله‌پوری

نشانی نویسنده مسئول: ایران - سنندج - بلوار پاسداران - دانشگاه کردستان - دانشکده مهندسی - گروه کامپیوتر

۱- مقدمه

زودرس» را دارند. بعضی از روش‌های مکاشفه‌ای مانند مرکزیت درجه [۸]، مرکزیت بینایی [۸]، مرکزیت نزدیکی [۹، ۱۰] و روش‌های حریصانه [۱۱، ۱۲]، با وجود این‌که می‌توانند گره‌های تأثیرگذار را شناسایی کنند، در برابر شبکه‌هایی با ساختار جامعه از خود محدودیت نشان می‌دهند. از این‌رو، الگوریتم‌های فراابتکاری، کاربرد بیشتری در حل این نوع مسائل دارند. این الگوریتم‌ها دارای راه‌کارهایی برای مقابله با گیرکردن در بهینه محلی هستند. معروف‌ترین الگوریتم‌های تقریبی برای حل مسائل بهینه‌سازی، روش‌های مبتنی بر هوش جمعی^۵ هستند. این الگوریتم‌ها دارای جمعیتی از عامل‌های مصنوعی هستند که رفتار گروهی حیوانات در دنیای واقعی را شبیه‌سازی می‌کنند. نظم میان این عامل‌ها و کار گروهی‌شان، منجر به حل مسائل پیچیده می‌گردد. الگوریتم ژنتیک^۶ (GA) [۱۳] یکی از محبوب‌ترین روش‌های هوش جمعی است که مؤثر بودن آن در حل مسائل بهینه‌سازی ثابت شده است.

در این مقاله، یک روش شناسایی گره‌های پرنفوذ در شبکه‌های اجتماعی به نام الگوریتم چندهدفه مبتنی بر اطلاعات ساختاری^۷ (MOSI) معرفی شده است. روش پیشنهادی شامل چهار گام و مبتنی بر الگوریتم چندهدفه ژنتیک (NSGA-II) [۱۴] است. در گام اول، یک جمعیت اولیه تصادفی تولید می‌شود. در گام دوم، برای هر راه‌حل (کروموزوم) توابع هدف محاسبه می‌شود. در گام سوم، یک الگوریتم مرتب‌سازی نامغلوب برای مشخص کردن جبهه پرتو استفاده می‌شود. در گام چهارم، عملگرهای تقاطع و جهش اعمال می‌شود تا جمعیت جدید تولید شود. در ادامه، اگر شرط خاتمه الگوریتم (تعداد تکرار الگوریتم) برقرار باشد، الگوریتم متوقف می‌شود. در نهایت تمامی زیرمجموعه گره‌های نامغلوب به دست آمده با مدل انتشار SIR^۸ مورد ارزیابی قرار می‌گیرند و بهترین مجموعه گره به عنوان خروجی الگوریتم انتخاب می‌شود. ارزیابی‌ها بر روی مجموعه داده‌های واقعی انجام گرفته و با چندین الگوریتم شناخته شده مقایسه شده است. نتایج آزمایش‌ها، کارایی الگوریتم پیشنهادی را از لحاظ قدرت نفوذ تأیید می‌کنند.

ادامه این مقاله به صورت زیر سازمان‌دهی شده است. در بخش ۲ کارهای مرتبط بیان می‌شود. در بخش ۳ مراحل روش پیشنهادی با جزئیات توضیح داده می‌شود. در بخش ۴ روش پیشنهادی بر روی مجموعه داده واقعی ارزیابی می‌شود و در بخش ۵ نتیجه‌گیری مقاله ارائه می‌گردد.

۲- کارهای مرتبط

با افزایش کاربرد شناسایی گره‌های پرنفوذ در بسیاری از زمینه‌ها، الگوریتم‌های متنوعی بدین منظور ارائه شده‌اند. ریچاردسون و دمینگوس [۱۵] پیشگامان مطالعه گره‌های هدف در شبکه‌های اجتماعی هستند. کمپ و همکاران [۱۶] سعی کردند مسئله بیشینه‌سازی نفوذ را به عنوان یک مسئله بهینه‌سازی گسسته بر اساس مدل آبخاری مستقل حل کنند. برخی از روش‌ها، از شبیه‌سازی مونت کارلو جهت برآورد میزان گسترش نفوذ گره‌ها استفاده می‌کنند، اما حتی یک‌بار اجرا از این الگوریتم‌ها،

امروزه، شبکه‌های اجتماعی، جهان را از جنبه‌های مختلفی تغییر داده‌اند. در یک شبکه اجتماعی، کاربران ارتباطات و جوامع مختلفی را ایجاد می‌کنند و نظرات کاربران بر روی تصمیمات خرید دیگران تأثیر می‌گذارد. یکی از مسائل مطرح در شبکه‌های اجتماعی، مسئله نفوذ اجتماعی است. به دلیل این‌که که افراد به دوستان خود بیشتر از تبلیغات سایر منابع اعتماد دارند، پذیرش یا رد تبلیغ یک کالا توسط یک فرد، به پذیرش یا رد دوستان آن فرد بستگی دارد. بسیاری از شرکت‌ها به سمت این روش که بازاریابی ویروسی^۱ نامیده می‌شود، متمایل شده‌اند [۱، ۲]. تبلیغات دهان‌به‌دهان^۲ با استفاده از سایت‌های شبکه اجتماعی برای دسترسی سریع به تعداد زیادی از افراد بسیار مفید است. از آنجایی که فرآیند بازاریابی از طریق تعداد محدودی از کاربران آغاز می‌شود، به منظور گسترش بیش‌تر تبلیغات می‌بایست در ابتدا افراد مؤثر انتخاب شوند.

با وجود تعداد بسیار زیاد کاربران شبکه‌های اجتماعی، انتخاب ارزشمندترین کاربران به عنوان کاربران هدف^۳ که بتوان از طریق آن‌ها به بیش‌ترین میزان گسترش در شبکه با کم‌ترین هزینه دست‌یافت، از اهمیت زیادی برخوردار است [۵-۳]. شبکه به صورت گراف مدل می‌شود که در آن گره‌ها و یال‌ها به ترتیب کاربران و ارتباطات آن‌ها می‌باشند. به منظور شناسایی افراد مؤثر، اغلب از ویژگی ساختاری گراف شبکه و موقعیت گره‌ها استفاده شده است. در بیش‌تر مطالعات انجام شده، اهمیت یک گره را با تعداد گره‌هایی که می‌تواند فعال کند، تخمین می‌زنند. اغلب روش‌ها، گره‌هایی را به عنوان گره‌های پرنفوذ انتخاب می‌کنند که میزان گسترش نفوذ بیشتری داشته باشند. اما در فعالیت‌های اقتصادی، کاربرانی ارزشمند هستند که بتوان از طریق آن‌ها در قبال هزینه کم، بیش‌ترین سود ممکن را به دست آورد [۶]. انتخاب کاربران پرنفوذ را می‌توان به عنوان فرآیندی از شناسایی کاربران دارای قدرت گسترش زیاد که کم‌ترین شباهت ممکن را به یکدیگر دارند، در نظر گرفت. دو کاربر با یکدیگر دارای افزونگی هستند اگر از لحاظ کاربرانی که تحت نفوذ و گسترش خود قرار می‌دهند مشابهت داشته باشند. انتخاب دو کاربر مشابه از لحاظ اقتصادی مقرون به صرفه نیست. در واقع، انتخاب کاربران مشابه، نه تنها به گسترش اطلاعات کمک نمی‌کند، بلکه باعث کاهش عملکرد الگوریتم شناسایی کاربران پرنفوذ و افزایش پیچیدگی محاسباتی می‌شود. روش‌های کنونی شناسایی کاربران پرنفوذ اغلب قادر به حذف کاربران مشابه نیستند. لذا، ارائه روشی که به صورت کارا قادر به شناسایی کاربران با کمترین افزونگی در راستای پوشش کل شبکه، قابلیت دسترسی بالا و در نتیجه افزایش سرعت در فرآیند انتشار باشد، از چالش‌های عمده در مسئله شناسایی کاربران پرنفوذ است.

از آنجایی که مسئله شناسایی کاربران پرنفوذ از نوع NP-سخت^۴ است [۷]، ارائه یک روش دقیق برای آن به دلیل زیاد بودن مجموعه جواب محتمل، غیرممکن است. بنابراین، روش‌های تقریبی اکتشافی و فراابتکاری، راه‌حل مناسبی برای این قبیل مسائل هستند. الگوریتم‌های اکتشافی دو مشکل اصلی «گیرکردن در بهینه محلی» و «همگرایی

در نهایت، تمامی گره‌های گراف بر اساس مجموع کوتاه‌ترین فاصله آن‌ها به بهترین گره‌های محلی به دست آمده، رتبه‌بندی می‌شوند. با توجه به کارهای ذکر شده و چالش‌های موجود در آن‌ها، روش پیشنهادی در این مقاله بر اساس الگوریتم ژنتیک و بر مبنای دو هدف مؤثر ارائه می‌شود تا بتواند چالش‌های ذکر شده را تا حدودی بهبود بخشد.

در ادامه، برخی از روش‌های شناسایی گره‌های پرنفوذ در گراف شبکه‌های اجتماعی که با روش پیشنهادی مقایسه شده‌اند به طور مختصر توضیح داده می‌شوند:

- مرکزیت درجه (DC) [۸]: یک مرکزیت محلی است که در آن گره‌هایی به عنوان مؤثرترین گره‌ها انتخاب می‌شوند که بیشترین تعداد ارتباط را دارند. به عبارت دیگر، تعداد بیش‌تر دوستان متعلق به یک گره، درجه بالاتر و در نتیجه قدرت بیشتر آن را نتیجه می‌دهد. مشکل اصلی روش مرکزیت درجه، این است که در آن موقعیت گره‌ها به صورت محلی در نظر گرفته می‌شود. در این روش درجه هر گره v با استفاده از رابطه (۱) محاسبه می‌شود.

$$DC(v) = \sum_{u=1}^n a_{uv} = |\Gamma_1(v)| \quad (1)$$

که n تعداد گره‌های شبکه است، a_{uv} یک است، اگر یالی بین u و v وجود داشته باشد و در غیر این صورت صفر است. $|\Gamma_1(v)|$ تعداد گره‌های همسایه سطح اول گره v است

- مرکزیت میانی (BC) [۸]: یک مرکزیت کلی است که در آن گره‌های مؤثر آن‌هایی هستند که توسط بیش‌ترین تعداد کوتاه‌ترین مسیرها از همه گره‌ها به دیگر گره‌ها در شبکه ملاقات می‌شوند.

$$BC(v) = \sum_{s \neq v, t \in V} \frac{\sigma_{st}(v)}{\sigma_{st}} \quad (2)$$

که σ_{st} نشان‌دهنده تعداد تمامی کوتاه‌ترین مسیرهای موجود بین گره s و t است. $\sigma_{st}(v)$ نشان‌دهنده تعداد کوتاه‌ترین مسیرهای موجود بین گره s و t است که از گره v عبور می‌کنند.

- مرکزیت نزدیکی (CC) [۹، ۱۰]: یک مرکزیت کلی است که در آن رتبه هر گره به صورت مجموع کوتاه‌ترین مسیر گره مورد نظر به تمامی گره‌های گراف شبکه اجتماعی در نظر گرفته می‌شود. مرکزیت نزدیکی هر گره v از رابطه (۳) محاسبه می‌شود.

$$CC(v) = \frac{1}{\sum_{u \neq v} d(v, u)} \quad (3)$$

که در آن $d(v, u)$ طول کوتاه‌ترین مسیر بین گره v و u است.

- مرکزیت k هسته (ShellRank) [۲۸]: در این روش، ابتدا تمامی گره‌های با درجه ۱ از گراف حذف می‌شوند و به آن‌ها رتبه $k=1$ تعلق می‌گیرد، پس از حذف گره‌های با درجه ۱ و به‌روزرسانی شبکه، اگر گره‌های دیگری با درجه ۱ وجود داشته باشد، حذف و به آن‌ها نیز رتبه $k=1$ داده می‌شود. پس از حذف تمامی گره‌ها با درجه ۱، گره‌ها با درجه

بسیار زمان‌گیر است. لسکوویک و همکاران [۱۷] با ارائه یک طرح بهینه‌سازی شناخته‌شده به عنوان انتخاب روبه‌جلوی مقرون‌به‌صرفه تنبل (CELf)، در ابتدا گسترش هر گره را محاسبه می‌کنند. در این روش، در تکرارهای بعدی تنها گسترش گره‌های محدودی باید به‌روز شود. این امر باعث می‌شود CELf بسیار سریع‌تر از الگوریتم‌های حریم‌بندی باشد، زیرا تعداد تکرارها را کاهش می‌دهد. با این وجود، الگوریتم CELf هنوز هم نیاز به ۷۰۰ تکرار دارد، که آن را کند و برای گراف‌های بزرگ غیرقابل استفاده می‌کند. گوپال و همکاران [۱۸] یک نسخه سریع‌تر و کارآمدتر از الگوریتم CELf با نام CELf++ را پیشنهاد کرده‌اند که تعداد تکرارهای الگوریتم را کاهش می‌دهد.

در روش سیمپس (Simpath) [۱۹]، از پوشش رئوس^۹ گراف جهت محاسبه گسترش هر گره در مجموعه پوشش رئوس گراف و جلوگیری از محاسبه گسترش هر گره در گراف استفاده می‌شود. این الگوریتم تمام مسیرهای موجود بین مجموعه گره انتخاب‌شده و سایر گره‌های گراف ورودی را جهت به‌حداکثر رساندن نفوذ بررسی می‌کند. سیمپس از روش پوشش سطحی برای کاهش تعداد تکرارها استفاده می‌کند ولی از آنجایی که تعداد گره‌های انتخاب‌شده پس از هر تکرار الگوریتم افزایش می‌یابد، روند به‌حداکثر رساندن نفوذ در این روش کند است. اگرچه الگوریتم‌های حریم‌بندی تضمین تقریبی خوبی دارند، اما از پیچیدگی بالا [۲۰-۲۲] و مقیاس‌پذیری کم [۲۳] رنج می‌برند.

از آنجایی که رویکردهای فرااکتشافی نتایج بهتری در مقایسه با رویکردهای حریم‌بندی، قطعی و تصادفی ارائه می‌دهند، یانگ و همکاران [۲۴] الگوریتمی بر مبنای کلونی مورچه‌ها (ACO) بر اساس یک مدل احتمالاتی پارامتری جهت بررسی مسئله بهینه‌سازی نفوذ پیشنهاد داده‌اند. آن‌ها از مرکزیت درجه، مرکزیت فاصله و روش‌های شبیه‌سازی نفوذ جهت تعیین مقادیر اکتشافی استفاده کرده‌اند. در [۲۵]، یک مدل انتشار جستجوی فاخته (CSDM) پیشنهاد شده است که بر مبنای الگوریتم فرااکتشافی جستجوی فاخته است. وانگ و همکاران [۲۲] یک الگوریتم ژنتیک کدشده مبتنی بر مجموعه (SGA) را ارائه کرده‌اند که به صورت احتمالاتی به راه‌حل مطلوب مسئله^{۱۰} انتخاب زیرمجموعه گره هدف^{۱۱} همگرا می‌شود. علاوه بر این، در [۲۶] رویکردی مبتنی بر شبیه‌سازی تبرید (SA) برای مسئله حداکثرسازی نفوذ ارائه شده است که از دو روش اکتشافی جهت سرعت بخشیدن به فرآیند همگرایی SA و روش جدیدی جهت محاسبه نفوذ برای سرعت بخشیدن به الگوریتم استفاده می‌کند. نویسندگان [۲۷] الگوریتمی فراشناختی پیشنهاد داده‌اند که می‌تواند مجموعه‌های بهینه (مجموعه گره پرنفوذ) را از نتایج الگوریتم NSGA-II انتخاب کند. آن‌ها گسترش نفوذ و زمان را به عنوان اهداف بهینه‌سازی در فرآیند انتشار در نظر گرفته‌اند. در [۵] روشی جهت رتبه‌بندی و شناسایی گره‌های پرنفوذ ارائه شده است که از ساختار و ویژگی‌های محلی و کلی و بخصوص نقش پل بودن گره‌ها استفاده می‌کند. در این روش برای یافتن بهترین گره‌های محلی، ابتدا جوامع شناسایی می‌شوند؛ سپس از هر جامعه بهترین گره استخراج می‌گردد.

Input	G: graph G (V,E) S: similarity matrix N: size of population MaxIt: maximum number of generations pMutation: percentage of mutation pCrossover: percentage of crossover q: probability of activation k: number of genes in each chromosome P: seed set
Output	1: Begin algorithm 2: Generate a random population of N chromosomes (P_t) 3: Calculate value of objectives for each chromosome in the population using (6) and (8) 4: Rank population by using nondominated sorting algorithm 5: Determine crowding distance between points on each front using (10) 6: Sort population according to rank and crowding distance 7: For t=1:MaxIt 8: Apply crossover and mutation to create a new population Q_t 9: $R_t = P_t \cup Q_t$ 10: $P_{t+1} = \emptyset, i = 1$ 11: $F = \text{Rank}R_t$ by using nondominated sorting algorithm into several front F 12: While ($(P_{t+1} + F_i) < N$) 13: Determine crowding distance between points on each front (F_i) using (10) 14: $P_{t+1} = P_{t+1} \cup F_i$ 15: $i = i + 1$ 16: End while 17: Sort F_i in descending order using crowding distance 18: Choose the first $(N - P_{t+1})$ elements of F_i 19: ($P_{t+1} = P_{t+1} \cup F_i [1: N - P_{t+1}]$). 20: $t=t+1$; 21: End for 22: Return best solution in P as final seed set 23: End algorithm

شکل ۱: شبه‌کد روش پیشنهادی

۳-۱- بازنمایی و مقداردهی اولیه

روش پیشنهادی با یک جمعیت تصادفی شروع می‌شود. هر راه‌حل (کروموزوم)، به‌صورت یک بردار به طول تعداد گره‌های پرفوذ از پیش تعیین‌شده، در نظر گرفته می‌شود. شماره هر گره (ژن) در این بردار، نشان‌دهنده شماره گره است. به‌عنوان مثال، برای شناسایی ۱۰ گره پرفوذ در یک گراف شبکه اجتماعی، کروموزوم‌هایی وجود خواهند داشت که در هر کدام از آن‌ها ۱۰ ژن وجود دارد. نحوه نمایش ساختار کروموزومی به طول ۱۰ گره به‌صورت شکل ۲ است. گره‌های با شماره‌های مشخص‌شده یک مجموعه هدف (seed set) را تشکیل می‌دهند.

۸	۱۰	۸۵	۳	۲۲	۱۵	۷۳	۵۲	۳۳	۴۳
---	----	----	---	----	----	----	----	----	----

شکل ۲: کروموزوم

۳-۲- محاسبه برازندگی

در این مرحله، الگوریتم به هر جواب، یک مقدار برازندگی اختصاص می‌دهد، که این مقدار نشان‌دهنده اهمیت آن راه‌حل است. عدد برازندگی که مقدار هزینه مربوط به هر کدام از کروموزوم‌ها است با توجه به توابع

از گراف حذف شده و به آن‌ها رتبه $k=2$ تعلق می‌گیرد و مجدداً پس از به‌روزرسانی، حذف و اعطای رتبه ۲ به دیگر گره‌های درجه ۲، روند مذکور برای گره‌های با درجه ۳ و بالاتر ادامه می‌یابد تا زمانی که تمامی گره‌های شبکه حذف و رتبه‌بندی شوند.

- مرکزیت بردار ویژه^{۱۱} (EC): در این معیار مرکزیت، یک گره زمانی یک گره با اهمیت محسوب می‌شود که به دیگر گره‌های مهم در شبکه لینک داشته باشد. در این روش به هر گره امتیازی بر اساس مجاورت آن با گره‌های دارای امتیاز بالا اختصاص می‌یابد [۲۹].

- مرکزیت محلی^{۱۲} (LC): یک روش مرکزیت نیمه‌محلی است که تعداد اولین و دومین نزدیک‌ترین همسایه‌ها را در محاسبات خود در نظر می‌گیرد [۳۰]. مرکزیت محلی هر گره v به‌صورت رابطه (۴) محاسبه می‌شود.

$$LC(v) = \sum_{u \in \Gamma_1(v)} \sum_{w \in \Gamma_1(u)} N(w) \quad (4)$$

که در آن $N(w)$ ، تعداد اولین و دومین نزدیک‌ترین همسایه‌های گره w است. $\Gamma_1(u)$ مجموعه نزدیک‌ترین گره‌های همسایه u است.

- مرکزیت ساختاری محلی^{۱۳} (LSC): این روش، هم تعداد و هم توپولوژی ارتباطات همسایه‌های هر گره را در نظر می‌گیرد [۳۱]. مرکزیت ساختاری محلی هر گره v به‌صورت رابطه (۵) تعریف می‌شود.

$$LSC(v) = \sum_{u \in \Gamma_1(v)} \left(\alpha N(u) + (1-\alpha) \sum_{w \in \Gamma_2(u)} c_w \right) \quad (5)$$

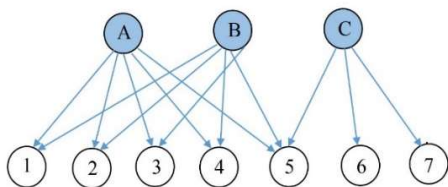
که در آن $N(u) = |\Gamma_2(u)|$ تعداد اولین و دومین نزدیک‌ترین همسایه‌های گره u است، c_w ضریب خوشه‌بندی محلی گره w است و α یک پارامتر تنظیمی بین ۰ و ۱ است

۳- روش پیشنهادی

در این بخش، ابتدا بازنمایی مناسبی برای کروموزوم ارائه می‌گردد. سپس، برازندگی هر کروموزوم (راه‌حل) با تعاریف مناسبی برای «سود ناشی از مجموعه گره‌های موجود در هر راه‌حل» به‌عنوان اولین تابع هدف و «شباهت بین گره‌ها» به‌عنوان تابع هدف دوم به‌دست می‌آید که در آن ارتباط بین گسترش نفوذ گره‌ها با شباهت میان گره‌ها توضیح داده می‌شود. دو تابع هدف ذکرشده در بهینه‌سازی چند هدفه اعمال می‌شوند و عملگرهای ژنتیکی بر روی هر راه‌حل اعمال می‌گردند. در روش پیشنهادی، الگوریتم NSGA-II برای یافتن راه‌حل‌ها مورد استفاده قرار می‌گیرد و قدرت نفوذ مورد انتظار هر یک از کروموزوم‌ها با استفاده از مدل انتشار SIR محاسبه می‌شود. در نهایت، در میان تمام راه‌حل‌های موجود در جبهه پرتو^{۱۴} [۳۲]، یک راه‌حل با حداکثر توانایی گسترش، به‌عنوان مجموعه گره نهایی انتخاب می‌شود. مراحل روش پیشنهادی در بخش‌های زیر با جزئیات بیشتر توضیح داده می‌شود. در شکل ۱ شبه‌کد الگوریتم پیشنهادی نشان داده شده است.

۲-۳-۲- تابع هدف دوم

در فرآیند بازاریابی ویروسی به کمک شبکه‌های اجتماعی، با در نظر گرفتن تعامل کاربران و ارتباطات آن‌ها، شباهت میان کاربران از نظر هم‌پوشانی کاربرانی که توسط آن‌ها فعال شده‌اند از اهمیت خاصی برخوردار است. لذا کاربران هدف باید طوری انتخاب شوند که کمترین شباهت ممکن را به هم دارند به این مفهوم که تا حد ممکن از هم فاصله داشته باشند، طوری که قادر باشند کل شبکه را پوشش دهند. هم‌چنین یکی از مواردی که میزان شباهت کاربران در آن تأثیرگذار است، سرعت انتشار در شبکه است. به این صورت که در مرحله زمانی t ، دو کاربر که از لحاظ قدرت گسترش شبیه به هم هستند ولی کاربران یکسانی را از لحاظ فریدیت فعال می‌کنند، مناسب نبوده و بهتر است کاربر دیگری با همان قدرت گسترش و شباهت کمتر انتخاب شود. به‌عنوان مثال شکل ۴ را در نظر بگیرید. دو کاربر A و B در مرحله زمانی یکم هر کدام حداکثر می‌توانند پنج کاربر به شماره‌های ۱ و ۲ و ۳ و ۴ و ۵ را فعال کنند. میزان گسترش دو کاربر A و B یکسان و بیشتر از کاربر C است، ولی اگر این دو کاربر به‌عنوان گره‌های پرنفوذ انتخاب شوند، در مجموع حداکثر پنج کاربر را می‌توانند فعال کنند. اگر از میان دو کاربر مشابه A و B به‌عنوان مثال کاربر A انتخاب گردد، سپس به‌جای کاربر B که قدرت گسترش مشابه A و درعین حال بیشتر نسبت به C دارد، کاربر C که حداکثر ۳ کاربر را می‌تواند فعال کند انتخاب شود، در این صورت با انتخاب A و C، در مجموع قادر خواهیم بود هفت کاربر را فعال کنیم، در حالی که با انتخاب A و B نهایتاً فقط پنج کاربر فعال می‌شوند. در نتیجه انتخاب کاربران با در نظر گرفتن شباهت میان آن‌ها می‌تواند به افزایش میزان گسترش در شبکه منجر شود.



شکل ۴: تفاوت قدرت نفوذ میان کاربران

همان‌طور که گفته شد، انتخاب گره‌های دور از هم، سرعت انتشار در شبکه را افزایش می‌دهد و باعث می‌شود هم‌پوشانی گره‌ها از لحاظ فریدیت گره‌هایی که فعال می‌کنند کمتر باشد؛ به طوری که بتوانند سراسر شبکه را پوشش دهند. لذا، میزان شباهت دوبره‌دوی گره‌های هر کروموزوم، به‌عنوان هدف دوم در نظر گرفته می‌شود. در این بخش یک معیار شباهت دوسطحی وزن‌دار بر روی گراف‌های ساده و بدون جهت به‌عنوان تابع هدف دوم تعریف می‌گردد که در ادامه توضیح داده می‌شود. برای به‌دست‌آوردن میزان شباهت گره‌ها در گراف‌های بدون جهت، یک روش نیمه‌محلی وزن‌دار ارائه می‌گردد. با توجه به این که میزان نفوذ هر گره با افزایش سطح‌ها کم و کمتر می‌شود، به‌همین دلیل فقط دو سطح همسایگی کاربران برای محاسبه میزان شباهت گره‌ها در نظر

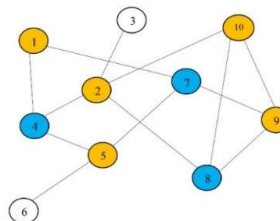
هدف مشخص می‌شود. چون در هر مرحله دو تابع هدف مدنظر است، لذا برای هر کروموزوم دو عدد به‌دست می‌آید که بیانگر میزان شایستگی کروموزوم نسبت به دو تابع هدف در نظر گرفته‌شده است. در روش پیشنهادی، میزان سود به‌دست‌آمده از گره‌ها و شباهت بین گره‌ها به‌عنوان توابع هدف در نظر گرفته می‌شوند.

۱-۲-۳- تابع هدف اول

تابع هدف اول میزان سود به‌دست‌آمده از هر کروموزوم در نظر گرفته می‌شود. اختلاف هزینه انتخاب گره‌ها از درآمد حاصل از انتخاب آن‌ها، مقدار سود حاصل از هر کروموزوم را نتیجه می‌دهد. درآمد نیز به‌عنوان تعداد گره‌هایی که تحت‌تأثیر گره‌های موجود در کروموزوم فعال می‌شوند، در نظر گرفته می‌شود. یک گره در گراف که جزء همسایه‌های سطح اول گره‌های موجود در کروموزوم باشد، با احتمال q توسط گره‌های موجود در کروموزوم فعال می‌شود، در غیر این صورت، غیرفعال می‌ماند. شکل ۳ را به‌عنوان یک گراف نمونه در نظر بگیرید. فرض کنید یک کروموزوم شامل سه گره به شماره‌های v_4 و v_7 و v_8 باشد. در ابتدا این سه گره فعال هستند. اکنون باید دید که چه تعداد گره می‌توانند توسط این سه گره فعال شوند. تمامی همسایه‌های آن‌ها که شامل گره‌های $v_1, v_2, v_3, v_5, v_6, v_9, v_{10}$ هستند، با احتمال q شانس فعال شدن دارند. فرض کنید که اگر چند گره قادر به فعال‌سازی یک گره باشند، هر کدام از آن‌ها به ترتیب شانس خود را امتحان می‌کنند. به‌عنوان مثال، در شکل موردنظر، گره v_5 می‌تواند توسط هر کدام از گره‌های v_4 و v_7 با احتمال q فعال شود. سرانجام، پس از اتمام فرآیند فعال‌سازی، تعدادی و یا تمامی گره‌های $v_1, v_2, v_3, v_5, v_6, v_9, v_{10}$ که در معرض ابتلا بودند، فعال می‌شوند. فرض می‌شود که به‌ازای گره‌های فعال موجود در کروموزوم می‌بایست هزینه‌ای پرداخت شود و از گره‌هایی که در نتیجه فعال‌سازی آن‌ها فعال می‌شوند، درآمدی به‌دست می‌آید. در نتیجه، می‌توان تابع هدف اول را به‌عنوان سود به‌دست‌آمده از هر کروموزوم (p_{cr}) به‌صورت رابطه (۶) فرموله کرد که می‌بایست به‌شینه شود.

$$p_{cr} = r \sum_{u \in I'} y_u - ck \quad (6)$$

در این رابطه r درآمد به‌دست‌آمده از فعال‌سازی یک گره است، y_u برابر با ۱ است اگر گره u توسط یکی از گره‌های موجود در کروموزوم فعال شود و در غیر این صورت صفر است. c هزینه هر گره کروموزوم و k اندازه کروموزوم است.



شکل ۳: نحوه فعال‌سازی گره‌ها در شبکه ساختگی

۳-۳- رتبه‌بندی جمعیت

زمانی که یک الگوریتم تک‌هدفه مطرح است، معیار برتری جواب‌ها نسبت به هم بسیار ساده و بدیهی است؛ زیرا تنها یک تابع هدف مدنظر است. به‌عنوان مثال، در صورتی که مسئله مورد بحث یک مسئله کمینه‌سازی باشد، جوابی که کمترین مقدار تابع هدف را دارا باشد، مطلوب است و بر سایر جواب‌ها برتری دارد. اما زمانی که برای حل مسئله‌ای از یک الگوریتم چندهدفه استفاده شود، به این معناست که حداقل دو تابع هدف مدنظر است و دیگر به‌آسانی نمی‌توان در مورد بعضی از جواب‌ها نظر قطعی داد. در اکثر موارد، نقاطی یافت می‌شود که هیچ‌کدام بر دیگری برتری کامل ندارند و نمی‌توان با مفهوم غلبه، دوه‌دو بین آن‌ها مقایسه‌ای انجام داد. لذا برای به‌دست‌آوردن بهترین جواب‌ها، باید آن‌ها را بر اساس معیاری مرتب کرد که در اینجا معیار، الگوریتم مرتب‌سازی نامغلوب [۱۴] است. در این الگوریتم به هر راه‌حل یک رتبه اختصاص داده می‌شود که بر اساس تعداد مغلوبان آن راه‌حل نسبت به سایر نقاط، انجام می‌شود. در پایان الگوریتم، نقاطی که بهترین رتبه را دارا باشند، به‌عنوان مجموعه جواب یا نقاط جبهه پرتو انتخاب می‌شوند. لذا، هدف رتبه‌بندی جمعیت، اختصاص دادن رتبه به هر راه‌حل موجود در جمعیت به‌منظور یافتن راه‌حل‌های نامغلوب است. سپس برای هر عضو در هر جبهه، معیار فاصله ازدحامی محاسبه می‌شود که بیانگر اندازه‌ای از نزدیکی نمونه مورد نظر به دیگر اعضای جمعیت است. به این صورت که ابتدا برای هر تابع هدف، نتایج مرتب می‌شود. این معیار تراکم، راه‌حل‌ها را به‌وسیله محاسبه میانگین فاصله بین هر دو نقطه تخمین می‌زند. از نظر ریاضی عملگر فاصله ازدحامی برای نقطه i در مسئله‌ای دو هدفه به‌صورت شکل ۵ است.

جدول ۱: ویژگی‌های اصلی مجموعه داده‌های واقعی

Network	n	m	$\langle K \rangle$	k_{max}	$\langle C \rangle$	$\langle d \rangle$	H	β_{rand}^c
Email	1133	5451	9.62	71	0.220	3.6028	4.1465	0.0535
PGP	10680	24316	4.55	205	0.440	7.4855	4.1465	0.0530
Ca-HepTh	9877	25998	5.748	65	0.471	5.9454	2.2624	0.0769

فاصله ازدحامی برای راه‌حل i به‌صورت رابطه (۱۰) تعریف می‌شود [۱۳].

$$d_i^1 = \frac{|f_1^{i+1} - f_1^{i-1}|}{f_1^{max} - f_1^{min}}, d_i^2 = \frac{|f_2^{i+1} - f_2^{i-1}|}{f_2^{max} - f_2^{min}} \quad (10)$$

$D = d_i^1 + d_i^2$
در رابطه (۱۰) مقادیر مربوط به f_1^{i+1}, f_1^{i-1} و f_2^{i+1}, f_2^{i-1} در شکل مشخص شده است و d_i^1 نسبت ناحیه مربوط به نقطه i با توجه به تابع هدف f_1 و d_i^2 نسبت ناحیه مربوط به همین نقطه با توجه به تابع هدف f_2 می‌باشد و D که

گرفته شده است. میزان شباهت وزن‌دار دوسطحی گره u و گره v به‌صورت رابطه (۷) بازنمایی می‌شود.

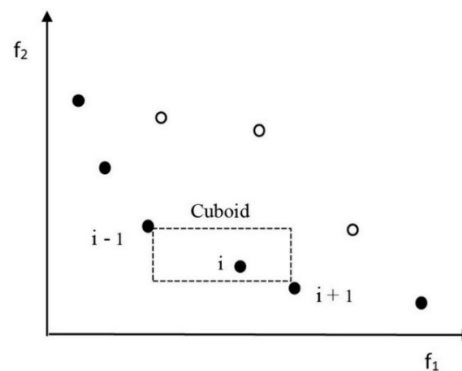
$$Similarity(u, v) = \alpha \frac{|\Gamma_1(u) \cap \Gamma_1(v)|}{|\Gamma_1(u) \cup \Gamma_1(v)|} + \beta \frac{|\Gamma_2(u) \cap \Gamma_2(v)|}{|\Gamma_2(u) \cup \Gamma_2(v)|} \quad (7)$$

در رابطه (۷)، $\Gamma_1(u)$ همسایه‌های سطح اول گره u است (نزدیک‌ترین همسایه‌های u) و $\Gamma_2(u)$ همسایه‌های سطح دوم گره u است (همسایه همسایه‌های u). $|\Gamma_1(u) \cap \Gamma_1(v)|$ تعداد همسایه‌های مشترک سطح اول گره u و v است. طبق رابطه ارائه‌شده، شباهت هر گره به خود ۱ می‌شود. هم‌چنین، با توجه به اینکه گراف بدون جهت است، شباهت گره‌ها متقارن است. به این صورت که میزان شباهت گره u به v با میزان شباهت گره v به u برابر است. به‌طور کلی، شباهت دو گره، مجموع وزن‌دار تعداد همسایه‌های مشترک سطح اول دو گره بر تعداد کل همسایه‌های سطح دوم دو گره در نظر گرفته می‌شود. α و β به‌عنوان مقادیری ثابت جهت وزن‌دهی به سطح اول و دوم هستند. از آنجایی که تأثیر مستقیم کاربران بر روی یکدیگر بیشتر است، لذا مقدار α را در آزمایش‌های بزرگ‌تر از β خواهد بود. در نهایت، تابع هدف دوم به‌عنوان میزان شباهت دوه‌دوی گره‌های هر کروموزوم (S_{cr}) به‌صورت رابطه (۸) فرموله می‌شود که می‌بایست کمینه گردد.

$$S_{cr} = \sum_u \sum_v similarity(u, v) k_u k_v, \quad u, v \in cr \quad (8)$$

در نهایت، مدل دو هدفه با در نظر گرفتن دو تابع هدف ذکر شده، به‌صورت رابطه (۹) تعریف می‌شود.

$$\begin{aligned} \max \quad & p_{cr} = r \sum_{u \in cr} y_u - ck \\ \min \quad & S_{cr} = \sum_u \sum_v similarity(u, v) k_u k_v, \quad u, v \in cr \end{aligned} \quad (9)$$

شکل ۵: مفهوم فاصله ازدحامی برای نقطه i [۱۴]

از بین فرزندان حاصل از تقاطع، تعدادی برای عمل جهش انتخاب می‌شوند. عمل جهش برای فرار از بهدام‌افتادن الگوریتم در بهینه محلی بسیار مفید است. در عمل جهش، هر گره در کروموزوم فرزندان به صورت تصادفی با احتمال مشخص شده به گره‌های انتخاب‌نشده تغییر می‌کند. بعد از تولید فرزندان حاصل از تقاطع و جهش، میزان برازندگی آن‌ها محاسبه می‌شود. در ادامه، این جمعیت با جمعیت اصلی ادغام می‌شود. در این مرحله، از بین جمعیت والدین و فرزندان باید تعدادی حذف شوند تا تعداد جمعیت اصلی ثابت بماند. لذا، مجدداً اعضای جمعیت جدید همانند مرحله قبل براساس رتبه و فاصله ازدحامی مرتب می‌شوند. حال اعضای جمعیت در درجه اول برحسب رتبه و در درجه دوم برحسب فاصله ازدحامی مرتب شده‌اند. سپس به تعداد اعضای جمعیت اصلی، اعضای از بالای جمعیت مرتب‌شده انتخاب و بقیه اعضا از جمعیت حذف می‌شوند. اعضای انتخاب‌شده جمعیت، نسل بعدی را تشکیل می‌دهند و این روال تا رسیدن به تعداد تکرار مشخص ادامه می‌یابد. در نهایت، جواب‌های نامغلوب به دست می‌آیند و مجموعه گره‌هایی که بهترین عملکرد را داشته باشند، به عنوان خروجی نهایی الگوریتم در نظر گرفته می‌شوند.

۴- ارزیابی و آزمایش‌ها

در این بخش، روش پیشنهادی بر روی مجموعه داده‌های مختلف ارزیابی می‌شود. مجموعه جواب‌های جبهه پرتوی به دست آمده در هر مجموعه داده با مدل انتشار SIR [۳۳] ارزیابی و جواب با بیشترین میزان قدرت گسترش انتشار به عنوان جواب نهایی انتخاب می‌شود. در ادامه، تنظیم پارامترها و نتایج آزمایش‌ها و ارزیابی‌ها از لحاظ میزان سود، شباهت میان گره‌ها و قدرت انتشار بررسی می‌شوند و با سایر الگوریتم‌ها مقایسه می‌شوند.

۴-۱- مجموعه داده‌ها

روش پیشنهادی بر روی شبکه‌های واقعی بدون جهت و بدون وزن، ارزیابی گردیده و کارایی آن با چندین الگوریتم شناخته شده مقایسه می‌شود. شبکه‌های استفاده شده را می‌توان از سایت <http://snap.stanford.edu/data/> دانلود کرد که شامل سه مجموعه داده زیر است:

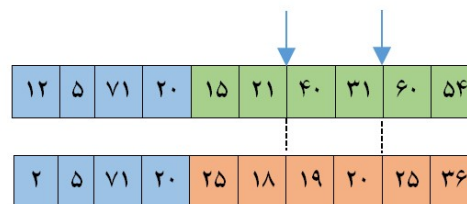
- Email: شبکه تبادل ایمیل بین اعضاء دانشگاه Rovira
- PGP: یک شبکه ارتباطی رمزگذاری شده است. الگوریتم‌های Pretty-Good-Privacy به منظور حفظ حریم خصوصی بین افراد توسعه یافته‌اند، که به آن وب اعتماد PGP نیز گفته می‌شود.
- Ca-HepTh: شبکه همکاری مابین همکاران علمی، مؤلفان مقالات ثبت شده در گروه تئوری فیزیک با انرژی بالا.

در جدول ۱ ویژگی‌های اصلی توپولوژی این شبکه‌ها شامل اندازه شبکه n ، تعداد یال‌ها m میانگین درجه $\langle K \rangle$ ، بیشترین درجه k_{max} ، میانگین ضریب خوشه‌بندی $\langle C \rangle$ ، میانگین طول کوتاه‌ترین مسیر $\langle d \rangle$ ، درجه

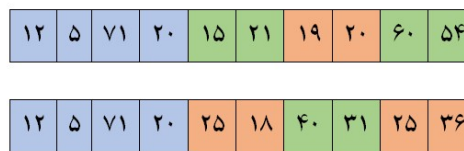
مجموع این دو نسبت است، فاصله ازدحامی نامیده می‌شود. هر نقطه‌ای که مقدار فاصله ازدحامی بیش‌تری داشته باشد، به این معنی است که محدوده بیش‌تری را پوشش می‌دهد و حذف آن منجر به از دست رفتن تنوع جواب در محدوده وسیعی از پاسخ‌ها می‌شود. رابطه (۱۰) برای هر مسئله با چندین تابع هدف نیز قابل تعمیم می‌باشد.

۴-۳- عملگر تقاطع و جهش

برای انجام عمل تقاطع، باید دو والدین در هر مرحله، انتخاب شوند. انتخاب در این مرحله به این دلیل که هر کروموزوم دارای دو مقدار برازندگی است، با الگوریتم ژنتیک تک هدفه متفاوت است. در این مرحله، از روش انتخاب رقابتی دودویی^{۱۵} برای انتخاب هر والد استفاده می‌شود. در این روش برای انتخاب هر والد، دو راه‌حل به تصادف از میان جمعیت موجود انتخاب شده و پس از مقایسه راه‌حل‌های انتخاب‌شده، بهترین آن‌ها انتخاب می‌شود. معیارهای انتخاب در درجه اول، رتبه جواب و در درجه دوم، فاصله ازدحامی جواب است. هرچه رتبه جواب کمتر باشد و دارای فاصله ازدحامی بیشتر باشد، مطلوب‌تر است. به بیان دیگر، برای انتخاب هر والد، ابتدا دو کروموزوم X و Y به صورت تصادفی انتخاب می‌شوند. سپس در صورتی که رتبه X کمتر از رتبه Y باشد، کروموزوم X به عنوان والد انتخاب می‌شود و اگر رتبه هر دو یکی باشد، آنگاه کروموزومی که فاصله ازدحامی بیشتری داشته باشد، انتخاب می‌شود. این عمل به تعداد والدین مورد نیاز جهت تقاطع تکرار می‌شود. برای عمل تقاطع، گره‌های مشترک دو والد مستقیماً به کروموزوم‌های فرزندان منتقل می‌شود و سپس گره‌های متفاوت بین دو والد به وسیله روش تقاطع دوقطه‌ای ادغام می‌شوند. شکل ۶ عملکرد تقاطع را نشان می‌دهد. در هر مرحله و والد انتخاب شده و با استفاده از تقاطع دومرحله‌ای عمل تقاطع بر اساس احتمال تقاطع انجام می‌شود. شکل ۶ (الف) والدین را نشان می‌دهد که در نتیجه عملگر تقاطع، فرزندان نشان داده شده در قسمت (ب) به دست آمده‌اند.



(الف)



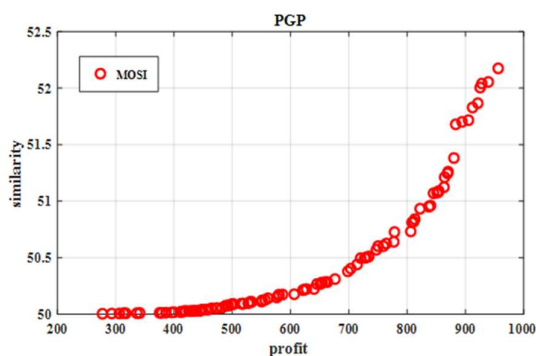
(ب)

شکل ۶: نمونه‌ای از عملکرد تقاطع روی کروموزوم‌ها

زمانی پایان می‌یابد که دیگر هیچ گره مبتلایی در شبکه وجود نداشته باشد. قدرت انتشار گره v ، S_v^β تعداد گره‌هایی است که در انتهای فرآیند انتشار توسط گره v مبتلا شده است. در نهایت لیستی رتبه‌بندی شده مطابق قدرت انتشار گره‌ها در شبکه به دست می‌آید.

۴-۴- نتایج آزمایش‌ها

در این بخش، نتایج اجرای الگوریتم MOSI پیشنهادی بر روی مجموعه داده‌های ذکر شده نشان داده می‌شود. جبهه پرتوی به دست آمده بر روی مجموعه داده PGP در شکل ۷ نشان داده شده است. در ادامه، آزمایش‌های مختلفی بر اساس مدل SIR انجام می‌گیرد. با توجه به اینکه آستانه ابتلا (β_{rand}^c) طبق جدول ۱، در شبکه Email مقدار 0.0535 ، در PGP برابر با 0.053 و در Ca-HepTh مقدار 0.0769 را دارد. در این بخش از مقدار 0.1 در مدل SIR برای هر سه شبکه استفاده شده است. بنابراین، نرخ مبتلا شدن گره‌ها، کوچک باقی می‌ماند. شایان ذکر است که اگر β مقدار بزرگی داشته باشد، هویت و نقش گره‌ها اهمیت چندانی ندارد. در این حالت فرآیند انتشار، کل شبکه را در برمی‌گیرد. روش پیشنهادی با VC++ 2013 و بر روی سیستمی با مشخصات Intel® Core™ i5 CPU 2.3 GHz و ۱۶ گیگابایت حافظه شبیه‌سازی و اجرا شد.



شکل ۷- جبهه پرتوی حاصل از اجرای MOSI در PGP

مقادیر انتشار به دست آمده با مدل انتشار اپیدمی SIR از میانگین ۱۰۰۰ اجرا بر روی سیستم با مشخصات ذکر شده به دست آمده‌اند. در ارزیابی‌ها، روش پیشنهادی با چند الگوریتم BC, DC, EC, LC, LSC, CC و shellrank و BridgeRank مقایسه شده است. ابتدا تأثیر مجموعه گره‌ها در اندازه‌های مختلف از لحاظ قدرت گسترش آن‌ها بررسی می‌گردد. شکل ۸ (الف) میزان نفوذ مجموعه گره‌های یافت شده را با تغییر اندازه از ۱ تا ۵۰، با روش‌های مختلف بر روی شبکه PGP نشان می‌دهد. همان‌طور که دیده می‌شود، MOSI به نسبت سایر الگوریتم‌ها کارایی بهتری دارد. از تعداد گره‌های ۱۰ به بالا، تفاوت روش پیشنهادی با سایر روش‌ها و کارایی بهتر آن به وضوح قابل مشاهده است. شکل ۸ (ب)، شبیه‌سازی نتایج تجمعی تعداد گره‌های مبتلا شده، ($F(t)$)، به صورت تابعی از زمان (t) در شبکه مذکور را نشان می‌دهد. با توجه به شکل ۸ می‌توان نتیجه گرفت که تعداد تجمعی گره‌های مبتلا شده با افزایش t

ناهمگونی $H = \frac{\langle k^2 \rangle}{\langle k \rangle^2}$ و آستانه ابتلا $\beta_{rand}^c = \frac{\langle k \rangle}{\langle k^2 \rangle}$ نمایش داده شده است.

۴-۲- تنظیم پارامترها

در این قسمت، پارامترهای تنظیمی الگوریتم، شامل اندازه جمعیت، تعداد تکرار، درصد ترکیب، درصد جهش، تعداد مجموعه گره‌های پرنفوذ k ، که در این بخش با تعداد ژن‌های هر کروموزوم مشخص می‌شود، و مقادیر α و β که به ترتیب میزان اهمیت همسایه‌های سطح اول و سطح دوم هر گره را در فرمول محاسبه شباهت میان گره‌ها در رابطه (۲) نشان می‌دهند. دی‌ژانگ [۳۴] نشان داد که برای این که الگوریتم ژنتیک عملکرد مناسبی داشته باشد، بهتر است که پارامتر نرخ تقاطع دارای مقدار نسبتاً بالایی باشد. به همین منظور در روش پیشنهادی، احتمال تقاطع برابر 0.8 مقداردهی شده است. هزینه انتخاب هر گره و درآمد به دست آمده از فعال سازی گره‌ها توسط انتخاب گره‌ها (گره‌های هر کروموزوم) و سایر پارامترهای الگوریتم پیشنهادی MOSI مطابق جدول ۲ تنظیم می‌شوند.

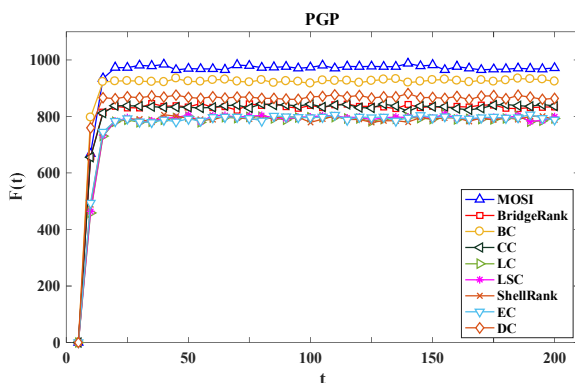
جدول ۲: تنظیم پارامترهای MOSI

مقدار	نماد	پارامتر
۱۰۰	N	اندازه جمعیت
۱۰۰	MaxIt	حداکثر تعداد نسل (تکرار)
۰/۲	pMutation	احتمال جهش
۰/۸	pCrossover	احتمال تقاطع
۰/۷	α	اهمیت نسبی همسایگان سطح اول
۰/۳	β	اهمیت نسبی همسایگان سطح دوم
۰/۱	Mu	نرخ جهش
۰/۷	q	احتمال فعال شدن
۰/۱	c	هزینه انتخاب گره
۱	r	درآمد حاصل از یک گره فعال
۵۰	k	تعداد ژن‌ها در هر کروموزوم
۱	r	سود بازگشتی از یک مشتری

۴-۳- مدل SIR

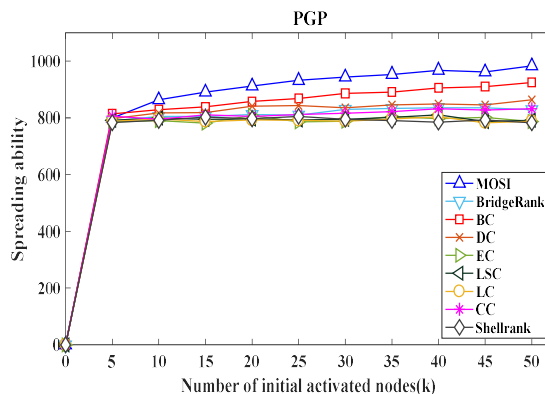
مدل SIR [۳۴] یکی از مدل‌های انتشار اپیدمی در شبکه‌های پیچیده است. در این مدل افراد در ۳ حالت قرار می‌گیرند: ۱- در معرض ابتلا $S(t)$ ، نشان‌دهنده تعداد افرادی است که در معرض ابتلا به بیماری هستند ولی هنوز مبتلا نشده‌اند. ۲- مبتلا شده $I(t)$ ، نشان‌دهنده تعداد افرادی است که مبتلا شده‌اند و قادر هستند بیماری را به افراد در معرض ابتلا سرایت دهند. ۳- حذف شده‌ها $R(t)$ ، نشان‌دهنده افرادی است که قبلاً مبتلا شده‌اند و هرگز دوباره مبتلا نخواهند شد. در مدل SIR در ابتدا همه گره‌ها در معرض ابتلا در نظر گرفته می‌شوند (S) به جز یک گره که مبتلا شده است (I). در هر مرحله زمانی، گره‌های مبتلا شده، همسایه‌های در معرض ابتلا خود را با احتمال β مبتلا می‌کنند و خود با احتمال ۱ وارد مجموعه حذف شده‌ها (R) می‌شوند و دیگر از ابتلا به بیماری مصون می‌مانند و نمی‌توانند دوباره مبتلا شوند. فرآیند انتشار

آمده بر روی شبکه PGP، روش پیشنهادی نسبت به سایر روش‌ها و به‌ویژه در مقایسه با EC و Shellrank، عملکرد بسیار بهتری دارد.



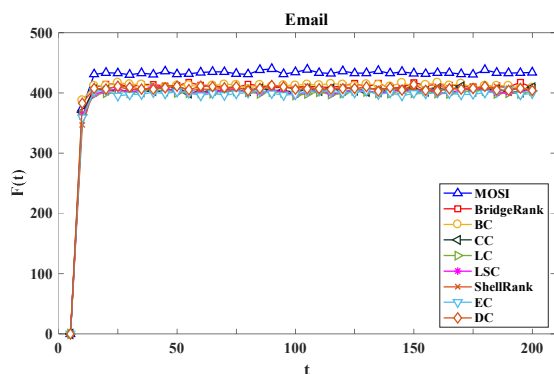
(ب)

افزایش یافته و در نهایت به مقدار ثابتی می‌رسد. در این نوع ارزیابی تعداد ۵۰ گره برتر هر الگوریتم استفاده شده است که با توجه به نتایج به‌دست

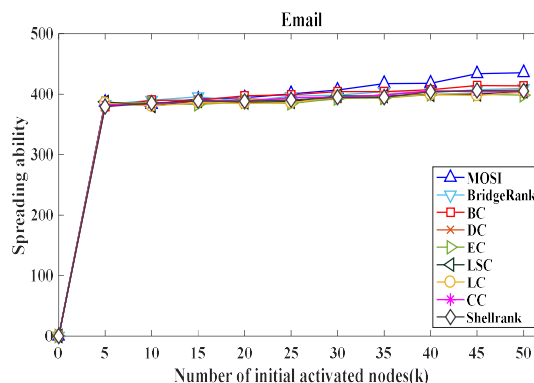


(الف)

شکل ۸: مقایسه قدرت گسترش الگوریتم پیشنهادی و سایر الگوریتم‌ها در PGP: (الف) مجموعه گره‌ها در اندازه‌های مختلف از ۱ تا ۵۰، (ب) تابعی از زمان



(ب)



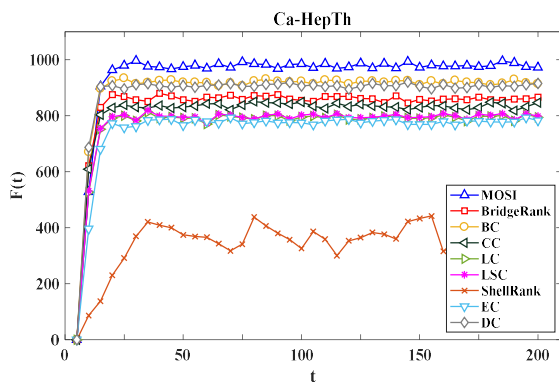
(الف)

شکل ۹: مقایسه قدرت گسترش الگوریتم پیشنهادی و سایر الگوریتم‌ها در Email: (الف) قابلیت گسترش مجموعه گره‌ها در اندازه‌های مختلف از ۱ تا ۵۰، (ب) تابعی از زمان

در ادامه، روش پیشنهادی از لحاظ توابع هدف به‌صورت جداگانه مورد بررسی قرار گرفته و با سایر الگوریتم‌ها مقایسه می‌شود. شکل ۱۱ مقادیر تابع هدف اول، تحت الشعاع سود به‌دست‌آمده از مجموعه ۵۰ گره برتر شناسایی شده توسط تمامی الگوریتم‌ها در هر سه شبکه Email، PGP و Ca-HepTh را نشان می‌دهد. در مجموعه داده Email، سود به‌دست‌آمده از الگوریتم MOSI پیشنهادی از الگوریتم‌های EC، KS، CC، LC و LSC بیشتر است و به نسبت الگوریتم‌های DC و BC کمتر است. دلیل اینکه روش پیشنهادی سود کمتری نسبت به روش DC دارد این است که در روش Degree Centrality گره‌های با درجه بالاتر که دارای تعداد همسایه‌های بیشتری هستند، گره‌های مهم‌تری هستند. بنابراین، ۵۰ گره برتر شناسایی شده با DC، تعداد گره‌های همسایه سطح اول بیشتری دارند. در نتیجه، با توجه به تعریف سود در روش پیشنهادی، در DC سود بیشتری از ۵۰ نود برتر آن حاصل می‌شود. در شبکه PGP

در شکل ۹ (الف) میزان نفوذ مجموعه گره‌های انتخاب شده (از ۱ تا ۵۰ گره)، بر روی شبکه Email با روش‌های مختلف مقایسه شده است. همان‌طور که مشخص است، روش پیشنهادی به نسبت سایر الگوریتم‌ها قابلیت گسترش بیشتری دارد. به‌ویژه اگر مجموعه انتخابی بیش از ۳۵ گره داشته باشد، تفاوت روش پیشنهادی با سایر روش‌ها بهتر مشاهده می‌شود. شکل ۹ (ب)، تعداد تجمعی گره‌های مبتلا شده $(F(t))$ ، به‌صورت تابعی از زمان را نشان می‌دهد. در اینجا تعداد ۵۰ گره برتر در هر الگوریتم به‌عنوان هسته اولیه استفاده شده است. نتایج نشان می‌دهد در شبکه Email، روش پیشنهادی نسبت به سایر روش‌ها عملکرد بهتری داشته و هسته انتخاب شده، تعداد گره‌های بیشتری را تحت تأثیر قرار می‌دهد. همچنین برتری الگوریتم پیشنهادی در شکل ۱۰ بر روی شبکه Ca-HepTh به‌وضوح دیده می‌شود.

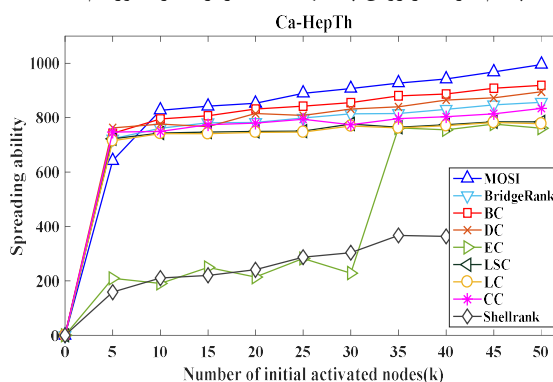
در $Ca-HepTh$ و PGP میزان شباهت میان ۵۰ گره برتر شناسایی شده با روش $ShellRank$ به نسبت سایرین بسیار زیاد است و در شبکه $Email$ روش LSC در جایگاه نخست قرار دارد. به علاوه در شبکه های $Ca-HepTh$ و $HepTh$ روش $BridgeRank$ دارای کمترین شباهت است و روش پیشنهادی بسیار نزدیک به آن در رتبه دوم قرار دارد. همان طور که قبلاً ذکر شد، میزان شباهت گره های انتخابی در میزان گسترش نفوذ آن ها در گراف شبکه اجتماعی مؤثر است. به این صورت که شباهت کمتر منجر به داشتن مجموعه گره های با قدرت انتشار بیشتر می شود.



(ب)

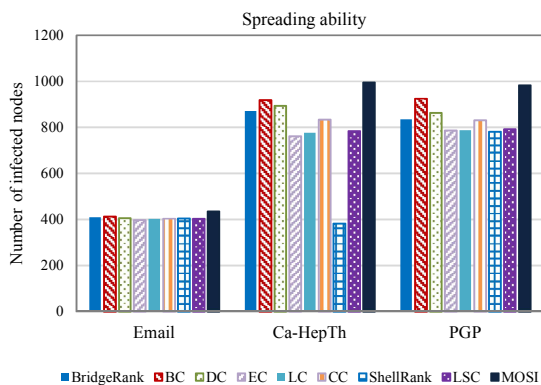
نیز، ۵۰ گره برتر شناسایی شده با روش پیشنهادی به نسبت EC ، CC و LC ، LSC سود بسیار بیشتری به دست آورده است ولی در مقایسه با BC و DC میزان سود $MOSI$ کمتر است. در $Ca-HepTh$ روش پیشنهادی بعد از DC ، CC و BC در رتبه چهارم قرار دارد.

شکل ۱۲ مقادیر توابع هدف دوم، تحت عنوان میزان شباهت دوهوی ۵۰ گره برتر شناسایی شده با الگوریتم های مختلف را نشان می دهد. همان طور که قابل مشاهده است، در PGP میزان شباهت گره ها با اختلاف چشم گیری در روش پیشنهادی کمتر از سایر الگوریتم ها است.



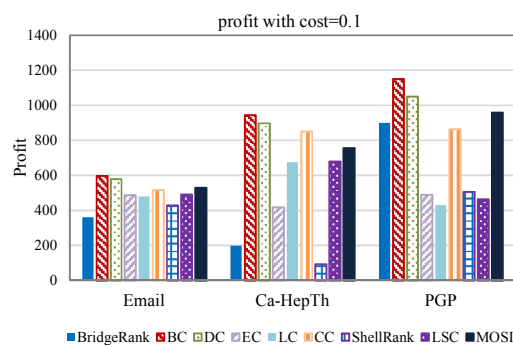
(الف)

شکل ۱۰: مقایسه قدرت گسترش الگوریتم پیشنهادی و سایر الگوریتم ها در $Email$: (الف) قابلیت گسترش مجموعه گره ها در اندازه های مختلف از ۱ تا ۵۰، (ب) تابعی از زمان

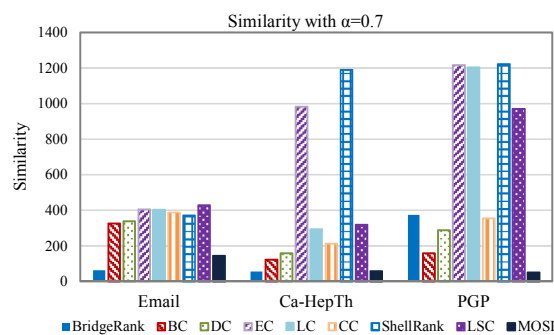


شکل ۱۳: مقایسه قدرت گسترش در الگوریتم های مختلف

در ادامه، شکل ۱۳ قدرت گسترش ۵۰ گره برتر شناسایی شده با الگوریتم های مختلف را نشان می دهد. همان طور که در شکل دیده می شود، روش پیشنهادی در هر سه مجموعه داده طبق مقادیر به دست آمده از مدل انتشار SIR ، قدرت گسترش بیشتر و در نتیجه کارایی بسیار بهتری به نسبت سایر الگوریتم ها دارد. از مقایسه سه نمودار نشان داده شده در شکل های ۱۱، ۱۲ و ۱۳ می توان تأثیر در نظر گرفتن شباهت میان گره های انتخابی را در میزان قدرت گسترش نودها مشاهده کرد.



شکل ۱۱: مقایسه مقادیر تابع هدف اول در الگوریتم های مختلف



شکل ۱۲: مقایسه مقادیر تابع هدف دوم در الگوریتم های مختلف

۵-۴- نتایج تست آماری

روش‌های شناسایی گره‌های پرنفوذ ذکر شده را نشان می‌دهد. همان‌طور که دیده می‌شود روش پیشنهادی در رتبه اول تست فریدمن قرار دارد.

۶-۴- پیچیدگی محاسباتی

پیچیدگی محاسباتی یک تکرار از الگوریتم شامل سه بخش (۱) مرتب‌سازی نامغلوب، (۲) فاصله ازدحامی و (۳) مرتب‌سازی است. پیچیدگی محاسباتی بخش اول $O(M(N)^2)$ است که N اندازه جمعیت و M تعداد توابع هدف است. برای بخش دوم نیاز به محاسبه معادله (۱۰) است که پیچیدگی محاسباتی آن $O(M(N) \log(N))$ است. بنابراین پیچیدگی محاسباتی بخش سوم $O(N \log(N))$ است. بنابراین پیچیدگی کلی الگوریتم $O(MN^2)$ است. فرض کنید تعداد گره‌های گراف n و تعداد یال‌های گراف m باشد. در مقایسه با سایر روش‌ها، در BridgeRank ابتدا گراف شبکه اجتماعی به بخش‌هایی تقسیم شده و از هر بخش بهترین گره شناسایی می‌شود. سپس سایر گره‌ها بر اساس کوتاه‌ترین فاصله آن‌ها از گره‌های شناسایی شده محلی رتبه‌بندی می‌شوند. پیچیدگی زمانی این الگوریتم $O(n \log n)$ است. پیچیدگی محاسباتی CC، EC و BC برابر با $O(n^3)$ است که در BC این مقدار برای گراف‌های خلوت به $O(n^2 \log n)$ کاهش می‌یابد. به علاوه روش ShellRank دارای هزینه محاسباتی $O(m)$ ، DC از مرتبه $O(n)$ و در مورد روش‌های LC و LSC پیچیدگی محاسباتی برابر با $O(n(k)^2)$ است که $\langle k \rangle$ میانگین درجه است.

جدول ۳: نتایج آزمون فریدمن جهت بررسی استقلال روش‌ها

P-value	df	Chi-Square	N
1.9e-270	8	3069.103	600

توانایی گسترش

جدول ۴: رتبه‌بندی روش‌ها با آزمون فریدمن

MOSI	BridgeRank	BC	CC	LC	LSC	ShellRank	EC	DC
۸/۷۹	۶	۷/۸۵	۵/۰۴	۲/۹۱	۳/۹۸	۱/۰۴	۲/۲۲	۷/۱۸

توانایی گسترش

قدرت نفوذ بالاتر نسبت به سایر روش‌ها انتخاب می‌کند. از آنجایی که یکی از چالش‌های الگوریتم پیشنهادی زمان اجرای آن است، در کارهای آینده می‌توان از سایر الگوریتم‌های فرااکتشافی از جمله الگوریتم کلونی مورچه‌ها به دلیل داشتن سرعت و قابلیت انعطاف بیش‌تر آن نسبت به الگوریتم ژنتیک، جهت بهبود پیچیدگی محاسباتی روش پیشنهادی استفاده کرد. به علاوه، الگوریتم مورچه‌ها در تحلیل مسائلی از جمله شبکه‌های اجتماعی که می‌توان آن را به صورت گراف بازنمایی کرد سازگارتر است. از سوی دیگر، می‌توان توابع هدف روش پیشنهادی را در راستای افزایش کارایی بهبود بخشید.

مراجع

- [1] R. Narayanam and Y. Narahari, "A shapley value-based approach to discover influential nodes in social networks," IEEE

در این بخش با استفاده از آزمون فریدمن نتایج به دست آمده از روش‌های مختلف، تجزیه و تحلیل می‌شود. آزمون فریدمن یک آزمون آماری غیر پارامتری است که از آن می‌توان برای ارزیابی نتایج N روش مختلف بر روی k مجموعه داده‌ای، استفاده کرد. این آزمون وضعیت یک متغیر را در چندین حالت وابسته به هم، ارزیابی می‌کند. این آزمون فرضیه‌هایی را تعریف می‌کند که با استفاده از داده‌های مورد آزمون، سعی بر آن دارد یکی از فرضیه‌ها را با احتمال بالا ثابت کند. آزمون فرضیه در این قسمت بدین صورت تعریف می‌گردد: فرضیه صفر برابر است با این که بین روش‌های مختلف تفاوت وجود ندارد. فرضیه یک برابر است با این که بین روش‌های ارائه شده، تفاوت وجود دارد. با اعمال این آزمون، می‌توان هم‌رتبه عملکرد الگوریتم‌ها را مشخص کرد و هم این که یکی از دو فرضیه بالا را به اثبات رساند. برای اثبات فرضیه یک، اگر سطح پوشش آماره آزمون (value-P) از ۵ درصد کمتر باشد، می‌توان وجود تفاوت بین روش‌ها را نتیجه گرفت. در غیر این صورت فرضیه صفر ثابت است. در این مقاله برای انجام آزمون فریدمن از نرم‌افزار SPSS استفاده شده است. در این آزمون، روش‌های شناسایی گره‌های پرنفوذ استفاده شده، بر اساس معیار کارایی آن‌ها رتبه‌بندی می‌شوند و چون معیار کارایی مقدار گسترش گره‌ها بر اساس مدل انتشار SIR است، در نتیجه روش با رتبه بیشتر، کارایی بالاتری دارد. جدول ۳، مقدار P-value برای میزان نفوذ گره‌های انتخابی در ۲۰۰ مرحله، در دیتاست Email، بسیار کمتر از ۰/۰۵ است که می‌توان نتیجه گرفت که فرض استقلال روش‌ها با درصد اطمینان بالایی پذیرفته شده است. به علاوه، جدول ۴، رتبه‌بندی

۵- نتیجه‌گیری

امروزه، شبکه‌های اجتماعی به طور گسترده در کاربردهای بسیاری مانند بازاریابی و پروسی استفاده می‌شوند. با توجه به تعداد زیاد کاربران، نحوه انتخاب هسته اولیه (کاربران هدف) تصمیم کلیدی برای شرکت‌ها به شمار می‌رود. برای پیدا کردن کاربران با نفوذ جهت بازاریابی و پروسی، روش‌های متعددی استفاده شده است. یکی از ضعف‌های روش‌های قبلی، انتخاب کاربران هدفی است که کاربران مشابه را فعال می‌کنند. در این مقاله، یک مدل بهینه‌سازی برای شناسایی با ارزش‌ترین کاربران که شباهت کمی به همدیگر دارند، پیشنهاد شده است. این مدل، هم‌زمان دو هدف بهینه‌سازی سود و کمینه‌سازی شباهت کاربران را در نظر می‌گیرد. ارزیابی بر روی داده‌های واقعی نشان می‌دهد که روش پیشنهادی مجموعه مناسب‌تری از کاربران هدف را برای دستیابی به سود بیشتر و

- model”, Data Mining (ICDM), IEEE 11th International Conference on, pp. 211-220, 2011.
- [20] E. Cohen, D. Delling, T. Pajor and R.F. Werneck, “Sketch-based influence maximization and computation: Scaling up with guarantees”, Proceedings of the 23rd ACM International Conference on Information and Knowledge Management, pp. 629-638, 2014.
- [21] S. Cheng, H. Shen, J. Huang, W. Chen and X. Cheng, “IMRank: influence maximization via finding self-consistent ranking”, Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval, pp. 475-484, 2014.
- [22] C. Wang, L. Deng, G. Zhou and M. Jiang, “A global optimization algorithm for target set selection problems”, Information Sciences, pp. 101-118, 2014.
- [23] M. Heidari, M. Asadpour and H. Faili, “SMG: Fast scalable greedy algorithm for influence maximization in social networks”, Physica A: Statistical Mechanics and its Applications, pp. 124-133, 2015.
- [24] W.-S. Yang, S.-X. Weng, C. Guestrin, C. Faloutsos, J. VanBriesen and N. Glance, “Application of the ant colony optimization algorithm to the influence-maximization problem”, Int J Swarm Intell Evol Comput, pp. 1-8, 2012.
- [25] N. Sinha and B. Annappa, “Cuckoo Search for Influence Maximization in Social Networks”, Proceedings of 3rd International Conference on Advanced Computing, Networking and Informatics, pp. 51-61, 2016.
- [26] Q. Jiang, G. Song, G. Cong, Y. Wang, W. Si and K. Xie, “Simulated Annealing Based Influence Maximization in Social Networks”, AAAI, pp. 127-132, 2011.
- [27] A. Mohammadi and M. Saraei, “Finding influential users for different time bounds in social networks using multi-objective optimization”, Swarm and evolutionary computation, pp. 158-165, 2018.
- [28] 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”, vol. 6, pp. 888-893, 2010.
- [29] P. Bonacich and P. Lloyd, “Eigenvector-like measures of centrality for asymmetric relations”, Social networks, vol. 23, no. 3, pp. 191-201, 2001.
- [30] D. Chen, L. Lü, M.-S. Shang, Y.-C. Zhang and T. Zhou, “Identifying influential nodes in complex networks”, Physica a: Statistical mechanics and its applications, vol. 391, no. 4, pp. 1777-1787, 2012.
- [31] S. Gao, J. Ma, Z. Chen, G. Wang and C. Xing, “Ranking the spreading ability of nodes in complex networks based on local structure”, Physica A: Statistical Mechanics and its Applications, vol. 403, pp. 130-147, 2014.
- [۳۲] محمدامیر عباسیان و حسین نظام‌آبادی‌پور، «الگوریتم جستجوی گرانشی چند هدفه مبتنی بر مرتب‌سازی جبهه‌های مغلوب نشده»، مجله مهندسی برق دانشگاه تبریز، دوره ۴۱، شماره ۱، صفحه ۸۰-۶۸، ۱۳۹۱
- [33] R. M. Anderson, R. M. May and B. Anderson, Infectious diseases of humans: dynamics and control. Wiley Online Library, vol. 28, 1992.
- [34] D. Jong and K. Alan, “Analysis of the behavior of a class of genetic adaptive systems, Engineering,” College of-Technical Reports, University of Michigan, 1975.
- Transactions on Automation Science and Engineering, vol. 8, no. 1, pp. 130-147, 2011.
- [2] W. Chen, Y. Wang and S. Yang, “Efficient influence maximization in social networks,” in Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 199-208, 2009.
- [3] D. Easley and J. Kleinberg, Networks, crowds, and markets: Reasoning about a highly connected world. Cambridge University Press, 2010.
- [4] D. Easley and J. Kleinberg, “Networks, crowds, and markets: Reasoning about a highly connected world,” Significance, vol. 9, pp. 43-44, 2012.
- [5] C. Salavati, A. Abdollahpouri and Z. Manbari, “BridgeRank: A novel fast centrality measure based on local structure of the network”, Physica A: Statistical Mechanics and its Applications, 2017.
- [6] K. Xu, J. Li and Y. Song, “Identifying valuable customers on social networking sites for profit maximization,” Expert Systems with Applications, vol. 39, no. 17, pp. 13009-13018, 2012.
- [7] W. Chen, Y. Yuan and L. Zhang, “Scalable influence maximization in social networks under the linear threshold model,” in Data Mining (ICDM), 2010 IEEE 10th International Conference on, pp. 88-97, 2010.
- [8] L. C. Freeman, “Centrality in social networks conceptual clarification,” Social networks, vol. 1, no. 3, pp. 215-239, 1978.
- [9] G. Sabidussi, “The centrality index of a graph,” Psychometrika, vol. 31, no. 4, pp. 581-603, 1966.
- [10] C. Dangalchev, “Residual closeness in networks,” Physica A: Statistical Mechanics and its Applications, vol. 365, no. 2, pp. 556-564, 2006.
- [11] P. Domingos and M. Richardson, “Mining the network value of customers,” in Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 57-66, 2001.
- [12] M. Richardson and P. Domingos, “Mining knowledge-sharing sites for viral marketing,” in Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 61-70, 2002.
- [13] C. M. Fonseca and P. J. Fleming, “Genetic Algorithms for Multiobjective Optimization: Formulation Discussion and Generalization,” in ICGA, 1993, vol. 93, pp. 416-423. Citeseer.
- [14] N. Srinivas and K. Deb, “Multiobjective optimization using nondominated sorting in genetic algorithms,” Evolutionary computation, vol. 2, no. 3, pp. 221-248, 1994.
- [15] M. Richardson and P. Domingos, “Mining knowledge-sharing sites for viral marketing”, Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 61-70, 2002
- [16] D. Kempe, J. Kleinberg and É. Tardos, “Influential nodes in a diffusion model for social networks”, International Colloquium on Automata, Languages, and Programming, pp. 1127-1138, 2005.
- [17] J. Leskovec, A. Krause, C. Guestrin, C. Faloutsos, J. VanBriesen and N. Glance, “Cost-effective outbreak detection in networks”, Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 420-429, 2007.
- [18] A. Goyal, W. Lu and L.V. Lakshmanan, “Celf++: optimizing the greedy algorithm for influence maximization in social networks”, Proceedings of the 20th international conference companion on World Wide Web, pp. 47-48, 2011.
- [19] A. Goyal, W. Lu and L.V. Lakshmanan, “SimpPath: An efficient algorithm for influence maximization under the linear threshold

زیر نویس ها

^۱ Vertex cover^۱ Seed set^{۱۱} Eigenvector Centrality^{۱۲} Local Centrality^{۱۳} Local Structural Centrality^{۱۴} Pareto front^{۱۵} Binary tournament selection^۱ Viral Marketing^۲ Word-of-Mouth^۳ Target users^۴ NP-Hard^۵ Swarm Intelligence^۶ Genetic Algorithm^۷ Multi-Objective algorithm based on Structural Information^۸ Susceptible-Infected-Recovered