

A Dynamic Metaheuristic Algorithm for Influence Maximization in Social Networks

J. Jabbari Lotf, M. Abdollahi Azgomi*, M. R. Ebrahimi Dishabi

* Professor, Computer Department, Miyaneh Branch, Islamic Azad University, Miyaneh, Iran

(Received: 02/08/2022, Accepted: 29/01/2023)

ABSTRACT

During the very last decade, people have been spending lots of time working with social networks to interact with friends and to share information, thoughts, news, and etc. These social networks comprise a very important part of our daily lives. Along with the exploitation of the development of social networks, finding influential individuals in a social network has many practical functions in marketing, politics, and even control of the diseases. In the present research, a novel method called the dynamic generalized vulture algorithm has been proposed to solve influence maximization problems. Regarding the fact that in real world social networks own very dynamic and scalable nature, through our proposed algorithm, we have considered two important criteria which have been rarely taken into consideration in previous projects. The first criterion is due to the network structure change during time pass and the other refers to scalability. The suggested algorithm was measured considering standard data sets. The results showed that the proposed algorithm has been more scalable and has had higher precision in locating the most influential tops in such networks compared with other algorithms due to the reduction of search area and using several different mechanisms during navigation and optimization, balance creation and moving through these stages.

Keywords: Influence maximization, Social networks, Network dynamicity, Diffusion model, Metaheuristic Algorithms

* Corresponding Author Email: azgomi@iust.ac.ir

یک الگوریتم پویای فرا ابتکاری برای بیشینه‌سازی نفوذ در شبکه‌های اجتماعی

جلیل جباری لطف^۱، محمد عبداللهی ازگمی^{۲*}، محمدرضا ابراهیمی دیشابی^۳

۱- دانشجوی دکترا، ۲- استاد، ۳- استادیار، گروه کامپیوتر، واحد میانه، دانشگاه آزاد اسلامی، میانه، ایران

(دریافت: ۱۴۰۱/۰۵/۱۱، پذیرش: ۱۴۰۱/۱۱/۰۹)

چکیده

در دهه گذشته، مردم زمان زیادی را در شبکه‌های اجتماعی برای تعامل با دوستان و به اشتراک‌گذاری اطلاعات، افکار، اخبار و غیره صرف می‌کنند. این شبکه‌های اجتماعی بخش مهمی از زندگی روزمره ما را تشکیل می‌دهند. با بهره‌برداری از توسعه شبکه‌های اجتماعی، یافتن افراد تأثیرگذار در یک شبکه اجتماعی کاربردهای عملی زیادی در بازاریابی، سیاست و حتی کنترل بیماری‌ها دارد. در این مقاله، روش جدیدی با عنوان الگوریتم کرکس توسعه‌یافته پویا برای حل مسئله بیشینه‌سازی نفوذ ارائه کرده‌ایم. با توجه به این نکته که در دنیای واقعی، شبکه‌های اجتماعی ماهیت بسیار پویا و مقیاس‌پذیر دارند. در الگوریتم پیشنهادی ما دو معیار مهم که در کارهای انجام‌شده قبلی کمتر مورد توجه قرار گرفته است را در نظر می‌گیریم. یکی تغییر ساختار شبکه در طول زمان و دیگری مقیاس‌پذیری است. الگوریتم پیشنهادی روی مجموعه داده‌های استاندارد مورد ارزیابی قرار گرفته شده است. نتایج به‌دست‌آمده نشان می‌دهد که الگوریتم پیشنهادی به دلیل کاهش فضای جستجو و استفاده از چندین مکانیسم مختلف و متفاوت در مراحل اکتشاف و بهره‌وری و ایجاد تعادل و گذار بین این مراحل نسبت به دیگر الگوریتم‌های مورد مقایسه، مقیاس‌پذیرتر بوده و از دقت بالاتری در پیدا کردن رؤس بانفوذ در این شبکه‌ها را برخوردار است.

کلیدواژه‌ها: بیشینه‌سازی نفوذ، شبکه‌های اجتماعی، پویایی شبکه، مدل انتشار، الگوریتم‌های فرا ابتکاری

تمرکز دارند. با این حال، اخیراً محققان بیشتر بر جنبه پویای شبکه‌های اجتماعی تمرکز می‌کنند.

۱- مقدمه

یکی از مسائل چالش‌برانگیز، بیشینه‌سازی نفوذ در شبکه‌های اجتماعی پویا است که در سال‌های اخیر توجه زیادی را به خود جلب کرده است. شبکه‌های اجتماعی مانند اینستاگرام یا توییتر دارای ساختار پویا هستند که در آن رؤس و یال‌هایی در طول زمان به شبکه اضافه یا از آن حذف می‌شوند؛ بنابراین به‌عنوان گراف‌های لحظه‌ای ارائه می‌شوند که در طول زمان تغییر می‌یابند. شبکه‌های اجتماعی در دنیای واقعی پویا هستند و در طول زمان تغییر می‌کنند. کاربران روابط خود را در شبکه‌های اجتماعی ایجاد می‌کنند، درحالی‌که تعاملات کاربر در طول زمان تغییر می‌کند. یک شبکه اجتماعی پویا می‌تواند به‌عنوان مجموعه‌ای از شبکه‌های ایستا مدل‌سازی شود. با توجه به پیشرفت در فناوری‌های جمع‌آوری داده‌ها، مطالعه شبکه‌های پویا به طور فزاینده‌ای رایج شده است. یک موضوع تحقیقاتی مهم این است که چگونه رؤس تأثیرگذار در شبکه‌های اجتماعی پویا شناسایی شوند [۱].

امروزه شبکه‌های اجتماعی نقش بسیار مهمی در تأثیرگذاری و گسترش نفوذ بروی افراد مختلف را دارند. شناسایی گروهی از افراد مهم در یک شبکه اجتماعی یک مسئله مهم است، به‌طوری‌که این افراد در ابتدا هدف قرار می‌گیرند (به‌عنوان مثال، برای اتخاذ یک محصول جدید) و می‌تواند پذیرش محصول جدید را به حداکثر برساند. شبکه‌های اجتماعی به‌عنوان شبکه‌های اجتماعی ایستا و پویا طبقه‌بندی می‌شوند. شناسایی افراد تأثیرگذار در شبکه‌های ایستا نسبتاً ساده است. برای انجام این کار، الگوریتم‌های بسیاری پیشنهاد شده است. اخیراً، تشخیص افراد تأثیرگذار در یک شبکه اجتماعی پویا اهمیت بیشتری پیدا کرده است. زمان نقش مهمی در تغییر توپولوژی شبکه‌های اجتماعی ایفا می‌کند. هرچقدر در تحلیل شبکه‌های اجتماعی نقش زمان پررنگ‌تر باشد شبکه از مدل ایستا به مدل پویا تغییر پیدا می‌کند و این تغییر از مدل ایستا به مدل پویا، پیچیدگی مدل را افزایش می‌دهد زیرا تغییرات ساختار شبکه در طول زمان بیشتر افزایش پیدا می‌کند و برعکس. همچنین با افزایش اندازه شبکه، رأس و یال‌های جدیدی ایجاد می‌گردد. تحقیقات و راه‌حل‌های موجود در زمینه بیشینه‌سازی نفوذ عمدتاً بر توسعه الگوریتم‌های مؤثر و کارآمد بر روی شبکه‌های اجتماعی ایستا

* رایانامه نویسنده مسئول: Azgomi@iust.ac.ir

دیگر کاربران شبکه اجتماعی به حداکثر برسانند. این مسئله به عنوان مسئله بیشینه سازی نفوذ مدل سازی شده و هدف آن انتخاب مجموعه ای متشکل از K رأس است که بیشترین گسترش نفوذ را در شبکه دارند [۳-۶]. با توجه به اینکه محققان الگوریتم های متفاوتی را برای یافتن افراد بانفوذ ارائه کرده اند، دسته بندی های متفاوتی نیز برای این الگوریتم ها ارائه شده است [۷-۱۱]. در این مقاله، این الگوریتم ها بر اساس روش های ارائه شده در سه گروه تقسیم بندی می شوند: (۱) الگوریتم های حریمانه، (۲) الگوریتم های ابتکاری، (۳) الگوریتم های فرا ابتکاری. الگوریتم های حریمانه اغلب برای حل مسائل بهینه سازی سخت مورد استفاده قرار می گیرند. ایده اصلی در این الگوریتم ها استفاده از شبیه سازی مونت کارلو برای محاسبه گسترش نفوذ است. انتخاب بهینه مجموعه بذر بستگی به تعداد تکرارهای شبیه سازی مونت کارلو در هر دور شبیه سازی مونت کارلو دارد و در هر دور یک رأس به مجموعه بذر اولیه خالی اضافه می شود که بیشترین سهم حاشیه ای را مطابق رابطه (۱) دارد.

$$S^* = \operatorname{argmax}_{S \subset V, |S|=k} \sigma(S) \quad (1)$$

S مجموعه ی رؤس فعال شده اولیه، V مجموعه کل رؤس، k تعداد مجموعه بذر، $\sigma(S)$ گسترش نفوذ از مجموعه S را نشان می دهد. این الگوریتم ها بهترین انتخاب ممکن را با توجه به شرایط مسئله انجام می دهند. البته هدف آن ها بهینه سازی مسئله با ادامه این رویکرد است. در بسیاری از موارد، یک فرآیند حریمانه، تقریب خوبی از نتایج قابل قبول را با دقت بالا ارائه می دهد و یک ضریب تقریبی بهینه را تضمین می کند، اما این الگوریتم اغلب زمان اجرای بالایی دارند، به خصوص در مقیاس بزرگ.

الگوریتم های [15].CELFF, [14].CELFF, [13].SMG, [12].TIM, [16].BCT, [17].RIS, [18].NewGreedyIC, [19].StaticGreedy و [20].MixedGreedy نمونه هایی از الگوریتم های حریمانه برای مسئله بیشینه سازی نفوذ هستند. هرچند این روش ها تا حد قابل توجهی زمان اجرای الگوریتم حریمانه را بهبود بخشیده اند اما این روش ها پیچیدگی زمانی بالایی دارد و از این رو قابل تعمیم به شبکه های بزرگ و پویا نمی باشد.

الگوریتم های ابتکاری می توانند راه حل های خوب و نزدیک به بهینه را در زمان اجرای کم برای مسائل بهینه سازی سخت پیدا کنند. رأس های تأثیرگذار دارای ویژگی های توپولوژیکی خاصی در شبکه های اجتماعی هستند، بنابراین ایده اصلی در این الگوریتم ها استفاده از ویژگی های توپولوژی گراف در انتخاب مجموعه بذر است. الگوریتم های [۲۱-۲۵] تنها از یک ویژگی توپولوژی گراف استفاده می کنند، در حالی که الگوریتم های [۲۶-۲۸] و از ترکیبی از چندین ویژگی توپولوژی گراف برای انتخاب رؤس بذر استفاده می کنند. این الگوریتم ها اغلب برای کاهش زمان اجرای

دلیل ناتوان بودن روش های بهینه سازی دقیق در حل مسائل پیچیده با ابعاد بالا و چندحالتی، الگوریتم های تقریبی به عنوان یک رویکرد جدید برای حل این گونه مسائل ارائه شده اند. الگوریتم های تقریبی به دودسته مهم شامل روش های ابتکاری و فرا ابتکاری تقسیم می شود. الگوریتم های ابتکاری به دلیل گیر افتادن در دام محلی و استفاده در حل مسائل بهینه سازی خاص کمتر مورد توجه قرار گرفته اند. به طوری که اخیراً بیشتر از الگوریتم های فرا ابتکاری برای حل اکثر مسائل بهینه سازی پیچیده دنیای واقعی که چندحالتی و غیرخطی هستند، استفاده می شود. این الگوریتم ها می توانند راه حل های بهینه را برای مسائل بهینه سازی پیچیده و سخت مختلف پیدا کنند. البته الگوریتم های فرا ابتکاری راه حل های قابل قبول در زمان منطقی ارائه می دهند ولی هیچ تضمینی برای راه حل های بهینه نمی دهند.

در مقاله [۲] ما با ارائه یک روش مبتنی بر ساختار پویایی شبکه های اجتماعی و محدود کردن فضای مسئله، محدودیت های الگوریتم ژنتیک را در این ساختار پویا و مقیاس پذیر کاهش داده و سرعت همگرایی را نسبت به روش های موجود افزایش داده ایم؛ اما در این مقاله ما چندین مکانیسم مختلف و متفاوت را در مراحل اکتشاف و بهره وری برای ایجاد تعادل و گذار بین این مراحل با توجه به مقیاس پذیری و پویایی شبکه های اجتماعی جهت افزایش دقت شناسایی افراد بانفوذ ارائه کرده ایم. با وجود افزایش دقت شناسایی افراد بانفوذ و ارائه یک مجموعه بذر پویا در یک شبکه نسبتاً بزرگ، هزینه های محاسباتی ارزیابی نفوذ تحت مدل انتشار مورد استفاده بالا بوده و زمان گیر می باشد.

ساختار این مقاله به این صورت تنظیم شده است: در بخش دوم، شرحی از فعالیت های مرتبط انجام شده در زمینه بیشینه سازی نفوذ در شبکه های اجتماعی ارائه شده است در بخش سوم، بیان دقیق مسئله به همراه جزئیات مورد توجه این مقاله آمده است مراحل دقیق پیاده سازی هر یک از روش های شناسایی و انتشار نفوذ به کاررفته در این مقاله نیز در بخش چهارم بیان شده اند در بخش پنجم نیز نتایج حاصل از پیاده سازی و کارهای آتی بیان شده است.

۲- روش تحقیق

اخیراً، شبکه های اجتماعی به طور فزاینده ای در میان کاربران محبوب شده است. با افزایش کاربران شبکه های اجتماعی، تأثیر این شبکه ها بر جامعه افزایش یافته است. مثال های مربوط به این اثرات عبارتند از: جلوگیری از گسترش یک بیماری خاص، فروش محصولات کارخانه و تبلیغات نامزدهای ریاست جمهوری. در این مثال ها، مشکل اساسی این است که چگونه کاربران تأثیرگذار را انتخاب کنیم تا این کاربران بتوانند تأثیر خود را بر

سرعت بالا، مقیاس‌پذیری، مدل‌های انتشار پویا و پیچیدگی محاسبات بالا است. نوآوری ما در این کار تحقیقاتی شامل موارد زیر می‌باشد:

- ما یک سازوکار جدید با توجه به تغییر ساختار شبکه‌های اجتماعی در طول زمان برای شناسایی رأس‌ها بانفوذ ارائه کرده‌ایم.
- ما یک رویکرد جدید جهت کاهش هزینه‌های محاسباتی و حفظ فرایند بهینه‌سازی با مقیاس بزرگ را پیشنهاد کرده‌ایم.
- ما چندین مکانیسم مختلف و متفاوت در مراحل اکتشاف و بهره‌وری با توجه به تغییرات شبکه‌های اجتماعی در طول زمان برای بهینه‌سازی نفوذ ارائه کرده‌ایم.

الگوریتم (۱). الگوریتم فراابتکاری

Input:

Dynamic Network G_t
Timestamps T
The size of seed nodes k
Probability of influence node λ_{ij}

Output:

Seed set S_t at $t = 1, 2, 3, \dots, T$

- 1 Create an initial population
- 2 Update the initial population with a local search
- 3 Update local solutions with global search
- 4 Select best fitness solutions
- 5 Repeat steps 2 to 4 until reaching the stop condition
- 6 Update the final solution
- 7 Calculate the influence spread of the solution in step 6 with the diffusion models

۳- روش پیشنهادی

شبکه‌های اجتماعی ماهیت بسیار پویا و تکاملی دارند و در طول زمان به سرعت در حال تغییر هستند و در نتیجه نتایج مربوط به بهینه‌سازی و برآورد نفوذ افراد در شبکه‌های اجتماعی ممکن است به سرعت از بین بروند و از طرفی با افزایش محبوبیت شبکه‌های اجتماعی تعداد افرادی که از این شبکه‌ها استفاده می‌کنند روبه‌افزایش است و این امر باعث می‌شود که اکثر الگوریتم‌ها به دلیل افزایش پیچیدگی محاسباتی غیر قابل استفاده باشند؛ بنابراین توسعه یک راه‌حل برای شبکه‌های پویا و مقیاس‌پذیر با هزینه کم‌تر و دقت بالاتر ضروری است. ما در این مقاله با در نظر گرفتن مقیاس‌پذیری و پویای شبکه‌های اجتماعی روش جدیدی مبتنی بر الگوریتم کرکس تعمیم‌یافته پویا^۱ (DGVA) برای مسئله بهینه‌سازی نفوذ ارائه می‌دهیم که در یک برجسب زمانی مشخص یک مجموعه بذر پویا با کارایی بالاتر نسبت به روش‌های قبلی ارائه می‌دهد.

الگوریتم‌های حریمانه استفاده می‌شوند. الگوریتم‌های [18] DD، [25] PageRank، [22] VoteRank، [24] LIR، [26] HybridRank، [27] LGIM، [28] IMMRA و [29] GIN نمونه‌هایی از الگوریتم‌های ابتکاری برای مسئله بهینه‌سازی نفوذ را نشان می‌دهد. مهم‌ترین مزیت روش‌های ابتکاری، کاهش تعداد ارزیابی‌های انجام‌شده برای ارزیابی گسترش نفوذ رئوس می‌باشد که نسبت به رویکردهای حریمانه در شبکه‌های بزرگ‌تر، مقیاس‌پذیرتر و سریع‌تر هستند. با این حال، مطالعات قبلی نشان دادند که انتخاب رأس بر اساس معیارهای توپولوژیکی می‌تواند نتایج نادرستی را ایجاد کند زیرا رویکردهای ابتکاری برای همه انواع شبکه‌های اجتماعی مناسب نیستند [۳۰]. به‌منظور حل این مشکل، بعضی از تحقیقات انجام‌شده [۳۱-۳۳] معیارهای جدیدی از جمله مرکزیت مبتنی بر تعاملات اجتماعی، مبتنی بر همسایگان رأس و همسایگان همسایگان، تعاملات زمان‌بندی‌شده، ارتباطات توپولوژی و مبتنی بر انجمن را ارائه داده‌اند که دقیق‌تر هستند. با وجود بهبود زمان اجرا و افزایش مقیاس‌پذیری، روش‌های ابتکاری هیچ تضمینی برای دقت شناسایی رأس‌های تأثیرگذار ندارند.

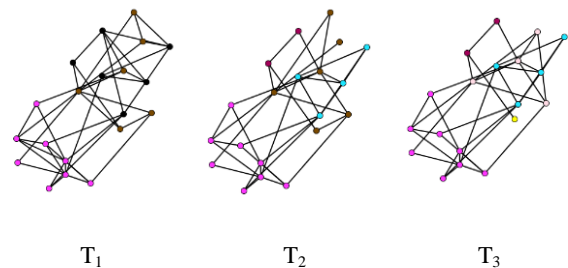
یکی از مشکلات اصلی الگوریتم‌های ابتکاری، همگرایی سریع و قرار گرفتن آن‌ها در نقاط بهینه محلی است، بنابراین محققان الگوریتم‌های فرا ابتکاری را برای حل مسئله الگوریتم‌های ابتکاری ارائه کرده‌اند. الگوریتم‌های فرا ابتکاری نتایج قابل قبول و کارآمدی را با جواب‌های بهینه و کاربردهای مختلف فراهم می‌کنند. الگوریتم‌های [34] SA، [35] GNA، [36] DPSON، [37] DDSE، [38] DBA، [39] AVOA، [40] DSFLA و [41] GTO نمونه‌هایی از الگوریتم‌های فرا ابتکاری برای مسئله بهینه‌سازی نفوذ هستند. در الگوریتم‌های فرا ابتکاری، پیچیدگی محاسباتی با حذف شبیه‌سازی مونت‌کارلو و با استفاده از مقدار نفوذ مورد انتظار به‌عنوان تابع هزینه کاهش می‌یابد. الگوریتم‌های فرا ابتکاری از الگوریتم (۱) برای حل مسئله بهینه‌سازی نفوذ با استفاده از مراحل خاص و با جستجوی فضای مسئله برای یافتن راه‌حل بهینه استفاده می‌کنند.

در حل مسائل بهینه‌سازی توسط الگوریتم‌های فرا ابتکاری، تعیین فضای جستجو و ایجاد تابع هدف برای هر مسئله ضروری است. در ادامه فرایند بهینه‌سازی از عملیات تکاملی به‌صورت تکراری برای یافتن راه‌حل بهینه استفاده می‌شود. به‌طورکلی در این الگوریتم‌ها با تعریف تابع هزینه، مسئله به‌عنوان یک مسئله بهینه‌سازی مدل شده و برای حل آن از روش‌هایی مانند الگوریتم‌های بهینه‌سازی تکاملی استفاده می‌شود. با توجه به مسائل مطرح‌شده، مسئله بهینه‌سازی نفوذ در شبکه‌های اجتماعی پویا، یک حوزه پژوهشی جدیدی است که دارای چالش‌های همچون تغییرات توپولوژی شبکه، انتقال داده با

¹ Dynamic generalized vulture algorithm

۱-۳- بیان مسئله

شبکه‌های اجتماعی پویا، چندین گراف ایستا متناظر با رئوس و یال‌های فعال در یک‌زمان معین در شبکه‌های اجتماعی پویا است. شبکه اجتماعی را می‌توان به مجموعه‌ای از تصاویر لحظه‌ای‌ها تقسیم کرد که در آن، هر گراف لحظه‌ای برای توصیف وضعیت شبکه در زمان t یا برای جمع‌آوری تعاملات مشاهده‌شده در یک دوره زمانی معین مورد استفاده قرار می‌گیرند. ما شبکه اجتماعی پویا را به‌عنوان مجموعه‌ای از گراف‌های لحظه‌ای $\{G_1, G_2, \dots, G_n\}$ مدل‌سازی می‌کنیم که در آن یال‌ها و رئوس در هر گراف لحظه‌ای در بازه‌های زمانی مختلف تغییر می‌کنند. گراف لحظه‌ای را به‌صورت $G_t = (V_t, E_t)$ نشان می‌دهیم که $V_t = \{V_1, V_2, \dots, V_n\}$ نشان‌دهنده رئوس و E_t نشان‌دهنده یال‌هایی است که در طول دوره‌های مورد بررسی ظاهر می‌شوند. نماد G_t ، بیانگر گراف‌های لحظه‌ای در طول زمان است. هدف پیدا کردن افراد بانفوذ در هر کدام از این گراف‌ها است که باعث تغییر اعضای مجموعه بذر در طول زمان مشخص می‌شود البته تعداد عناصر مجموعه بذر برای هر کدام از گراف‌های لحظه‌ای ثابت است. مثالی از شبکه اجتماعی که شامل سه تصویر لحظه‌ای در زمان‌های T_1, T_2, T_3 است در شکل (۱) نشان داده شده است. واضح‌ترین تغییراتی که بین دونقطه زمانی T_i و T_{i-1} رخ می‌دهد حذف و اضافه شدن یال‌ها و رئوس می‌باشد.



شکل (۱). گراف تصویر لحظه‌ای [۱]

همان‌طور که در الگوریتم (۳) نشان داده شده است، ابتدا ما گراف لحظه‌ای G_t را مطابق با ساختار شبکه استخراج کرده و جهت کاهش پیچیدگی محاسبات و بهبود زمان اجرا، رئوس گراف لحظه‌ای G_t را برحسب تمام رئوس قابل‌دسترس از این رئوس را هرس کرده و فهرست‌سازی می‌کنیم. سپس مطابق با الگوریتم پیشنهادی (الگوریتم کرکس تعمیم‌یافته پویا) مجموعه بذر گراف لحظه‌ای G_t را استخراج و جهت استفاده در G_{t+1} ذخیره می‌کنیم. با توجه به بررسی و شبیه‌سازی‌های صورت‌گرفته بر روی مجموعه داده‌های استاندارد و شبکه‌های اجتماعی واقعی [۴۲]، تغییرات یال‌ها در فاصله زمانی G_t, G_{t+1} زیاد نبوده است. بنابراین ما مجموعه بذر استخراج‌شده و رئوس کاندید در گراف لحظه‌ای G_t را به‌عنوان مجموعه رئوس کاندید فهرست گذاری و ذخیره می‌کنیم و مراحل قبلی را برای گراف G_{t+1} تکرار می‌کنیم که

درنهایت ما به یک مجموعه بذر پویا می‌رسیم که بیان‌کننده مجموعه‌ای با بیشترین نفوذ در هر لحظه می‌باشد.

الگوریتم (۲). الگوریتم پیشنهادی

Input:

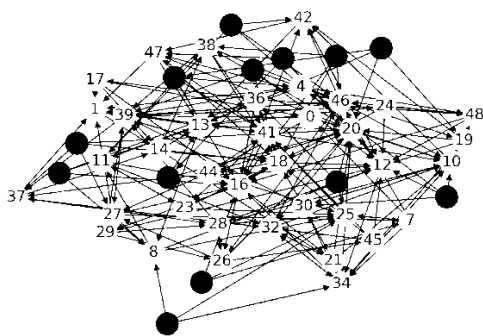
Dynamic Network G_t
Timestamps T
The size of seed nodes k
Probability of influence node λ_{ij}

Output:

Seed set S_t at $t = 1, 2, 3, \dots, T$
 1 for $t = 1$ to T do
 2 $G \leftarrow G_t$
 3 for each $v \in V_t$ do
 4 $C \leftarrow$ Calculate all nodes reachable from v in G .
 5 $P \leftarrow$ Pruning and indexing(C)
 6 Set $S_t \leftarrow$ Proposed DGVA (P)
 7 Extraction and storage
 8 return Seed set S_t at $t = 1, 2, 3, \dots, T$

۲-۳- هرس کردن و فهرست‌سازی

یکی از چالش‌های مهم در تحلیل شبکه‌های اجتماعی مقیاس‌پذیری این نوع از شبکه‌ها است؛ زیرا با افزایش رئوس و یال‌ها اغلب روش‌های ارائه‌شده به دلیل افزایش هزینه پردازش، کارایی و دقت خود را در شناسایی مجموعه بذر از دست می‌دهند.



شکل (۲). رئوس قابل‌دسترس از رأس ۴ (نارنجی)

یکی از راه‌های موجود جهت کاهش هزینه پردازش، هرس کردن و فهرست‌سازی جمعیت اولیه مناسب با گراف می‌باشد. ما در الگوریتم پیشنهادی ابتدا رئوس با درجه خروجی صفر را (برگ‌ها) از مجموعه رئوس هرس کرده و سپس در گراف لحظه‌ای T_0 جمعیت اولیه را بر اساس تعداد رئوس قابل‌دسترس هرس و فهرست‌سازی می‌کنیم. شکل (۲) تعداد رئوس قابل‌دسترس از رأس مشخص را نمایش می‌دهد. سپس در گراف لحظه‌ای بعدی T_1 مجموعه بذر و رئوس کاندید در گراف لحظه‌ای T_0 را به جمعیت اولیه در گراف لحظه‌ای T_1 اضافه کرده و عملیات هرس و فهرست‌سازی را انجام می‌دهیم. این کار را در گراف‌های لحظه‌ای بعدی تا اتمام برچسب زمانی مشخص تکرار

مستقل [۴۳] بررسی کرده‌ایم. مدل انتشار آبخاری مستقل در الگوریتم (۳) شرح داده شده است.

۳-۴- الگوریتم کرکس تعمیم یافته پویا

عمل جستجو و تولید راه‌حل‌های بهینه برای یک مسئله بهینه‌سازی در الگوریتم‌های فرا ابتکاری بر پایه دو مفهوم مهم اکتشاف و بهره‌وری است [۴۴][۴۴][۴۳]. به طوری که هر الگوریتم فرا ابتکاری موفق تلاش می‌کند بین اکتشاف و بهره‌وری تعادل برقرار کند. در این مرحله ما ابتدا بر اساس خروجی مرحله قبل، چندین جواب به صورت تصادفی برای مسئله بیشینه‌سازی نفوذ تولید می‌کنیم. سپس برای تولید راه‌حل‌های جدید از عمل اکتشاف استفاده می‌کنیم. با کاهش فرآیند اکتشاف، قابلیت بهره‌وری رفته‌رفته افزایش پیدا می‌کند. در واقع فرآیند بهره‌وری باعث تولید راه‌حل جدید بر اساس بهترین راه‌حل موجود در جمعیت می‌شود. در نتیجه الگوریتم پیشنهادی از دو فرآیند مهم اکتشاف و بهره‌وری برای گیر نیفتادن در دام محلی و همگرا شدن به سوی هدف (بیشینه‌سازی نفوذ) استفاده می‌کند. الگوریتم پیشنهادی شامل سه مرحله جداگانه بوده و ساختار اصلی الگوریتم کرکس تعمیم یافته در الگوریتم (۴) نشان داده شده است.

مرحله اول: بعد از تشکیل جمعیت اولیه، تابع برازندگی^۳ تمام راه‌حل‌ها محاسبه می‌شود و بهترین راه‌حل به عنوان بهترین جواب بهینه گروه اول و دومین بهترین راه‌حل به عنوان بهترین جواب بهینه گروه دوم انتخاب می‌شود و دیگر راه‌حل‌ها با استفاده از فرمول (۲) به سمت بهترین راه‌حل‌های گروه‌های اول و دوم حرکت می‌کند و در هر تکرار تابع برازندگی کل جمعیت دوباره محاسبه می‌شود.

الگوریتم (۴). DGVA پیشنهادی

Input:

Dynamic network G_t
The size of seed nodes k
Set of Candidate Nodes
Set of Seed Previous

Output:

Seed set S_t

```

1 Initialize the population
2 while (stopping condition is not met) do
3   Calculate the fitness function
4   Find First best category
5   Find Second best category
6   Update the  $pI$ 
7   for (each Initialize population) do
8     Update the  $F$ 
9      $S_B \leftarrow$  Select Random of the Best categories
10    if ( $|F| \geq 1$ ) then
11      if ( $p1 \leq rand$ ) then
12        Change of the vertex with the least
          number of children.
```

می‌کنیم. در حقیقت با این کار ما تعداد رأس‌های گراف لحظه‌ای G_t را حدوداً ۵۰٪ کاهش می‌دهیم.

۳-۳- مدل انتشار آبخاری مستقل^۱

در این مدل کل شبکه اجتماعی به صورت یک گراف نمایش داده می‌شود که در آن رئوس نشان‌دهنده افراد و یال‌ها نشان‌دهنده ارتباط میان آن‌ها هستند. در ابتدا برخی از رأس‌ها به صورت پیش‌فرض فعال (مجموعه بذر) و بقیه غیرفعال هستند. وزن یال‌ها عددی بین صفر و یک است که نشان‌دهنده احتمال فعال‌سازی رأس مقصد توسط رأس مبدأ را نشان می‌دهد.

الگوریتم (۳). مدل آبخاری مستقل

Input:

Dynamic Network G_t
The Set of Seed Nodes A_0
Number of Monte-Carlo simulations MC
Probability of influence node λ_{ij}

Output:

Average Number of Nodes Influenced

```

1 spread = 0
2 repeat:
3    $\tau \leftarrow$  False  $\tau$ : Has The Propagation Ended?
4    $A \leftarrow A_0$   $A$ : Active Nodes
5    $B \leftarrow A$   $B$ : the set of nodes activated
6   while not  $\tau$  do
7     nextB  $\leftarrow$   $\emptyset$ ;
8     for each  $n \in B$  do
9       for each neighbour  $m$  of  $n$ , where  $m \notin A$ ,
do
10        with probability  $\lambda_{ij}$ , add  $m$  to nextB
11         $B \leftarrow$  nextB
12         $A \leftarrow A \cup B$ 
13        if  $B$  is empty then
14           $\tau \leftarrow$  True
15        Spread. Append(Len( $A$ ))
16 until MC
17 return mean(Spread)
```

هر رأس فعالی در زمان فعال‌سازی، رئوس همسایه غیرفعال خود را با احتمال متناظر رابطه‌شان که به طور مستقل مشخص شده است فعال می‌کند و در زمان‌های بعدی در فرآیند فعال‌سازی شرکت نمی‌کند. در مدل آبخاری مستقل، تلاش برای فعال‌سازی یک بار صورت می‌گیرد به این معنی که اگر رأس v نتوانست رأس u را فعال کند دیگر رأس v شانس برای فعال کردن رأس u ندارد. فرآیند تا زمانی اجرا می‌شود که دیگر هیچ فعال‌سازی امکان‌پذیر نباشد. در این مقاله احتمالات نفوذ یکسان در نظر گرفته شده است. پایه‌ی این مدل بر این است که هر فرد در مواجهه با یک خبر بدون توجه به ساختار شبکه و دیگر دوستانش و تنها بر پایه‌ی ارتباط با خبرسان تصمیم بر قبول کردن یا رد آن خبر می‌کند. در این مقاله، ما بیشینه‌سازی نفوذ افراد را بر روی مدل انتشار مبتنی بر زمان‌های گسسته^۲ یعنی مدل آبخاری

^۱ Independent Cascade Diffusion Model

^۲ time-discrete

^۳ Fitness

تکرار این میزان کاهش پیدا می‌کند. زمانی که مقدار $|F|$ بیشتر از یک باشد در این صورت الگوریتم پیشنهادی وارد فاز اکتشاف می‌شود. اما در صورتی که مقدار $|F|$ کمتر از یک باشد الگوریتم پیشنهادی وارد فاز بهره‌برداری می‌شود.

مرحله دوم: اکتشاف

اکتشاف یکی از مهم‌ترین فرایندهای تکاملی برای رسیدن به جواب بهینه محسوب می‌شود. در مرحله اکتشاف اطلاعات جدیدی به فرایند جستجو اضافه می‌شود. چنین ویژگی مهمی به الگوریتم پیشنهادی کمک می‌کند تا از قرارگرفتن در دام بهینه محلی^۱ فرار کند. زمانی که در جواب‌های بهینه متوالی، از عملیات بهره‌وری جهت ایجاد جواب‌های بهینه استفاده می‌شود، جواب‌های کاندید به همگن^۲ شدن گرایش پیدا می‌کند. عملگر اکتشاف به الگوریتم پیشنهادی کمک می‌کند تا تنوع^۳ در جواب‌های کاندید افزایش پیدا کند. در این مرحله الگوریتم پیشنهادی می‌تواند مناطق تصادفی مختلفی را مورد بررسی قرار دهد که این بررسی‌ها می‌توانند بر اساس دو راهبرد مختلف باشد و برای انتخاب هر کدام از این دو راهبرد از پارامتری به نام p_1 استفاده شده است که این پارامتر باید قبل از انجام عملیات جستجو مقداردهی شود و مقداری بین ۰ و ۱ داشته باشد که این پارامتر تعیین‌کننده میزان استفاده هر یک از این دو راهبرد می‌باشد. برای انتخاب هر کدام از راهبردها یک عدد تصادفی بین ۰ و ۱ تولید می‌کنیم اگر این عدد بزرگ‌تر یا مساوی از پارامتر p_1 باشد ما بر اساس معادله (۱) یکی از دو راه‌حل بهینه را انتخاب کرده و سپس رأسی که کمترین تعداد فرزند را در این راحل بهینه را دارد با یک رأس جدید از تفاضل مجموعه کل رئوس با درجه ورودی صفر با مجموعه رئوس راه‌حل بهینه انتخابی به صورت تصادفی جایگزین می‌کنیم.

اما در صورتی که عددی تصادفی که تولید شده بود کوچک‌تر از پارامتر p_1 باشد از مرکزیت درجه ورودی استفاده می‌شود. مرکزیت درجه میزان ارتباط یک رأس با سایر رئوس را بررسی می‌کند. به عبارت دیگر رأس خوردگی هر رأس را به شبکه نشان می‌دهد. رئوس که مرکزیت درجه بالایی دارند با احتمال بیشتری دریافت‌کننده اطلاعات در شبکه خواهند بود. در شبکه‌های جهت‌دار هر رأس یک درجه یال‌های ورودی و یک درجه یال‌های خروجی دارد. درجه یال‌های ورودی به‌نوعی نشانگر شهرت و درجه یال‌های خروجی نشانگر اجتماعی بودن آن رأس است. رئوس با مرکزیت درجه پایین معمولاً در حاشیه شبکه یا شاخه‌های با جریان کم قرار دارند. برای یک رأس i ، مرکزیت درجه آن به صورت معادله (۵) تعریف می‌شود که در آن A_{ij} ماتریس مجاورتی گراف و n تعداد رئوس گراف است. ما در این

```

13 else
14 Change of the vertex with the highest
   input degree.
15 else ( $|F| < 1$ ) then
16   if ( $|F| < 0.5$ ) then
17     if ( $p_2 \geq rand$ ) then
18       Merge  $S_B$  and population $_i$  then
19       Update based on the degree
20     else
21       Merge first and second the best
   categories then Update based
22       on the Betweenness
23     else if ( $|F| \geq 0.5$ ) then
24       if ( $p_3 \geq rand$ ) then
25         Merge  $S_B$  and population $_i$  then
   Update based on the degree
26     else
27       Merge first and second the best
   categories then Update based on
   the Closeness
26 Return Seed set  $S_i$ 

```

$$R_S = \begin{cases} \text{Best Category (1)} & \text{if } p_1 = \alpha \\ \text{Best Category (2)} & \text{if } p_2 = \beta \end{cases} \quad (2)$$

در فرمول (۲)، احتمال انتخاب جواب بهینه منتخب برای حرکت دیگر راه‌حل‌ها به سمت یکی از بهترین راه‌حل‌های هر گروه محاسبه می‌کند که α و β پارامترهایی هستند که قبل از انجام عملیات جستجو مقداردهی شوند و مقدار این پارامترها بین ۰ و ۱ باشد و مجموع مقادیر هر دو پارامتر برابر با یک می‌باشد و احتمال انتخاب هر یک از بهترین راه‌حل‌ها با استفاده از چرخه رولت ویل می‌باشد که احتمال انتخاب هر یک از بهترین راه‌حل هر گروه با استفاده از معادله (۳) به دست می‌آید.

$$p_i = \frac{F_i}{\sum_{i=1}^n F_i} \quad (3)$$

در الگوریتم پیشنهادی ما از یک سازوکاری برای انتقال از فاز اکتشاف به فاز بهره‌برداری استفاده می‌کنیم که این سازوکار بر اساس میزان اکتشاف جواب‌های بهینه است که به طور مکرر در حال جستجوی می‌باشند و با افزایش تعداد این تکرارها که باعث می‌شود جواب‌های بهینه به هم نزدیک بوده و روندی نزولی در تولید جواب‌های بهینه جدید را داشته باشیم. برای مدل‌سازی این رفتار از معادله (۴) استفاده شده است.

$$F = 2F_0 \left(1 - \frac{\text{iteration}_i}{\text{max iterations}}\right) \quad (4)$$

در حقیقت معادله (۴) بیانگر میزان همگرا بودن جواب‌های بهینه می‌باشد؛ و F_0 عددی تصادفی بین ۱- تا ۱ می‌باشد که در هر بار تکرار تغییر می‌کند. زمانی که مقدار F_0 کمتر از ۰ می‌شود بدان معنی است که هنوز جواب‌های بهینه جدیدی تولید می‌شود و در صورت افزایش این مقدار از ۰ بدان معنی است که میزان تولید جواب‌های بهینه جدید روندی کاهشی دارد و در هر بار

¹ Local Optimum

² Homogeneous

³ Diversity

در تکرار جاری صورت می‌گیرد. سپس بر اساس درجه ورودی مرتب‌شده و هرس می‌شود. مرحله دوم داخلی: زمانی الگوریتم پیشنهادی وارد فاز دوم در بهره‌برداری می‌شود که مقدار که مقدار $|F|$ کمتر از ۱ و بزرگ‌تر و مساوی از ۰/۵ باشد. در این مرحله دو راهبرد متفاوت در نظر گرفته می‌شود. برای تعیین میزان انتخاب هریک از راهبردها از پارامتر $p3$ استفاده شده است که این پارامتر باید قبل از اجرای عملیات جستجو مقداردهی شود و مقدار این پارامتر باید بین ۰ و ۱ باشد. در ابتدای این فاز عددی تصادفی بین ۰ و ۱ تولید می‌شود در صورتی که این عدد بزرگ‌تر و یا مساوی پارامتر $p3$ باشد راهبرد انتخاب تصادفی یکی از راه‌حل‌های بهینه و ادغام با راه‌حل انتخاب‌شده جاری صورت می‌گیرد. سپس بر اساس مرکزیت درجه خروجی مرتب‌شده و هرس می‌شود.

اما در صورتی که این عدد تصادفی کوچک‌تر از پارامتر $p3$ باشد راهبرد ادغام راه‌حل‌های بهینه دسته اول و دوم در تکرار جاری صورت می‌گیرد. سپس بر اساس مرکزیت بینابینی مرتب شده و هرس می‌شود. مرکزیت بینابینی یک رأس خاص در شبکه عبارت است از تعداد کوتاه‌ترین مسیرهای میان رأس‌های شبکه که از یک رأس خاص عبور می‌کنند. در حقیقت، این معیار محاسبه می‌کند چه تعداد از رأس‌های شبکه با واسطه کمتر برای ارتباط سریع‌تر باهم به این رأس نیاز دارند. هرچه بینابینی رأس بیشتر باشد یعنی اینکه رأس در موقعیت راهبردی قرار گرفته است و قدرت بالایی در انتقال اطلاعات دارد و این معیار به صورت رابطه (۷) حساب می‌شود:

$$C_b(n_i) = \sum_{w \neq w' \in V} \frac{\delta_{ww'}(n_i)}{\delta_{ww'}} \quad (7)$$

که در آن $\delta_{ww'}$ تعداد کوتاه‌ترین مسیریابی است که بین دو رأس w و w' وجود دارد و $\delta_{ww'}(n_i)$ تعدادی از این کوتاه‌ترین مسیریاب است که از رأس n_i نیز رد می‌شود. از آنجاکه محاسبه دقیق مرکزیت بینابینی در شبکه‌های بسیار بزرگ هزینه‌بر است ما از [۴۶] جهت محاسبه مرکزیت بینابینی تقریبی استفاده کرده‌ایم. لیست کامل پارامترهای الگوریتم کرکس تعمیم‌یافته پویا در جدول (۱) گزارش شده است. ما این پارامترها را با آزمایش مقادیر مختلف به دست آورده‌ایم.

جدول (۱). پارامترهای الگوریتم کرکس تعمیم‌یافته پویا

Parameter	Value
Population size	۱۰۰
$P1$	۰/۶
$P2$	۰/۴
$P3$	۰/۶
α	۰/۸
β	۰/۲

مرحله راه‌حل بهینه جاری از جمعیت اولیه را بر اساس درجه ورودی مرتب کرده و رأسی که بیشترین درجه ورودی را دارد با یک رأس جدید از مجموعه رؤوس کاندید که شامل رؤوسی با بیشترین درجه بینابینی، نزدیکی و رؤوس ریشه می‌باشد به صورت تصادفی جایگزین می‌کنیم.

$$C_D(v_i) = \frac{1}{n+1} \sum_{j=1}^n A_{ij} \quad (5)$$

مرحله سوم: بهره‌وری

در صورتی که مقدار $|F|$ کمتر از یک باشد الگوریتم پیشنهادی وارد مرحله بهره‌وری می‌شود که همین مرحله نیز دارای دو مرحله داخلی دیگری است و در هر مرحله بهره‌وری از دو راهبرد متفاوت استفاده شده است که میزان انتخاب هر یک از راهبردها در هر مرحله داخلی با استفاده از دو پارامتر $p2$, $p3$ تعیین می‌شود. از پارامتر $p2$ برای فاز داخلی اول و از پارامتر $p3$ برای فاز داخلی دوم استفاده شده است هر دو پارامتر می‌بایستی قبل از انجام عملیات جستجو مقداردهی شده و مقداری بین صفر و یک قرار بگیرد. **مرحله اول داخلی:** در صورتی که مقدار $|F|$ کمتر از ۱ و کمتر از ۰/۵ باشد آنگاه این مرحله از الگوریتم اجرا می‌شود. در این مرحله یک عدد تصادفی بین صفر و یک تولید می‌شود در صورتی که این عدد بزرگ‌تر و یا مساوی پارامتر $p2$ باشد راهبرد انتخاب تصادفی یکی از راه‌حل‌های بهینه و ادغام با راه‌حل بهینه جاری صورت می‌گیرد. سپس بر اساس مرکزیت نزدیکی مرتب شده و هرس می‌شود. در یک گراف جهت‌دار اگر بخواهیم به سرعت اطلاعات را در شبکه پخش کنیم باید دنبال رأس‌هایی باشیم که به طور متوسط به دیگران نزدیک هستند و می‌تواند به سرعت اطلاعات را به همه رأس‌ها (به طور مستقیم یا غیرمستقیم) برسانند. نزدیکی عبارت است از عکس متوسط فاصله یک رأس تا رأس‌های دیگر گراف است. رأس ای که دارای بیشترین مقدار نزدیکی است، سرعت دسترسی بیشتری به رأس‌های دیگر دارد و می‌تواند در مدت‌زمان کمی به همه رأس‌ها اطلاعات ارسال نماید. این معیار به صورت رابطه (۶) محاسبه می‌شود:

$$C_c(n_i) = \left(\sum_{j=1}^n d_{ij} \right)^{-1} \quad (6)$$

که در آن d_{ij} کوتاه‌ترین مسیر میان دو رأس است. محاسبه معیارهای مرکزیت در گراف‌های متناظر با شبکه‌های بزرگ دارای پیچیدگی محاسباتی و زمانی بالاست. در این مقاله ما برای کاهش زمان محاسبه مرکزیت نزدیکی از الگوریتم [۴۵] که مقیاس‌پذیری و دقت بالاتری را از خود نشان می‌دهد استفاده می‌کنیم. در غیر این صورت اگر مقدار تولیدشده کوچک‌تر از پارامتر $p2$ باشد راهبرد ادغام راه‌حل‌های بهینه دسته اول و دوم

دیگر دریافت می‌کنند و کاربران ممکن است در هر دو حالت سؤال و پاسخ اظهارنظر کنند که در نهایت ما یک شبکه زمانی را در محدوده زمان مشخص به دست می‌آوریم.

جدول (۲). مشخصات شبکه‌های اجتماعی مورد ارزیابی

دیتاست	Wiki Talk	Stack Overflow
تعداد رأس‌ها	۱۱۴۰۱۴۹	۱۶۴۶۳۳۸
تعداد یال‌های پویا	۷۸۳۳۱۴۰	۲۵۴۰۵۳۷۴
تعداد یال‌های ایستا	۳۳۰۹۵۹۲	۱۱۳۷۰۳۴۲
محدوده زمان	۲۳۲۰ روز	۲۷۷۳ روز

دیتاست Wiki Talk: این دیتاست یک شبکه زمانی است که کاربران ویکی‌پدیا را در حال ویرایش صفحه بحث یکدیگر نشان می‌دهد. یال u به v در زمان t به این معنی است که کاربر u صفحه بحث کاربر v را در زمان t ویرایش کرده است.

چندین شیوه ابتکاری برای یافتن راه‌حل‌های مناسب برای مسئله بیشینه‌سازی نفوذ ارائه شده است. الگوریتم (ERIS) ارائه شده توسط هاوتاهی^۲ و همکاران از مفاهیم نمونه‌گیری تأثیر معکوس برای محاسبه پاسخ‌های دقیق نفوذ به منظور ارزیابی دقت تجربی راه‌حل‌های حریصانه در مسئله بیشینه‌سازی نفوذ استفاده می‌کند. درجه بالا (HIGHDEG) یک روش ابتکاری حریصانه است که به سادگی رؤسی را به مجموعه بذر اضافه می‌کند که به ترتیب بیشترین درجه خروجی را دارند. تخفیف درجه^۳ (SDISC) یک پالایشی^۴ از روش درجه بالا است که توسط چن و همکارانش پیشنهاد شده است. آن‌ها تعداد همسایگان رأس مورد ارزیابی را که درون مجموعه بذر قرار دارد را به عنوان مشخصه جدید در نظر گرفتند و این تعداد را از درجه آن رأس، کم کردند و از مقدار به دست آمده به جای درجه تنها استفاده کردند. روش نمونه‌گیری تأثیر معکوس^۵ (RIS) که توسط براگ^۶ و همکاران پیشنهاد شده است. یک روش مبتنی بر طرح^۷ می‌باشد که به جای شبیه‌سازی مستقیم فرایند نفوذ، شبیه‌سازی‌های معکوس را برای ساخت طرح‌هایی متشکل از یک مجموعه از رأس‌ها برای تخمین مؤثر گسترش نفوذ انجام می‌دهد. الگوریتم (DGA) با در نظر گرفتن پویایی شبکه‌های اجتماعی و با استفاده از روش‌های محاسبات نرم یک مجموعه بذر پویا برای بهبود گسترش نفوذ به دست می‌آورد.

شکل (۳) نتایج ارزیابی‌های صورت گرفته بر روی دیتاست Stack Overflow در سه لحظه‌ای متفاوت از یک برچسب زمانی مشخص و با مقادیر مختلفی از مجموعه بذر را نشان می‌دهد.

۳-۵- استخراج و ذخیره

زمان اجرا یکی از پارامترهای مهم در مسئله بیشینه‌سازی نفوذ است. با افزایش اندازه شبکه پیچیدگی محاسباتی مسئله بیشینه‌سازی نفوذ افزایش می‌یابد. در نتیجه زمان اجرا به صورت قابل ملاحظه‌ای افزایش پیدا می‌کند. شبکه‌های اجتماعی همچون توئیتر که یک شبکه اجتماعی پویا است حدود ۱۰ درصد از کل ارتباطات در یک ماه تغییر می‌کند [۴۲]. به عنوان مثال، یک کاربر معمولی با ۱۰۰ دنبال کننده، فقط حدود ۱۰ درصد دنبال کنندگان آن بیشتر می‌شود در حالی که در حدود ۳ درصد از دنبال کننده‌گان خود را در یک ماه از دست می‌دهد. با توجه به مطالب ذکر شده ما در این مرحله برای کاهش پیچیدگی محاسباتی نیاز به کاهش قابل قبول اندازه شبکه هستیم و همچنین با توجه به تغییرات حدود ۱۰ درصد می‌توانیم از مجموعه بذر مراحل قبل نیز استفاده کنیم در حقیقت زمانی که یک رأس به عنوان یک عضو مجموعه بذر در گراف لحظه‌ای G_t انتخاب می‌شود آن را فهرست گذاری و ذخیره می‌کنیم و در گراف‌های لحظه‌ای بعدی در مرحله‌ای ایجاد جمعیت اولیه و هرس کردن و فهرست سازی مورد استفاده قرار می‌دهیم. با توجه به تغییر ساختار شبکه‌های اجتماعی در طول زمان، مجموعه بذر نیز در طول زمان تغییر می‌یابد. که در نهایت ما با در نظر گرفتن مقیاس پذیری و پویایی شبکه‌های اجتماعی یک مجموعه بذر پویا با کارایی بالاتر نسبت به روش‌های قبلی برای یک برچسب زمانی مشخص داریم.

۴- نتایج و بحث

در این قسمت نتایج تجربی مقایسه الگوریتم پیشنهادی با روش‌های ERIS[47], HIGHDEG[48], SDISC[18], RIS[17], DGA[2] ارائه شده است. ما جهت ارزیابی الگوریتم پیشنهادی با سایر الگوریتم‌های ارائه شده از مجموعه داده‌ای (دیتاست) موجود در SNAP^۱ طبق جدول (۲) استفاده می‌کنیم. آزمایش‌های شبیه‌سازی شده توسط پایتون بر روی سرویس ابری Colab و با استفاده از پردازنده GPU صورت گرفته است و پارامترها در آزمایش‌ها به شرح زیر تنظیم شده‌اند. تعداد تکرار شبیه‌سازی‌ها برای گسترش نفوذ برابر ۵۰ و تعداد تکرار شبیه‌سازی مونت کارلو برابر ۱۰۰ و احتمال نفوذ ۰/۵ در نظر گرفته شده است. این سه پارامتر در همه آزمایش‌ها یکسان هستند همچنین اندازه مجموعه بذر از ۱۰ تا ۵۰ تنظیم شده است و الگوریتم پیشنهادی با سایر الگوریتم‌های ارائه شده با تعداد متغیری از رأس‌ها و یال‌ها در سه لحظه متفاوت t_a , t_b و t_c در بازه زمانی یکسان به طوری که $t_a < t_b < t_c$ می‌باشد مورد ارزیابی قرار گرفته‌اند.

دیتاست Stack Overflow: در وبسایت Stack Overflow، کاربران سؤالات خود را ارسال می‌کنند و پاسخ‌ها را از کاربران

^۲ Hautahi

^۳ Single discount

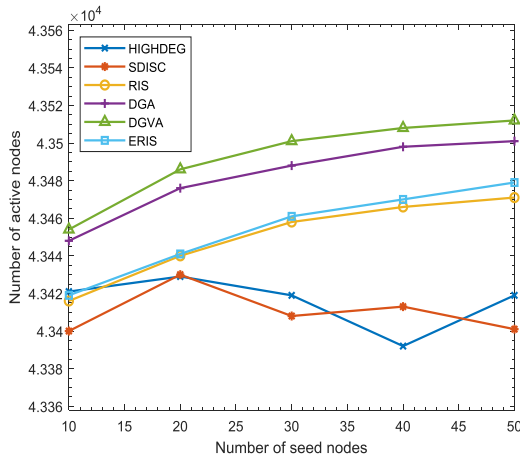
^۴ refinement

^۵ Reverse Influence Sampling

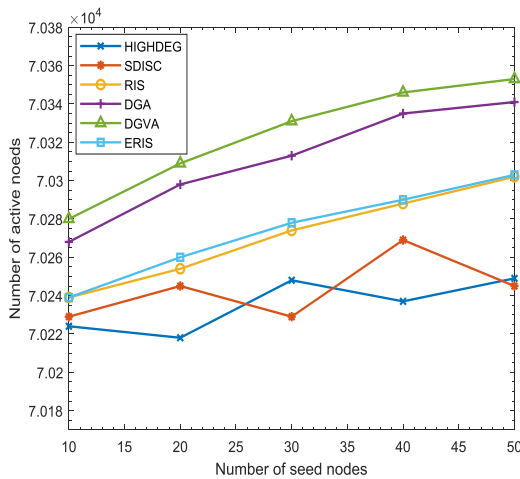
^۶ Brogs

^۷ Sketch-based methods

^۱ <https://snap.stanford.edu/index.html>

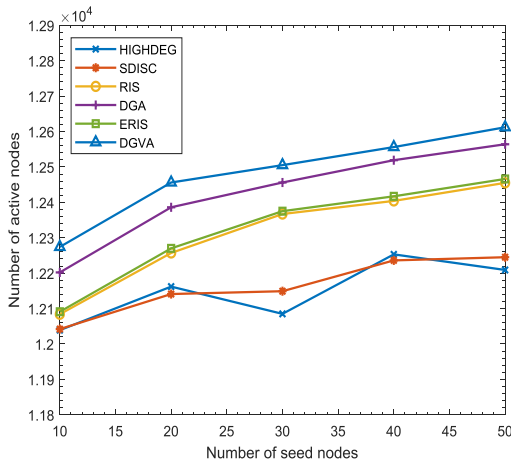


ب) نتایج گسترش نفوذ روی دیتاست در لحظه t_B



ج) نتایج گسترش نفوذ روی دیتاست در لحظه t_C

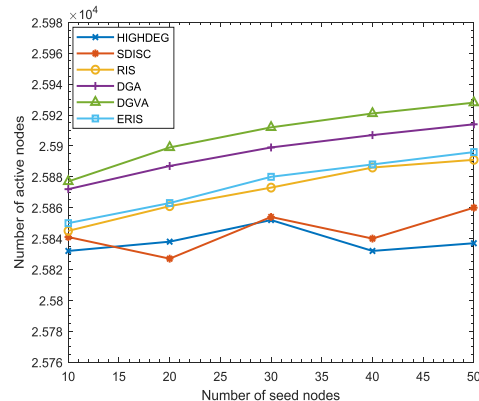
شکل (۳). مقایسه نتایج عددی بر روی دیتاست Stack Overflow در یک برچسب زمانی



الف) نتایج گسترش نفوذ روی دیتاست در لحظه t_A

الگوریتم پیشنهادی دقت بیشتری در یافتن جواب نزدیک به بهینه را از خود نشان می‌دهد. دلیل این امر استفاده از چندین مرحله اکتشاف و بهره‌وری برای به دست آوردن بهترین تعادل بین اکتشاف راه‌حل‌های جدید و بهره‌وری از راه‌حل‌های موجود است. استفاده از جمعیت متغیر این اجازه را به الگوریتم پیشنهادی می‌دهد که الگوریتم از مینیمم‌های محلی عبور کرده و فضای بیشتری را جستجو کند. همان‌طور که از شکل‌ها برمی‌آید الگوریتم پیشنهادی ما در فواصل زمانی مختلف بیشترین نفوذ (تعداد نواحی فعال) را در شبکه نسبت به روش‌های دیگر از جمله الگوریتم DGA که یک روش مبتنی بر محاسبات نرم جهت ارائه یک مجموعه بذر پویا در یک برچسب زمانی مشخص است و الگوریتم مبتنی بر روش نمونه‌گیری تأثیر معکوس RIS که توسط براگ، برای اجتناب از محدودیت‌های موجود در الگوریتم‌های حریصانه سنتی برای به حداکثر رساندن نفوذ تحت مدل آبخاری مستقل پیشنهاد شده بود و الگوریتم‌های ابتکاری حریصانه High-Degree و Single-Degree نشان می‌دهد.

همچنین با افزایش اندازه مجموعه رأس‌ها و یال‌ها الگوریتم پیشنهادی ما نتایج مناسب‌تری را با توجه به سایر نتایج نشان می‌دهد. با مقایسه نتایج با جزئیات بیشتر، می‌بینیم که وقتی اندازه مجموعه بذر و تعداد رأس‌های شبکه کم باشد، تفاوت بین نتایج ما و دیگر الگوریتم‌ها واضح نیست؛ اما با افزایش تعداد رأس‌ها، یال‌ها و اندازه مجموعه بذر، نتایج الگوریتم پیشنهادی ما بسیار بهتر از الگوریتم‌های ابتکاری حریصانه است البته الگوریتم‌های ابتکاری حریصانه مورد مقایسه دارای کمترین مرتبه زمانی هستند ولی در شبکه‌های بزرگ این الگوریتم‌ها با افزایش اندازه مجموعه بذر دارای دقت نسبتاً پایینی در شناسایی افراد بانفوذ به دلیل عدم انتخاب آگاهانه دارند و الگوریتم پیشنهادی در مقایسه با الگوریتم DGA عملکرد بهتری از خود نشان می‌دهد. نتایج شکل‌های الف، ب و ج شکل (۳) نشان می‌دهد که الگوریتم پیشنهادی مقیاس‌پذیرتر و دارای دقت بالاتری نسبت به سایر الگوریتم‌های شبیه‌سازی شده است.



الف) نتایج گسترش نفوذ روی دیتاست در لحظه t_A

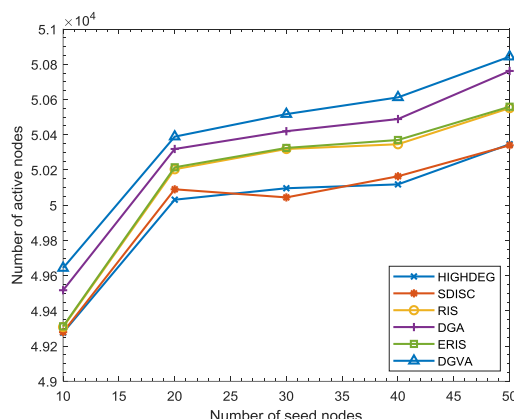
SDISC, RIS, DGA را با تعداد مجموعه بذر ۱۰ و ۳۰ و ۵۰ در دو برجسب زمانی متفاوت در دیتاست‌های Stack Overflow, Wiki Talk مقایسه می‌کنیم جدول (۳) نتایج را نشان می‌دهد. همان‌طور که در جدول (۳) نشان داده شده است، زمان اجرای الگوریتم پیشنهادی کندتر از الگوریتم‌های SDISC, RIS است و در مقایسه با الگوریتم DGA با افزایش تعداد مجموعه بذر بهتر عمل می‌کند، اما دقت شناسایی افراد بانفوذ در الگوریتم پیشنهادی نسبت به بقیه بیشتر می‌باشد. اگرچه عملکرد الگوریتم پیشنهادی تحت تأثیر پارامترهای آن قرار دارد اما نتایج آن بسیار بهتر از سایر الگوریتم‌های شبیه‌سازی شده است. در برخی موارد الگوریتم پیشنهادی تحت تأثیر ساختار شبکه قرار می‌گیرد، باوجود این تنوع راه‌حل را تضمین می‌کند اما فرایند بهینه‌سازی پیچیده‌تر شده و زمان اجرای الگوریتم افزایش پیدا می‌کند. در حالت کلی الگوریتم پیشنهادی در دقت شناسایی مجموعه بذر بهتر عمل می‌کند.

جدول (۳). زمان اجرای الگوریتم‌های مختلف

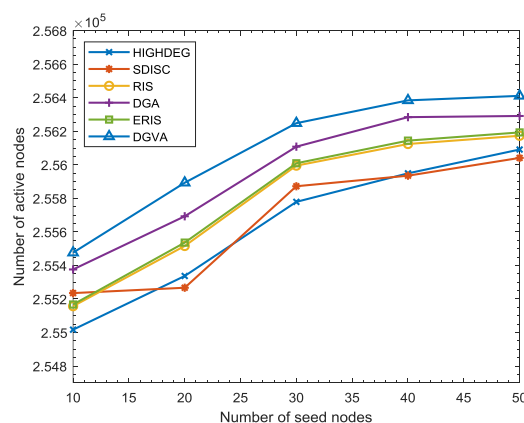
	SDISC			RIS			DGA			DGVA		
	۱۰	۳۰	۵۰	۱۰	۳۰	۵۰	۱۰	۳۰	۵۰	۱۰	۳۰	۵۰
Stack Overflow	۱۱۲/۴۳	۲۱۴/۱۷	۳۷۲/۹۱	۲۰۸/۹۴	۲۱۶/۷۶	۲۰۴/۷۰	۱۹۸/۴۸	۳۷۴/۵۹۶	۷۹۹/۹۴	۲۱۴/۲۲	۳۴۴/۹۲	۴۹۹/۲۴
	۱۰۶/۰۵	۱۸۲/۱۴	۳۰۰/۸۵	۱۸۶/۰۵	۲۰۳/۳۴	۲۰۱/۶۶	۱۲۱/۴۸	۲۸۳/۱۴۲	۵۰۲/۱۴۲	۱۵۳/۱۶۵	۲۷۴/۱۳۶	۴۹۴/۱۲
Wiki Talk	۱۱۱/۰۷	۱۶۶/۰۹	۲۵۷/۰۴	۱۳۲/۳۸	۱۳۰/۱۵	۱۳۶/۷۳	۱۵۲/۵۴	۳۲۲/۵۴	۶۸۲/۵۴	۱۹۸/۱۸۳	۳۲۲/۸۶	۶۳۹/۳۳
	۱۶۲/۱۴	۲۴۷/۸۴	۳۹۸/۲۲	۱۸۹/۷۲	۱۸۷/۵۱	۱۸۹/۶۴	۲۸۴/۴۲	۴۹۳/۱۰۶	۹۶۲/۱۴۲	۳۲۴/۲۱	۴۷۶/۸۷	۷۸۴/۱۶

۵- نتیجه‌گیری

شبکه‌های اجتماعی مانند فیس‌بوک، توئیتر، اینستاگرام، واتس‌آپ و غیره با توسعه اینترنت تا حد زیادی در مقیاس بی‌سابقه‌ای توسعه یافته‌اند. بر اساس آمار^۱ ژانویه سال ۲۰۲۱ از ۴/۶۶ میلیارد کاربر فعال در اینترنت ۴/۲ میلیارد در شبکه‌های اجتماعی فعالیت داشته‌اند. با بهره‌برداری از توسعه شبکه‌های اجتماعی



(ب) نتایج گسترش نفوذ روی دیتاست در لحظه t_B



(ج) نتایج گسترش نفوذ روی دیتاست در لحظه t_C

شکل (۴). مقایسه نتایج عددی بر روی دیتاست wiki-talk در یک برجسب زمانی

در شکل (۴) گسترش نفوذ بر روی دیتاست wiki-talk در زمان‌های متفاوت با تعداد رأس‌ها و یال‌های متفاوت مورد بررسی قرار گرفته است. زمانی که تعداد رأس‌ها و یال‌های شبکه افزایش پیدا می‌کند الگوریتم پیشنهادی به علت دارا بودن فضای جستجوی بهینه، رقابت و تکامل جمعیت اولیه‌ای که بر اساس فهرست‌سازی و هرس کردن انتخاب شده‌اند و همچنین به علت استفاده از سازوکارهای متفاوت در مراحل اکتشاف و بهره‌وری جهت اجتناب از قرارگرفتن در دام بهینه محلی، نتایج بهتری نسبت به سایر الگوریتم‌های مورد بررسی قرار گرفته از خود نشان می‌دهد.

لازم به ذکر است که با افزایش اندازه مجموعه بذر در الگوریتم پیشنهادی به دلیل تنوع راه‌حل‌ها و رقابت بین رأس‌ها گسترش نفوذ افزایش پیدا می‌کند اما بعد از مدتی به دلیل ساختار شبکه نمی‌توان این مقدار را به طور قابل‌ملاحظه افزایش داد.

از لحاظ کارایی، ما زمان اجرای چهار الگوریتم DGVA,

¹ <http://www.statista.com>

- [5] H. A. Beni, Z. Aghaee, A. Bouyer, and M. Vahidipour, "IMT: selection of top-k nodes based on the topology structure in social networks," in *2020 6th International Conference on Web Research (ICWR)*, 2020, pp. 84-88: IEEE.
- [6] Z. Aghaee, H. A. Beni, S. Kianian, and M. Vahidipour, "A heuristic algorithm focusing on the rich-club phenomenon for the influence maximization problem in social networks," in *2020 6th International Conference on Web Research (ICWR)*, 2020, pp. 119-125: IEEE.
- [7] S. Peng, Y. Zhou, L. Cao, S. Yu, J. Niu, and W. Jia, "Influence analysis in social networks: A survey," *Journal of Network and Computer Applications*, vol. 106, pp. 17-32, 2018.
- [8] B. Chang, T. Xu, Q. Liu, and E.-H. Chen, "Study on information diffusion analysis in social networks and its applications," *International Journal of Automation and Computing*, vol. 15, no. 4, pp. 377-401, 2018.
- [9] S. Banerjee, M. Jenamani, and D. K. Pratihari, "A survey on influence maximization in a social network," *Knowledge and Information Systems*, vol. 62, no. 9, pp. 3417-3455, 2020.
- [10] S. S. Singh, A. Kumar, K. Singh, and B. Biswas, "IM-SSO: Maximizing influence in social networks using social spider optimization," *Concurrency and Computation: Practice and Experience*, vol. 32, no. 2, p. e5421, 2020.
- [11] J. Shang, S. Zhou, X. Li, L. Liu, and H. Wu, "CoFIM: A community-based framework for influence maximization on large-scale networks," *Knowledge-Based Systems*, vol. 117, pp. 88-100, 2017.
- [12] 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*, vol. 420, pp. 124-133, 2015.
- [13] J. Leskovec, A. Krause, C. Guestrin, C. Faloutsos, J. VanBriesen, and N. Glance, "Cost-effective outbreak detection in networks," in *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2007, pp. 420-429: ACM.
- [14] A. Goyal, W. Lu, and L. V. S. Lakshmanan, "CELF++: optimizing the greedy algorithm for influence maximization in social networks," presented at the Proceedings of the 20th international conference companion on World wide web, Hyderabad, India, 2011.
- [15] Y. Tang, X. Xiao, and Y. Shi, "Influence maximization: Near-optimal time complexity meets practical efficiency," in *Proceedings of the 2014 ACM SIGMOD international conference on Management of data*, 2014, pp. 75-86.
- [16] H. T. Nguyen, M. T. Thai, and T. N. Dinh, "A billion-scale approximation algorithm for maximizing benefit in viral marketing," *IEEE/ACM Transactions On Networking*, vol. 25, no. 4, pp. 2419-2429, 2017.
- [17] C. Borgs, M. Brautbar, J. Chayes, and B. Lucier, "Maximizing social influence in nearly optimal

آنلاین، یافتن افراد تأثیرگذار در یک شبکه اجتماعی کاربردهای عملی زیادی در بازاریابی، سیاست و حتی کنترل بیماری‌ها دارد. مسئله بیشینه‌سازی نفوذ متکی بر مدل انتشار و الگوریتم شناسایی رئوس تأثیرگذار به‌منظور بیشینه‌سازی نفوذ است. با توجه به این نکته که اندازه اکثر شبکه‌های اجتماعی بسیار بزرگ و پویا است، شناسایی افراد بانفوذ جهت گسترش نفوذ با استفاده از روش‌های سنتی، پیچیده و زمان‌بر خواهد بود. در این مقاله، ما الگوریتم کرکس توسعه‌یافته پویا را پیشنهاد کرده‌ایم. الگوریتم پیشنهادی با ایجاد تنوع و تعادل بین مراحل اکتشاف و بهره‌وری باعث افزایش عملکرد مدل پیشنهادی می‌شود. همچنین این الگوریتم با افزایش ابعاد مسئله بهینه‌سازی از کارایی مناسبی برخوردار است که در هر لحظه از برچسب زمانی مشخص، یک مجموعه بهینه از رئوس مؤثر جهت بیشینه‌سازی نفوذ افراد در شبکه‌های اجتماعی را مشخص می‌کند. از طریق نتایج تجربی، عملکرد الگوریتم پیشنهادی خود را با سایر الگوریتم‌های ارائه‌شده مقایسه کردیم، اگرچه عملکرد الگوریتم پیشنهادی تحت تأثیر پارامترهای آن قرار دارد اما نتایج آن بسیار بهتر از الگوریتم‌های شبیه‌سازی‌شده است ولی باوجود افزایش دقت شناسایی افراد بانفوذ در یک شبکه نسبتاً بزرگ هزینه‌های محاسباتی جهت ارزیابی نفوذ تحت مدل انتشار مورد استفاده بالا بوده و زمان‌گیر می‌باشد. در کارهای آینده، می‌توان به مطالعه حساسیت پارامترها و تنظیم پارامترهای الگوریتم پیشنهادی و تأثیر آن‌ها بر حفظ تعادل مطلوب بین فازهای اکتشاف و بهره‌برداری از فرایند بهینه‌سازی الگوریتم و ارائه مدل‌های انتشار مقیاس‌پذیر جهت بهبود هزینه محاسباتی پرداخت. همچنین می‌توان از معیارهای دیگری برای اندازه‌گیری مرکزیت رئوس استفاده نمود و از سایر روش‌های کشف اجتماعات اعم از اجتماعات مجزا و همپوشان را بررسی کرد.

۶- مراجع

- [1] N. Hafiene, W. Karoui, and L. B. Romdhane, "Influential nodes detection in dynamic social networks: A survey," *Expert Systems with Applications*, vol. 159, p. 113642, 2020.
- [2] J. J. Lotf, M. A. Azgomi, and M. R. E. Dishabi, "An improved influence maximization method for social networks based on genetic algorithm," *Physica A: Statistical Mechanics and its Applications*, vol. 586, p. 126480, 2022.
- [3] D. Kempe, J. Kleinberg, and É. Tardos, "Maximizing the spread of influence through a social network," in *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, 2003, pp. 137-146: ACM.
- [4] Z. Aghaee and S. Kianian, "Efficient influence spread estimation for influence maximization," *Social Network Analysis and Mining*, vol. 10, no. 1, pp. 1-21, 2020.

- its Applications*, vol. 486, pp. 968-985, 2017.
- [32] D. Cai, Z. Wang, N. Wang, and D. Wei, "A new method for identifying influential nodes based on DS evidence theory," in *2017 29th Chinese Control And Decision Conference (CCDC)*, 2017, pp. 4603-4609: IEEE.
- [33] E. B. A. Karimi, M. Nemati, M. Saleh Esfehiani, "Identifying Influential Nodes in Social Networks by Integrating the Centrality Method and Node Activity," *Journal of Electronical & Cyber Defence*, vol. 8, no. 3, pp. 1-11, 2020. (In Persian)
- [34] Q. Jiang, G. Song, C. Gao, Y. Wang, W. Si, and K. Xie, "Simulated annealing based influence maximization in social networks," in *Twenty-fifth AAAI conference on artificial intelligence*, 2011.
- [35] C.-W. Tsai, Y.-C. Yang, and M.-C. Chiang, "A genetic newgreedy algorithm for influence maximization in social network," in *2015 IEEE International Conference on Systems, Man, and Cybernetics*, 2015, pp. 2549-2554: IEEE.
- [36] M. Gong, J. Yan, B. Shen, L. Ma, and Q. Cai, "Influence maximization in social networks based on discrete particle swarm optimization," *Information Sciences*, vol. 367, pp. 600-614, 2016.
- [37] L. Cui *et al.*, "DDSE: A novel evolutionary algorithm based on degree-descending search strategy for influence maximization in social networks," *Journal of Network and Computer Applications*, vol. 103, pp. 119-130, 2018.
- [38] J. Tang *et al.*, "Maximizing the spread of influence via the collective intelligence of discrete bat algorithm," *Knowledge-Based Systems*, vol. 160, pp. 88-103, 2018.
- [39] B. Abdollahzadeh, F. S. Gharehchopogh, and S. Mirjalili, "African vultures optimization algorithm: A new nature-inspired metaheuristic algorithm for global optimization problems," *Computers & Industrial Engineering*, vol. 158, p. 107408, 2021.
- [40] J. Tang, R. Zhang, P. Wang, Z. Zhao, L. Fan, and X. Liu, "A discrete shuffled frog-leaping algorithm to identify influential nodes for influence maximization in social networks," *Knowledge-Based Systems*, vol. 187, p. 104833, 2020.
- [41] B. Abdollahzadeh, F. Soleimani Gharehchopogh, and S. Mirjalili, "Artificial gorilla troops optimizer: a new nature-inspired metaheuristic algorithm for global optimization problems," *International Journal of Intelligent Systems*, vol. 36, no. 10, pp. 5887-5958, 2021.
- [42] S. A. Myers and J. Leskovec, "The bursty dynamics of the twitter information network," in *Proceedings of the 23rd international conference on World wide web*, 2014, pp. 913-924.
- [43] D. Kempe, J. Kleinberg, and É. Tardos, "Maximizing the spread of influence through a social network," *Theory of Computing*, vol. Volume 11, p. 42, 2015.
- [44] H. Shayanfar and F. S. Gharehchopogh, "Farmland fertility: A new metaheuristic algorithm for solving continuous optimization problems," *Applied Soft Computing*, vol. 71, pp. 728-746, 2018.
- [45] E. Cohen, D. Delling, T. Pajor, and R. F. Werneck, "Computing classic closeness centrality, at scale," in *Proceedings of the twenty-fifth annual ACM-SIAM symposium on Discrete algorithms*, 2014, pp. 946-957: SIAM.
- [18] W. Chen, Y. Wang, and S. Yang, "Efficient influence maximization in social networks," presented at the Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining, Paris, France, 2009.
- [19] S. Cheng, H. Shen, J. Huang, G. Zhang, and X. Cheng, "Staticgreedy: solving the scalability-accuracy dilemma in influence maximization," in *Proceedings of the 22nd ACM international conference on Information & Knowledge Management*, 2013, pp. 509-518.
- [20] J. Lv, J. Guo, and H. Ren, "Efficient Greedy Algorithms for Influence Maximization in Social Networks," *JIPS*, vol. 10, no. 3, pp. 471-482, 2014.
- [21] N. Samadi and A. Bouyer, "Identifying influential spreaders based on edge ratio and neighborhood diversity measures in complex networks," *Computing*, vol. 101, no. 8, pp. 1147-1175, 2019.
- [22] J.-X. Zhang, D.-B. Chen, Q. Dong, and Z.-D. Zhao, "Identifying a set of influential spreaders in complex networks," *Scientific reports*, vol. 6, no. 1, pp. 1-10, 2016.
- [23] K. Berahmand, A. Bouyer, and N. Samadi, "A new local and multidimensional ranking measure to detect spreaders in social networks," *Computing*, vol. 101, no. 11, pp. 1711-1733, 2019.
- [24] D. Liu, Y. Jing, J. Zhao, W. Wang, and G. Song, "A fast and efficient algorithm for mining top-k nodes in complex networks," *Scientific reports*, vol. 7, no. 1, pp. 1-8, 2017.
- [25] Z.-L. Luo, W.-D. Cai, Y.-J. Li, and D. Peng, "A pagerank-based heuristic algorithm for influence maximization in the social network," in *Recent progress in data engineering and internet technology*: Springer, 2012, pp. 485-490.
- [26] S. Ahajjam and H. Badir, "Identification of influential spreaders in complex networks using HybridRank algorithm," *Scientific reports*, vol. 8, no. 1, pp. 1-10, 2018.
- [27] L. Qiu, X. Tian, S. Sai, and C. Gu, "LGIM: A global selection algorithm based on local influence for influence maximization in social networks," *IEEE Access*, vol. 8, pp. 4318-4328, 2019.
- [28] W. Li *et al.*, "Three-hop velocity attenuation propagation model for influence maximization in social networks," *World Wide Web*, vol. 23, no. 2, pp. 1261-1273, 2020.
- [29] Z. Aghaee and S. Kianian, "Influence maximization algorithm based on reducing search space in the social networks," *SN Applied Sciences*, vol. 2, no. 12, pp. 1-14, 2020.
- [30] Y. Zhao, S. Li, and F. Jin, "Identification of influential nodes in social networks with community structure based on label propagation," *Neurocomputing*, vol. 210, no. C, pp. 34-44, 2016.
- [31] F. Ullah and S. Lee, "Identification of influential nodes based on temporal-aware modeling of multi-hop neighbor interactions for influence spread maximization," *Physica A: Statistical Mechanics and*

- accuracy of influence maximization algorithms," *Social Network Analysis and Mining*, vol. 10, no. 1, p. 70, 2020/08/24 2020.
- [48] J. K. David Kempe, and Éva Tardos, "Maximizing the Spread of Influence through a Social Network," *Theory of Computing*, vol. Volume 11 pp. 105-147, April 22, 2015.
- presented at the Proceedings of the second ACM conference on Online social networks, Dublin, Ireland, 2014. Available: <https://doi.org/10.1145/2660460.2660465>
- [46] M. Riondato and E. M. Kornaropoulos, "Fast approximation of betweenness centrality through sampling," *Data Mining and Knowledge Discovery*, vol. 30, no. 2, pp. 438-475, 2016.
- [47] H. Kingi *et al.*, "A numerical evaluation of the