Journal of Soft Computing and Information Technology (JSCIT)
Babol Noshirvani University of Technology, Babol, Iran
Journal Homepage: jscit.nit.ac.ir
DOI: 10.22034/jscit.2025.219059
Volume 14, Number 1, Spring 2025, pp. 11-26
Received: 12/17/2024, Revised: 03/09/2025, Accepted: 04/12/2025

Identifying Key Influencers in Social Networks: A New Neighborhood-Based Method

Yahya Dorostkar Navaei[1], Mohammad Hossein Rezvani[2*] and Amir Masoud Eftekhari Moghadam[3]
1- Department of Computer and Information Technology Engineering, Qa.C., Islamic Azad University, Qazvin, Iran.
2*- Department of Computer and Information Technology Engineering, Qa.C., Islamic Azad University, Qazvin, Iran.
3- Department of Computer and Information Technology Engineering, Qa.C., Islamic Azad University, Qazvin, Iran.
[1]y.dorostkar@iau.ac.ir, [2*]mh.rezvani@iau.ac.ir, and [3]amireftekhari@iau.ac.ir

Corresponding author's address: Mohammad Hossein Rezvani, Department of Computer and Information Technology Engineering, Qa.C., Islamic Azad University, Qazvin, Iran.

Abstract- One of the key challenges for social network designers is to identify influential users. Such users, by influencing other users, can play an important role in the rapid spread of information in the network. The goal here is influence maximization. This means finding the smallest number of users who have the greatest influence on other users. Although traditional metrics, such as degree, can indicate the importance of a node among its immediate neighbors in the network, they ignore the importance of a node's neighbors. An influential node found by traditional metrics is likely to be in the neighborhood of weakly connected nodes. This can slow down the spread of information, especially in advertising applications. To address this problem, in this paper, we present a new metric called neighborhood similarity. It determines the social importance of a user based on a combination of information about his/her neighbors. Due to the NP-hardness of the problem, we use our newly developed metric with a well-known metaheuristic method called Non-dominated Sorting Genetic Algorithm (NSGA-III). To reduce the cost, a centrality threshold is used so that only nodes with high connections act as influential users. We use sorting of users based on social similarity criteria to improve exploration and exploitation. Extensive tests are performed under different scenarios with a limited number of influential users on four well-known datasets from Caltech, Reed, Simmons, and Haverford. The results of the experiments show that our proposed neighborhood similarity metric can identify influential users better than traditional metrics. It improves the number of affected users by 9.5%, 16.5%, and 6.5%, respectively, in terms of centrality, closeness, and betweenness criteria.

Keywords- Social network; influence maximization; neighborhood similarity; metaheuristics; NSGA-III.

مجله علمی رایانش نرم و فناوری اطلاعات
دانشگاه صنعتی نوشیروانی بابل
صفحه مجله: jscit.nit.ac.ir
شناسه دیجیتال: 10.22034/jscit.2025.219059
جلد ۱۴، شماره ۱، بهار ۱۴۰۴ ، صفحه ۱۱-۲۶
دریافت: ۱۴۰۳/۰۹/۲۷، بازنگری: ۱۴۰۳/۱۲/۱۹، پذیرش: ۱۴۰۴/۰۱/۲۳

شناسایی تأثیرگذاران کلیدی در شبکه‌های اجتماعی: یک روش جدید مبتنی بر همسایگی

یحیی درستکار نوایی[۱]، محمدحسین رضوانی*[۲]، امیر مسعود افتخاری مقدم[۳]
۱- دانشکده مهندسی کامپیوتر و فناوری اطلاعات، دانشگاه آزاد اسلامی، واحد قزوین، قزوین، ایران.
*۲- دانشکده مهندسی کامپیوتر و فناوری اطلاعات، دانشگاه آزاد اسلامی، واحد قزوین، قزوین، ایران.
۳- دانشکده مهندسی کامپیوتر و فناوری اطلاعات، دانشگاه آزاد اسلامی، واحد قزوین، قزوین، ایران.
[1]y.dorostkar@iau.ac.ir, [2*]mh.rezvani@iau.ac.ir, [3]amireftekhari@iau.ac.ir

* نشانی نویسنده مسئول: محمدحسین رضوانی، قزوین، دانشکده مهندسی کامپیوتر و فناوری اطلاعات، دانشگاه آزاد اسلامی، واحد قزوین.

چکیده- یکی از چالش‌های کلیدی برای طراحان شبکه‌های اجتماعی، شناسایی کاربران تاثیرگذار است. چنین کاربرانی با تأثیرگذاری بر سایر کاربران می توانند نقش مهمی در انتشار سریع اطلاعات در شبکه داشته باشند. هدف در اینجا به حداکثر رساندن تأثیر است. این به معنای یافتن کمترین تعداد کاربرانی است که بیشترین تأثیر را بر سایر کاربران دارند. اگرچه معیارهای سنتی، مانند درجه، می‌توانند اهمیت گره در میان همسایگان بلافصل خود در شبکه را نشان دهند ولی اهمیت همسایگان یک گره را نادیده می‌گیرند. یک گره تأثیرگذار که توسط معیارهای سنتی یافت می‌شود احتمالاً در همسایگی گره‌های ضعیف متصل قرار دارد. این امر می‌تواند انتشار اطلاعات را به ویژه در برنامه‌های تبلیغاتی با مشکل روبرو کند. برای پرداختن به این مشکل، در این مقاله، معیار جدیدی به نام شباهت همسایگی ارائه می‌کنیم. این معیار، اهمیت اجتماعی یک کاربر را بر اساس ترکیبی از اطلاعات در مورد همسایگان او تعیین می کند. با توجه به سختی NP مسئله، ما از معیار جدید توسعه یافته خود با یک روش فراابتکاری معروف به نام الگوریتم ژنتیک مرتب‌سازی غیرمسلط (NSGA-III) استفاده می‌کنیم. برای کاهش هزینه، از آستانه مرکزیت استفاده می‌شود تا فقط گره‌هایی با اتصالات بالا به عنوان کاربران تأثیرگذار عمل کنند. ما از مرتب‌سازی کاربران بر اساس معیارهای شباهت اجتماعی برای بهبود اکتشاف و بهره‌برداری استفاده می‌کنیم. آزمایش‌های گسترده تحت سناریوهای مختلف با تعداد محدودی از کاربران تأثیرگذار بر روی چهار مجموعه داده معروف از Caltech، Reed، Simmons و Haverford انجام می‌شود. نتایج آزمایش‌ها نشان می‌دهد که معیار شباهت همسایگی پیشنهادی ما می‌تواند کاربران تأثیرگذار را بهتر از معیارهای سنتی شناسایی کند. معیار پیشنهادی ما، تعداد کاربران تحت تأثیر را به ترتیب ۹,۵٪، ۱۶,۵٪ و ۶,۵٪ از نظر معیارهای مرکزیت، نزدیکی و بینابین بودن بهبود می‌بخشد.

واژه‌های کلیدی: شبکه اجتماعی، به حداکثر رساندن تأثیر؛ شباهت همسایگی؛ فراابتکاری؛ NSGA-III.

۱- مقدمه

حداکثرسازی نفوذ[۱]، یک مسئله بهینه‌سازی چند هدفه برای یافتن کمترین تعداد کاربرانی است که بیشترین نفوذ را در شبکه دارند [۱]. یک چالش عمده در این مساله، یافتن روش‌های مقیاس‌پذیر[۲] است. افزایش تعداد کاربران در شبکه‌های اجتماعی می‌تواند منجر به کاهش مقیاس‌پذیری و افزایش هزینه محاسباتی شود [۲, ۳]. همچنین با افزایش تعداد کاربران، میزان محاسبه برای تعیین گره‌های تاثیرگذار افزایش می‌یابد. حداکثرسازی نفوذ، یک مساله

NP-hard است زیرا با رشد شبکه، تعداد راه‌حل‌ها به طور تصاعدی افزایش می‌یابد [۴].

در سال‌های اخیر، محققان علاقه زیادی به استفاده از روش‌های فراابتکاری[۳] نشان داده‌اند. این روش‌ها به طور کلی قادر به دستیابی به راه حل‌های نزدیک به بهینه با بهره‌گیری از سرعت همگرایی بالا هستند [۵]. روش‌های فراابتکاری به جای جستجوی کل فضای راه‌حل، تنها مکان‌های ممتاز را با ایجاد تعادل بین اکتشاف[۴] و بهره‌برداری[۵] جستجو می‌کنند. در اینجا، اکتشاف به جستجو به منطقه ناشناخته امکان‌پذیر اشاره دارد، در حالی که بهره برداری به جستجوی همسایگی یک منطقه امیدوارکننده اشاره دارد [۶]. ایجاد تعادل بین درجه اکتشاف و بهره‌برداری می‌تواند به یافتن راه‌حل‌های خوب منجر شود. هنگامی که یک گره تأثیرگذار پیدا می‌شود، گره‌های مجاور آن ممکن است تأثیرگذار نباشند. این یکی از مشکلات اصلی روش‌های فراابتکاری است[۷]. برای غلبه بر این مشکل می‌توان کاربران را بر اساس معیارهای مهم رتبه‌بندی کرد. با رتبه‌بندی و فهرست‌بندی مجدد کاربران، می‌توان آنها را با ارتباطات بیشتری گروه‌بندی کرد. یک گره تأثیرگذار که بر اساس این معیارها یافت شود ممکن است توسط گره‌هایی با اتصالات ضعیف احاطه شود. وجود چنین گره‌های ضعیفی ممکن است انتشار محتوا را کاهش دهد.

در این مقاله، ما روشی را برای به حداکثر رساندن تأثیر بر اساس الگوریتم ژنتیک مرتب‌سازی غیرمسلط NSGA-III[۶] ارائه می‌کنیم. برای این منظور، ما از معیار پیشنهادی خود، NS[۷]، در ارتباط با معیارهای سنتی مبتنی بر مرکزیت استفاده می‌کنیم. این یک رویکرد بهینه‌سازی حداقل-حداکثری است که اهداف آن شامل کاهش تعداد کاربران، افزایش توزیع محتوا در شبکه و کاهش هزینه محاسباتی است. ما معیار نوآورانه خود را با نام شباهت همسایگی (NS) [۸] با معیارهای مهمی مانند مرکزیت[۸]، نزدیکی[۹]، و بینابین بودن[۱۰] مقایسه می‌کنیم. برای کاهش پیچیدگی محاسباتی، از یک آستانه برای معیار مرکزیت برای انتخاب جمعیت اولیه استفاده می‌کنیم. جمعیت اولیه از کاربرانی که مرکزیت آنها بالاتر از آستانه است انتخاب می‌شود. در نهایت، یک مقایسه تحلیلی با سناریوهای مختلف برای ارزیابی عملکرد معیار NS در برابر معیارهای مبتنی بر مرکزیت ارائه شده است. مشارکت‌های اصلی این مقاله به شرح زیر است

• ما یک معیار تشابه همسایگی جدید برای تعیین کاربران تأثیرگذار در شبکه‌های اجتماعی تعریف می‌کنیم. این معیار به کاربران با نفوذی می‌پردازد که همسایگانشان ارتباطات قوی دارند. ما نشان می‌دهیم که انتشار محتوا در شبکه اجتماعی از طریق چنین گره‌هایی تسریع می‌شود.

• ما مسئله چندهدفه حداکثرسازی نفوذ را بر اساس مجموع وزنی معیارهای مرکزیت، نزدیکی، بینابین بودن و شباهت همسایگی مدل می‌کنیم. سپس آن را با استفاده از NSGA-III [۹] برای به حداکثر رساندن تأثیر و کاهش هزینه حل می‌کنیم. الگوریتم ارائه شده در این مقاله می‌تواند جایگزین الگوریتم‌های قبلی در شبکه‌های اجتماعی برای انتشار محتوا شود. روش پیشنهادی ما، دو تغییر مهم در NSGA-III کلاسیک ایجاد می‌کند: الف) مرتب‌سازی کاربران برای حل مشکل جستجوی محلی و ب) تعیین آستانه برای کاهش حجم اولیه جمعیت و در نتیجه کاهش هزینه.

• ما روش پیشنهادی را بر روی مجموعه داده‌های مختلف و با همه جایگشت‌های[۱۱] ممکن معیارها ارزیابی می‌کنیم. نتایج ما اثربخشی روش پیشنهادی را در تمام سناریوهای دنیای واقعی اثبات می‌کند.

ادامه این مقاله به شرح زیر سازماندهی شده است: بخش ۲ تحقیقات مرتبط قبلی را در خصوص بیشینه‌سازی تأثیر بررسی می‌کند. بخش ۳ پیشینه معیارهای سنتی را ارائه می‌دهد. بخش ۴ جزئیات روش پیشنهادی را ارائه می‌کند. بخش ۵ عملکرد روش پیشنهادی را ارزیابی می‌کند. در نهایت، بخش ۶ مقاله را به پایان می‌رساند.

۲- کارهای مرتبط

در این بخش، مقالات قبلی در مورد به حداکثر رساندن تأثیر، همراه با مزایا و معایب هر یک بررسی می‌شود. خلاصه‌ای از این مقالات در جدول ۱ آورده شده است. این موضوع دارای کاربردهایی با پتانسیل تجاری بالا است [۱۰]. محققان از معیارهای مختلفی با اهمیت اجتماعی مانند مرکزیت، نزدیکی و بینابین بودن استفاده کرده‌اند [۵].

عباس و نواف [۱۱] از NSGA-II برای به حداکثر رساندن نفوذ استفاده کردند. برای انتخاب جمعیت اولیه، آنها به صورت شبه‌تصادفی گره‌ها را با استفاده از معیارهای مرکزیت انتخاب می‌کنند. برای این منظور از گره‌هایی استفاده می‌کنند که مرکزیت، نزدیکی و بردار ویژه بهتری دارند. به حداکثر رساندن پوشش نفوذ و به حداقل رساندن تعداد گره‌های بذر[۱۲] دو هدف متضاد هستند که در این تحقیق با هم استفاده می‌شوند. NSGA-II برای مسائلی با حداکثر سه هدف بهتر عمل می‌کند. برای مسائل بزرگ-مقیاس با بیش از سه هدف، معمولا NSGA-III [۹] بهتر عمل می‌کند. یانگ و همکارانش [۱۰] با ملاحظات تأثیر بین گره‌ها و مبادله بین هزینه تعادل ساختاری و کیفیت جامعه در شبکه، یک مدل تعادل

جدول ۱: مقایسه تحقیقات در مورد حداکثرسازی تأثیر در شبکه‌های اجتماعی.

مرجع	مزایا	معایب
یانگ و همکاران [۱]	ارزیابی تأثیر همبستگی موجودیت بر انتشار اطلاعات بر اساس توالی رفتار تاریخی کاربران	مشکل جمع‌آوری داده‌های بزرگ در روش یادگیری عمیق
زارعی و همکاران [۲]	استفاده از یک متغیر فازی برای نشان دادن میزان تأثیرپذیری کاربر از یک پیام (سطح فعال سازی کاربر).	عدم تمرکز بر تعداد کاربران با نفوذ و همسایگان تحت نفوذ قرار گرفته
بیسواس و همکاران [۱۲]	ارزیابی تأثیر همبستگی موجودیت بر انتشار اطلاعات بر اساس توالی رفتار تاریخی کاربران	نادیده گرفتن تأثیر گره‌ها در همسایگی یک گره تأثیرگذار
لطفی و ازگمی[۱۰]	استفاده از معیار مرکزیت برای شناسایی کاربران تاثیرگذار با الگوریتم ژنتیک تعمیم یافته پویا	تمرکز بر هدف افزایش پوشش شبکه توسط کاربران تاثیرگذار و عدم توجه به تعداد کاربران تاثیرگذار
عباس و نواف [۱۱]	تلفیق معیارهای اهمیت اجتماعی و استفاده از آنها به عنوان ضریب تابع تناسب در NSGA-II	نادیده گرفتن تأثیر گره‌ها در همسایگی یک گره تأثیرگذار
چاترجی، بی، و همکاران. [۱۳]	استفاده از یک استراتژی جستجوی محلی برای یافتن گره‌های تاثیرگذار بر اساس جهش قورباغه مختلط	- تنها تمرکز بر هدف افزایش پوشش شبکه دارد - عدم استفاده از تلفیق معیارهای اهمیت اجتماعی در عملکرد برازندگی
بیسواس و همکاران [۴]	حذف گره‌های کم نفوذ در مرحله مقدماتی در رویکرد تصمیم‌گیری چند معیاره (MCDM) و نسخه اصلاح شده تبرید شبیه‌سازی شده برای کاهش هزینه	- عدم استفاده از تابع برازندگی چند هدفه برای بهینه‌سازی تعداد کاربران تأثیرگذار و پوشش شبکه - عدم توجه به گره‌های تحت نفوذ همسایه
لی و همکاران [۱۴]	استفاده از مرکزیت درجه و راه رفتن تصادفی برای استراتژی جستجو در الگوریتم جستجوی گسسته کلاغ	- عدم استفاده از تابع برازندگی چند هدفه برای بهینه‌سازی تعداد کاربران تأثیرگذار و پوشش شبکه - عدم توجه به گره‌های تحت نفوذ همسایه
کانتینی و همکاران [۳]	استفاده از درجه مرکزیت در کلونی زنبورهای مصنوعی وزن‌دار	تمرکز تنها بر هدف افزایش پوشش شبکه توسط کاربران تأثیرگذار و عدم توجه به تعداد کاربران تأثیرگذار و تحت نفوذ همسایگی
کارچمارچیک و همکاران [۶]	استفاده از ویژگی‌های کاربران در شبکه اجتماعی برای هدف قرار دادن بهتر گروه‌های خاصی از کاربران	- عدم تمرکز بر تعداد کاربران با نفوذ - عدم استفاده از تابع برازندگی چندهدفه برای به حداکثر رساندن تأثیر
هان و همکاران [۱۵]	استفاده از معیار مرکزیت به عنوان وزن تابع تناسب اندام در الگوریتم خفاش گسسته	فقط از تابع تناسب اندام تک هدف برای افزایش پوشش استفاده کنید

ساختاری چندهدفه جدید پیشنهاد دادند. آنها برای بهینه‌سازی مدل جدید، یک الگوریتم بهبود یافته بر اساس NSGA-II توسعه دادند. برای این منظور، آنها عملگر جستجوی محلی پیشنهادی را در NSGA-II اصلی ادغام کردند.

بیسواس و همکاران [۱۲] مسئله را با سه هدف متناقض با محدودیت‌های دنیای واقعی فرموله کردند. آنها از یک الگوریتم تکاملی مبتنی بر تسلط به نام DEA-II استفاده کردند. آنها از یک عملگر انتخاب مبتنی بر شکاف برای دستیابی به تعادل بهتر بین واگرایی و همگرایی در DEA-II استفاده کردند. آنها علاوه بر عادی‌سازی، تغییراتی در فرآیند مرتب‌سازی با استفاده از معیارهای اهمیت اجتماعی ایجاد کردند. چاترجی و همکاران [۱۳] از الگوریتم جهش قورباغه مختلط برای تشخیص جامعه و IM استفاده کردند. آنها از استراتژی‌های جستجوی محلی مانند تپه‌نوردی با بهره‌برداری از تمرکز کاربر استفاده می‌کنند. بیسواس و همکاران [۴] یک رویکرد تصمیم‌گیری چند معیاره (MCDM)[۱۳] برای حل مساله در شبکه‌های اجتماعی پیشنهاد می‌کنند. در مرحله مقدماتی، روش آنها گره‌های کم نفوذ را با استفاده از متریک مرکزیت برای کاهش هزینه محاسباتی حذف می‌کند. سپس، یک نسخه اصلاح شده از تبرید شبیه‌سازی شده با یک استراتژی جستجوی پیشرفته برای یافتن راه حل بهینه استفاده می شود. لی و همکاران [۱۴] از الگوریتم جستجوی کلاغ گسسته با مکانیزم کدگذاری جدید و قوانین تکامل بهبود قابلیت جستجو اتخاذ می کنند. کانتینی و همکاران [۳] از کلونی زنبورهای مصنوعی وزن‌دار برای شناسایی زیرمجموعه‌ای از کاربران تأثیرگذار استفاده کنند. آنها از تکنیک‌های رتبه‌بندی بر اساس معیارهای مرکزیت کلاسیک استفاده می کنند. کارچمارچیک و همکاران [۶] سعی می‌کنند با استفاده از ویژگی‌های کلیدی کاربران مانند سن و جنسیت، گره‌های تأثیرگذار را بیابند. آنها از تمام معیارهای اصلی برای رتبه‌بندی اهمیت کاربران در گروه‌های مختلف استفاده می‌کنند.هان و همکاران [۱۵] از الگوریتم خفاش گسسته[۱۴] و تقسیم بندی شبکه استفاده می کنند. نتایج آنها نشان می دهد که که این روش در هر اجرا به یک مقدار تخمین نفوذ محلی همگرا می‌شود. این، سبب حذف نوساناتی می‌شود که توسط الگوریتم خفاش بر روی مقدار تناسب ایجاد شده بود. در این تحقیق از معیارهای مرکزیت به عنوان عامل جستجوی محلی در تابع برازندگی استفاده شده است.

به طور خلاصه، بررسی مقالات فوق ما را به نتایج زیر می رساند:

• با توجه به روش‌های بررسی شده در جدول ۱، می‌توان به معیارهای اهمیت اجتماعی مانند مرکزیت و مشتقات آن دست یافت

که ابزاری شناخته شده برای رتبه‌بندی و وزن‌دهی کاربران است. محققان به معیار مرکزیت توجه زیادی دارند. این نشان می‌دهد که کاربرانی که تعداد ارتباط بالایی دارند می‌توانند تأثیر زیادی در انتشار محتوا داشته باشند.

• اگر علاوه بر مرکزیت، بتوانیم روشی پیدا کنیم که اهمیت گره‌های متصل به کاربر را نشان دهد، اثربخشی بسیار بالاتر خواهد بود. با توجه به ماتریس مجاورت، کاربران با درجه مرکزییت یکسان را می‌توان مشابه در نظر گرفت. این ایده اصلی ما در این مقاله است. کاربرانی که تعداد زیادی از کاربران هم‌سطح یا مشابه در همسایگی خود دارند، می‌توانند نفوذ بیشتری در شبکه داشته باشند. در بخش‌های بعدی، آزمایش‌های گسترده‌ای را برای اثبات ادعای خود با سناریوهای مختلف دنیای واقعی انجام خواهیم داد.

۳- مفاهیم بنیادین

در این بخش معیارهای سنتی مورد استفاده برای تعیین اهمیت کاربران در شبکه‌های اجتماعی را ارائه می‌کنیم. همانطور که در بخش قبل توضیح داده شد، مرکزیت، نزدیکی و بینابین بودن، سه معیاری هستند که در تحقیقات قبلی بسیار مورد استفاده قرار گرفته‌اند. بعداً، معیار جدیدی از شباهت همسایگی را که برای گراف‌های شبکه‌های اجتماعی ایجاد کرده‌ایم، معرفی خواهیم کرد.

۳-۱- معیار مرکزیت

معیار مرکزیت برای شناسایی تاثیرگذارترین کاربران در شبکه‌های اجتماعی استفاده می‌شود. در اینجا اهمیت یک گره با توجه به درجه آن مشخص می‌شود. فرض کنید $G=(V,E)$ یک شبکه اجتماعی را نشان دهد، که در آن، V مجموعه‌ای از n کاربر است ($|V|=n$). همچنین مجموعه E نشان دهنده ارتباط بین کاربران و m تعداد یال‌ها در شبکه است ($|E|=m$). اکنون، یک شبکه اجتماعی به صورت یک ماتریس متقارن A، به نام ماتریس مجاورت، با ابعاد $n \times n$ نمایش داده می‌شود. در این ماتریس، اگر مقدار هر عنصر a_{ij} برابر با ۱ باشد، نشان دهنده وجود رابطه بین گره i و گره j است. بر این اساس، معیار مرکزیت برای هر کاربر به صورت زیر محاسبه می‌شود [۷]:

$$C(i) = \sum_{i=1}^{n} \sum_{j=1}^{n-1} \frac{a_{ij}}{n-1} \qquad (۱)$$

با استفاده از رابطه فوق می توان مقدار درجه مرکزیت را برای همه کاربران شبکه به دست آورد. کاربرانی که تعداد اتصالات بالایی دارند، مرکزیت بالایی دارند. این نشان دهنده قدرت کاربر و فالوورهای او است. چنین کاربرانی به شدت در معرض ارسال پیام و محتوا در شبکه اجتماعی هستند. برعکس، کاربری با درجه مرکزییت پایین، چندان محبوب نیست و شخصیت درونگرای او را منعکس می‌کند. یکی از معایب معیار مرکزیت، دسترسی محلی به توپولوژی شبکه است. این معیار، از دانش محلی محدودی برای تعیین اهمیت کاربران استفاده می کند.

۳-۲- معیار نزدیکی

ایده «معیار نزدیکی» این است که "قدرت یک کاربر در یک شبکه اجتماعی با مجموع فاصله او از افراد دیگر نسبت معکوس دارد" [۷]. به عبارت دیگر، نزدیکی برای یک کاربر بر اساس مجموع مسیرهای موجود از سایر کاربران به او در شبکه اجتماعی به دست می‌آید. بر این اساس، نزدیکی یک کاربر i به صورت زیر محاسبه می‌شود [۷]:

$$CN(i) = \frac{1}{\sum_{i=1}^{n} \sum_{j=1}^{n-1} d_{ij}} \qquad (۲)$$

در رابطه فوق، d_{ij} نشان دهنده طول کوتاه ترین مسیر از یک گره i به گره j است. این معیار به طور کمی میانگین فاصله بین یک کاربر i و سایر کاربران در شبکه اجتماعی را نشان می‌دهد. کاربرانی که ارزش CN بالاتری دارند، محتوا را از هر بخشی از شبکه در زمان کوتاه‌تری دریافت می‌کنند. یکی از معایب این معیار این است که در شبکه‌های غیرهمبند کار نمی‌کند. به عبارت دیگر، این معیار، فقط بخشی از شبکه را که به آن متصل است پوشش می‌دهد.

۳-۳- معیار بینابین بودن

متریک بینابین بودن، اهمیت یک گره را متناسب با تعداد کوتاه‌ترین مسیرهای بین گره‌های دیگر که از آن عبور می‌کنند، تعریف می‌کند. به عبارت دیگر، بینابین بودن برای یک کاربر i متناسب با کسری از کوتاه‌ترین مسیرها بین همه کاربران ممکن در شبکه است، جایی که کاربر i در مسیر بین آنها قرار دارد. معیار بینابین بودن برای هر کاربر i به صورت زیر محاسبه می‌شود [۷]:

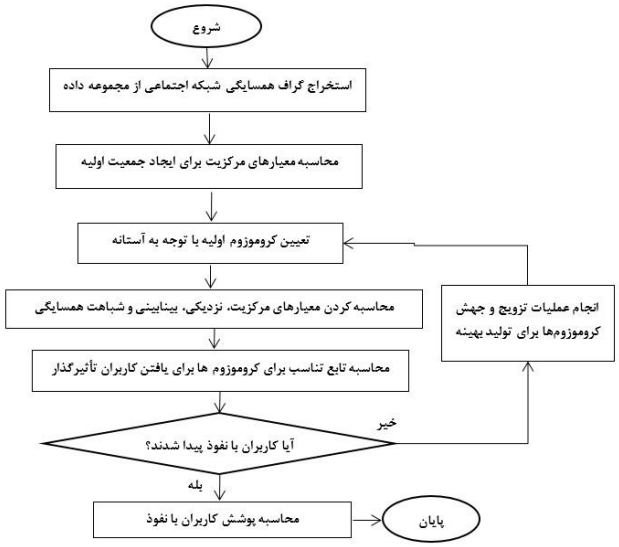
$$B(i) = \sum_{i=1}^{n} \sum_{j=1}^{n-1} \sum_{k=1}^{n-2} \frac{\sigma_{jk}(i)}{\sigma_{jk}} \qquad (۳)$$

در رابطه فوق، $\sigma_{jk}(i)$ نشان دهنده تعداد کوتاهترین مسیرها از گره j به مسیری که از گره k می‌گذرد است. همچنین، σ_{jk} تعداد کل کوتاه ترین مسیرها از گره j به k را نشان می‌دهد. معیار بینابین بودن فرض می‌کند که انتقال داده در شبکه اجتماعی بین کاربران از طریق کوتاه‌ترین مسیرهای بین آنها صورت می‌گیرد. این

معیار نشان دهنده قدرت دلالی یک فرد در یک شبکه اجتماعی است. به عبارت دیگر، نشان دهنده میزان بار ارتباطی مورد انتظاری است که کاربر باید تحمل کند. یک کاربر با مقدار بینابینی بالا ممکن است کنترل بیشتری بر اطلاعاتی که در سراسر شبکه جریان دارد داشته باشد. چنین کاربری ممکن است کاندیدای خوبی برای تبدیل شدن به یک فرد تأثیرگذار باشد.

۴- روش پیشنهادی

همانطور که گفته شد، یکی از مشارکت‌های ما در این مقاله طراحی یک معیار جدید به نام شباهت همسایگی است. در این قسمت پس از معرفی مشخصات این معیار، مشکل بیشینه‌سازی تاثیر را فرموله کرده و با NSGA-III حل می‌کنیم. در بخش‌های بعدی، جزئیات تغییراتی را که در NSGA-III انجام داده‌ایم شرح می‌دهیم. شکل ۱ نمودار جریانی روش پیشنهادی را نشان می‌دهد.



شکل ۱: نمودار جریانی روش پیشنهادی

بر اساس شکل ۱ می‌توان دید، در روش پیشنهادی در گام اول ماتریس مجاورتی از مجموعه داده‌های شبکه‌های اجتماعی استخراج می‌شود. در گام بعدی معیارهای اهمیت اجتماعی شامل مرکزیت درجه، نزدیکی، بینابینی و معیار شباهت همسایگی پیشنهادی برای هر یک از کاربران محاسبه می‌شود. سپس کاربرانی که احتمال کمی برای تبدیل شدن به کاربر تأثیرگذار در شبکه دارند، بر اساس آستانه مسانگین مرکزیت حذف می‌شوند. کاربران باقیمانده به عنوان جمعیت اولیه برای الگوریتم NSGA-III در نظر گرفته می شوند تا از میان آن‌ها، کاربران تأثیرگذار توسط الگوریتم-NSGA IIIشناسایی شوند. الگوریتم NSGA-III به منظور بررسی جمعیت از تابع تناسب چندهدفه ای استفاده می‌کند که سعی در کاهش تعداد کاربران، افزایش انتشار اطلاعات در شبکه و کاهش هزینه محاسباتی دارد. پارامترهای این تابع تناسب شامل ترکیبی از معیارهای اهمیت اجتماعی و میزان انتشار اطلاعات در شبکه اجتماعی است. در نهایت کاربران تأثیرگذار در شبکه بر اساس تعداد کاربرانی که تحت تأثیر قرار داده‌اند، ارزیابی می‌شوند.

۴-۱- معیار شباهت همسایگی پیشنهادی

ایده اصلی معیار شباهت همسایگی طراحی شده در این مقاله بسیار ساده است. اگر یک کاربر تأثیرگذار توسط کاربران مشابه و بسیار مرتبط احاطه شود، ممکن است انتشار محتوا توسط او تسریع شود. این به نوبه خود منجر به اطلاعات بیشتر کاربران در مورد شبکه می‌شود. از آنجایی که در گراف‌های وزنی غیرجهت‌دار، تنها راه بررسی شباهت گره‌ها درجه اتصال آنهاست، گره‌های مشابه دارای درجه یکسانی در نظر گرفته می‌شوند. شباهت همسایگی برای یک کاربر متناسب با تعداد گره‌های هم‌درجه است که مستقیماً به او متصل هستند. بنابراین، فرض می‌کنیم که کاربر i مستقیماً با p کاربر دیگر در ارتباط است. بنابراین میزان ارتباط کاربر iدر ارتباط مستقیم با کاربران همسایه به صورت زیر محاسبه می شود:

$$D(i) = \frac{|P|}{|N|},\ P \subseteq N \qquad (۴)$$

در رابطه فوق، P نشان دهنده همسایگان کاربر i است که زیرمجموعه‌ای از کل کاربران شبکه، یعنیN، است. حال می‌خواهیم تأثیر درجه همسایه‌ها را بر اینفلوئنسر i بر اساس استقرای ریاضی اثبات کنیم.

فرض کنید s_{ij} تعداد همسایگان هم‌درجه کاربر i را نشان دهد. همچنین، d_i نشان دهنده درجه کاربران هم‌درجه در همسایگی کاربر i است. ضمناً p نشان دهنده تعداد همسایگان مستقیم کاربر i باشد. در این حالت درجه شباهت همسایگی کاربر i به صورت زیر بدست می‌آید:

$$SN(i) = \sum_{i=1}^{n} \frac{\prod_{j=1}^{k} p_{ij} u_{ij}}{n} \qquad (۵)$$

در رابطه فوق، u_{ij} تعداد همسایگان هم‌درجه با کاربر i را نشان می‌دهد. همچنین، p_{ij} نشان دهنده احتمال انتشار اطلاعات در کاربران سطح j و n نشان دهنده تعداد همسایگان مستقیم کاربر i است. کاربران با مقدار NS بالا دارای همسایگان مشابه با اتصالات بالا هستند. این می‌تواند به انتشار اطلاعات در شبکه کمک کند. معیار جدیدی که ما ابداع کرده‌ایم می‌تواند کاربردهای جالبی در شبکه‌های اجتماعی داشته باشد. به طور مثال، می‌تواند به طور بالقوه برای کشف جوامع کاربران مشابه استفاده شود.

۴-۲- فرموله کردن مساله

اتصالات بین کاربران در شبکه اجتماعی با گراف$G=(V,E)$ نشان داده می‌شود که در آن،V و E به ترتیب رئوس و لبه‌های گراف هستند. هر رأس گراف G نشان دهنده یک کاربر و هر یال نشان دهنده ارتباط بین یک جفت کاربر است. اندازه شبکه بر اساس تعداد کاربران شبکه $|V|=N$، و تعداد پیوندها $|E|=M$، تعریف می‌شود. ساختار شبکه به صورت یک ماتریس مجاورت $\mathbf{A}=(a_{ij})_{n\times n}$ نشان داده می‌شود که در آن هر عنصر می‌تواند یکی از مقادیر {0, 1} را داشته باشد. اگر کاربر i به یک کاربر j متصل باشد آنگاه $a_{ij}=1$، در غیر این صورت $a_{ij}=0$. برای یافتن گره‌ها، یک مجموعه$S\subset V(G)$ متشکل از K گره انتخاب می‌شود. این گره‌ها می‌توانند اطلاعات را به بیشترین تعداد گره در سریع‌ترین زمان ممکن منتشر کنند. برای این منظور، تعداد K کاربر که اصطلاحاً کاربران بذر[۱۵] نامیده می‌شوند به عنوان کاربران تاثیرگذار بهینه در مجموعهS^* قرار می‌گیرند. اندازه مجموعهS^* با $|S^*|$ نشان داده می‌شود و مقدار آن برابر K است. مجموعهS^* به صورت زیر ایجاد می‌شود [۵]:

$$S^* = arg \max_{s\subset v,|s|=k} \sigma(S) \qquad (۱۲)$$

در رابطه فوق، $\sigma(S)$ یک تابع هدف است که در پایان فرآیند انتشار اطلاعات، گسترش مورد انتظار را روی گراف G ایجاد می‌کند. همچنین،S^* نشان دهنده K گره با حداکثر گسترش نفوذ در گراف شبکه اجتماعی است. مقادیر معیارهای اهمیت گره‌ها را می‌توان بر اساس درجه ورودی و خروجی آنها محاسبه کرد.

• مرتب سازی کاربران

روش‌های فراابتکاری به صورت محلی جستجو می‌کنند و پس از یافتن راه‌حل عملی، انتظار یافتن راه‌حل بهتر در همسایگی را دارند. به طور خاص، الگوریتم‌های ژنتیک با تطبیق والدین محلی همسایه، فرزندان جدیدی ایجاد می‌کنند. در مساله بیشینه‌سازی تاثیر، ممکن است روند عادی بهینه‌سازی به دلیل عدم نزدیکی به کاربران تأثیرگذار مختل شود. برای غلبه بر این مشکل از یک ترفند استفاده می‌کنیم. برای این منظور، کاربران را بر اساس معیارهای اهمیت مرتب می‌کنیم تا روند بهینه‌سازی عادی حفظ شود.

همانطور که قبلا ذکر شد، ما در روش پیشنهادی خود، از چهار معیار اهمیت شامل مرکزیت، نزدیکی، بینابین بودن و شباهت همسایگی برای رتبه‌بندی کاربران استفاده می‌کنیم. این معیارها به عنوان وزن‌های هدف در تابع برازندگی پیشنهادی استفاده می‌شوند. علاوه بر این، ما جمعیت اولیه را بر اساس این معیارها مرتب می‌کنیم و سپس آن را به عنوان ورودی به NSGA-III تغذیه می‌کنیم. ابتدا جمعیت اولیه بر اساس هر معیار مرتب شده و عملکرد مساله، ارزیابی می‌شود. سپس، تمام جایگشت‌های معیارها محاسبه شده و عملکرد مساله دوباره ارزیابی می‌شود. برای مرتب‌سازی کاربران از ترکیب وزنی معیارها استفاده می‌شود. پس از مرتب شدن لیست کاربران به ترتیب نزولی، کاربران تاثیرگذار از بالای لیست انتخاب می‌شوند.

• جمعیت اولیه

یکی از مهمترین دغدغه‌ها در روش‌های فراابتکاری، تنظیم جمعیت اولیه است. در اینجا، هر فرد (کروموزوم) در جمعیت اولیه یک راه حل را نشان می‌دهد. سطح شایستگی هر فرد با یک تابع برازندگی ارزیابی می‌شود. هدف ما یافتن K کاربر با بیشترین نفوذ در شبکه اجتماعی است. در روش پیشنهادی، تعداد جمعیت اولیه P برابر با ۱۰۰ در نظر گرفته شده است. هر فرد از جمعیت اولیه به عنوان یک کروموزوم به صورت برداری است که طول آن برابر با تعداد کاربران تأثیر گذار K در نظر گرفته می‌شود. به هر درایه از این بردار یک ژن گفته می‌شود. به یاد داشته باشید که K نشان دهنده تعداد کاربران تأثیرگذار و برابر با تعداد ژن ها در هر کروموزوم است. در روش‌های حداکثرسازی تأثیر، مقدار K به صورت بهینه توسط الگوریتم‌های فرااکتشافی تعیین می‌شود، اما در روش پیشنهادی به منظور بررسی میزان تأثیر معیار شباهت همسایگی بر تأثیر در شبکه‌های اجتماعی، تعداد کاربران تأثیرگذار در سناریوهای متفاوت به صورت محدود در نظر گرفته شده است. تعداد کاربران در این سناریوها از مقدار کمی شروع شده و رفته رفته افزایش می‌یابد تا تأثیر افزایش تعدادکاربران تأثیر گذار در شبکه به وضوح قابل رویت باشد. مقدار هر ژن نشان دهنده شاخص کاربر در شبکه اجتماعی است. شکل ۱ نمونه‌ای از کروموزوم را با ۵ کاربر تأثیرگذار نشان می‌دهد.

User	U_1	U_2	U_3	U_4	U_5
X	5	16	13	8	27

شکل ۲: نمونه ای از کروموزوم با ۵ کاربر تاثیرگذار

همانطور که در شکل ۲ نشان داده شده است، در این مثال کروموزوم اولیه دارای ۵ ژن است و هر ژن نشان دهنده شاخص یک کاربر است. کاربران در شبکه اجتماعی به صورت تصادفی به عنوان مقادیر ژن انتخاب می شوند. با توجه به تعداد زیاد کاربران در شبکه اجتماعی، تعداد راه‌های انتخاب ژن برای کروموزوم‌های اولیه و همچنین فضای جستجو برای NSGA-III بسیار زیاد است. برای حل این مشکل از یک آستانه بر اساس درجه مرکزیت کاربران استفاده می‌کنیم. شرط لازم برای اینفلوئنسر بودن هر کاربری، داشتن درجه بالایی از مرکزیت است. بنابراین، تنها کاربرانی که درجه مرکزیت آنها بیشتر از میانگین باشد، می‌توانند به عنوان کاندیدای جمعیت

اولیه انتخاب شوند. آستانه با استفاده از رابطه زیر محاسبه می شود:

$$\tau = \sum_{i=1}^{n} \frac{c_i}{n} \quad (۱۳)$$

در رابطه فوق، c_i مقدار درجه مرکزیت کاربران است و n تعداد کل کاربران شبکه اجتماعی است. کاربرانی با درجه مرکزیت کمتر از حد میانگین، احتمال بسیار کمی برای تبدیل شدن به کاربران تأثیرگذار دارند. بنابراین از فضای مساله حذف می‌شوند. با کاهش تعداد کاربرانی که در فضای مسئله الگوریتم NSGA-III قرار دارند، تعداد کاربرانی که باید برای یافتن کاربران تأثیر گذار مورد جستجو قرار گیرند، کاهش می یابد. از این رو هزینه الگوریتم از نظر پیچیدگی زمانی برای یافتن کاربران تأثیرگذار کاهش می‌یابد.

• تابع برازندگی پیشنهادی

پس از انتخاب تصادفی کاربران (کروموزوم‌ها) از بین افرادی که شرایط آستانه را برآورده می‌کنند، باید برازندگی آنها ارزیابی شود. ارزیابی کروموزوم‌های اولیه برای بررسی بهینه بودن راه‌حل‌های تصادفی اولیه انجام می‌شود. همانطور که قبلا توضیح داده شد، تابع برازندگی ترکیب وزنی از چهار معیار است: مرکزیت، نزدیکی، بینابین بودن و شباهت همسایگی. یک مشکل بالقوه با اجزای تابع برازندگی این است که آنها ممکن است دامنه‌های مختلفی از تغییرات داشته باشند. بنابراین، لازم است مقادیر هر یک از چهار عنصر تابع برازندگی، نرمال (یکنواخت) شود. برای این منظور از روش یکنواخت سازی min-max استفاده می‌کنیم که به صورت زیر تعریف شده است [۱۶]:

$$ND_i = \frac{D_i - D_{\min}}{D_{\max} - D_{\min}} \quad (۱۴)$$

در فرمول فوق، ND_i مقدار داده نرمال شده برای کاربر i را نشان می‌دهد و D_i مقدار داده اصلی را برای کاربر i نشان می‌دهد. همچنین، $D_{\min}$ و $D_{\max}$ به ترتیب حداقل و حداکثر داده را برای همه کاربران مشخص می‌کند. پس از نرمال‌سازی مقادیر معیارهای اهمیت، آنها به عنوان وزن برای تابع تناسب اعمال می‌شوند. وزن کاربر i به صورت زیر بدست می‌آید:

$$w_i = \sum_{i=1}^{n} \mu.(w_i^C + w_i^{CN} + w_i^B + w_i^{NS}) \quad (۱۵)$$

در فرمول فوق، w_i^C نشان دهنده وزن مرکزیت، w_i^{CN} وزن نزدیکی، و w_i^B وزن بینابینی و w_i^{NS} وزن معیار تشابه همسایگی است. همچنین μ نشان دهنده مقدار میانگین فاکتورهای وزنی است. برای یافتن کاربران تأثیرگذار، علاوه بر بررسی تناسب معیارهای فردی، از جایگشت آنها نیز استفاده می‌کنیم. برای ارزیابی تأثیر هر یک از معیارها به تنهایی، وزن بقیه معیارها باید برابر با صفر در نظر گرفته شود.

باتوجه به این که در روش پیشنهادی از مدل انتشار آبشاری استفاده شده است، ابتدا نیاز به تعریف رسمی این مدل به صورت زیر است:

- شبکه به صورت یک گراف G=(V,E) تعریف می‌شود که V گره‌ها (کاربران) و E یال‌ها (روابط) هستند.
- زمانی که گره u اقدام به انتقال اطلاعات بین همسایگان خود می‌نماید، یک مجموعه اولیه از گره‌ها (S) فعال می‌شوند (تحت تأثیر قرار می‌گیرند).
- هر گره فعال‌شده ، i، در یک گام زمانی، فرصتی مستقل برای فعال کردن همسایگان غیرفعال خود که با متغیر j مشخص می‌شوند با احتمال p_{ij} دارد.
- این فرآیند به‌صورت گسسته در گام‌های زمانی ادامه می‌یابد تا زمانی که هیچ گره جدیدی فعال نشود.
- هدف معمولاً بیشینه‌سازی تعداد گره‌های فعال‌شده (اندازه گسترش یا Spread) است.

از این رو برای برای تعریف تابع برازندگی بر اساس مدل آبشاری، فرض می‌کنیم که SP_i نشان‌دهنده احتمال تأثیرگذاری گره i بر همسایگانش با توجه به $deg(i)$، تعداد همسایگان گره i است (مشابه p_{ij} در مدل انتشار آبشاری). همچنین، ρ_i وزن تخصیص‌یافته به گره i است که می‌تواند به‌عنوان شدت تلاش برای فعال‌سازی گره تفسیر شود. بدین ترتیب مقدار مورد انتظار گسترش تأثیر در شبکه را می‌توان به صورت زیر تخمین زد.

$$\mathbb{E}[spread] = \sum_{i=1}^{n} \rho_i . SP_i . \deg(i) \quad (۱۶)$$

از این رو تابع برازندگی پیشنهادی به صورت زیر تعریف می‌شود:

$$f\ min = \sum_{i=1}^{n} W_i . Cost_i - \mathbb{E}[spread] \quad (۱۷)$$

s.t.

$$\sum_{i=1}^{n} w_i \leq 1 \quad (۱۸)$$

$$i > 0 \quad (۱۹)$$

در فرمول فوق، i نمایه کاربر در شبکه اجتماعی و K نشان دهنده تعداد کاربران تأثیرگذار است. همچنین،$\mathbb{E}[spread]$ مقدار مورد انتظار برای تعداد کاربران تحت تأثیر را نشان می‌دهد و $Cost_i$ هزینه انتخاب گره های تأثیر گذار و W_i وزن مربوط به معیارهای مرکزیت موثر در انتخاب گره های تأثیرگذار را نشان می‌دهد. وزن کاربران

تأثیرگذار در روش پیشنهادی بر اساس تخمینی از مقدار انتشار تعیین می‌شود. در روش پیشنهادی برای محاسبه مقدار انتشار از مدل آبشاری استفاده می‌شود. در این مدل در هر گام نیاز به احتمال انتشار کاربران تأثیر گذار وجود دارد. احتمال انتشار کاربران تأثیرگذار در هر گام بر اساس مقدار درجه مرکزیت به کار رفته نسبت به مقدار میانگین همان معیار مرکزیت در ازای کاربران صدق کرده در آستانه رابطه (۱۳) محاسبه شده است. از این رو می‌توان مقدار انتشار را بر اساس مقدار درجه معیار مرکزیت تخمین زد.
معادله (۱۶) بیان می‌کند که حداکثر مقدار وزن نباید از ۱ تجاوز کند. این محدودیت به این معنی است که یکی از دو مورد می‌تواند در هر اجرا اتفاق بیفتد: الف) تنها یک معیار می‌تواند در تعیین کاربران تأثیرگذار دخیل باشد، یا ب) به طور متوسط چندین مورد معیارهایی با مجموع حداکثر ۱ می‌توانند در تعیین کاربران تأثیرگذار دخیل باشند. همچنین معادله (۱۷) بیان می‌کند که هزینه راه‌حل‌ها باید حداکثر ۱ باشد. وقتی کل شبکه اجتماعی جستجو شود هزینه راه‌حل برابر با ۱ می‌شود. با توجه به آستانه اعمال شده در معادله (۱۲)، در جمعیت اولیه، تنها کاربرانی که درجه مرکزیت آنها بیشتر از میانگین است، می‌توانند راه‌حل‌های قابل اجرا باشند. این به طور شهودی نشان می‌دهد که محدودیت معادله (۱۷) همیشه معتبر است.

۵- ارزیابی کارایی

در این بخش ابتدا مدل‌های ارزیابی تأثیر اجتماعی، مورد بررسی قرار می‌گیرند. این مدل‌ها بستگی به ترکیب‌های مختلفی از عوامل دارند که هر کدام به نوعی می‌توانند تأثیرگذار باشند، مانند روابط فردی، زمان، فاصله و غیره.

۵-۱- انواع مدل‌های ارزیابی

اخیراً تلاش‌های زیادی صرف مدل‌سازی تأثیر اجتماعی در شبکه‌های اجتماعی شده است. این مدل‌ها به طور کلی در سه نوع اصلی دسته‌بندی می‌شوند که به شرح زیر هستند [۱۷]

الف) مدل‌های مبتنی بر موضوع: برخی از طرح‌ها که برای ارزیابی تأثیر اجتماعی ارائه شده‌اند، مبتنی بر موضوع هستند. در این مدل‌ها، تأثیر اجتماعی با شمارش میزان اطلاعاتی که مربوط به یک موضوع خاص می‌تواند در شبکه‌ها منتشر شود، اندازه‌گیری می‌شود.

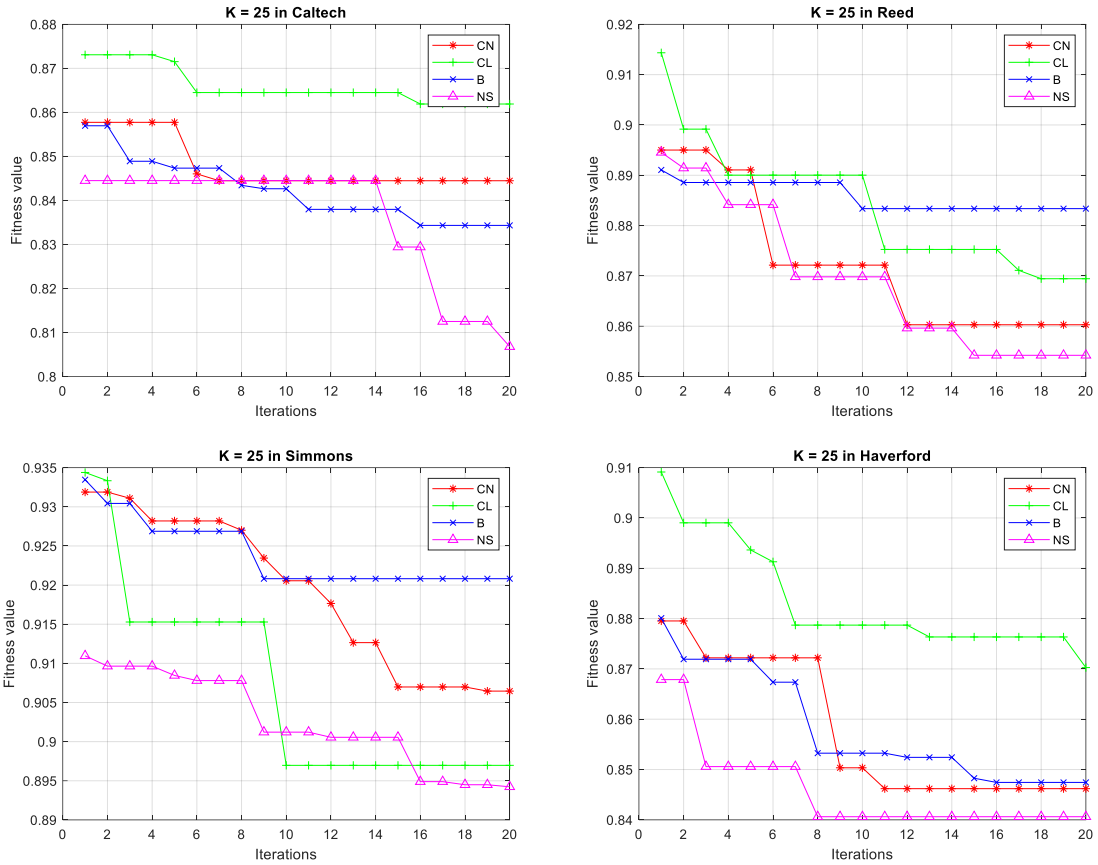
ب) مدل‌های نادیده‌گیرنده موضوع: برخی از طرح‌ها در ارزیابی تأثیر اجتماعی به موضوع توجهی ندارند. در این مدل‌ها، تأثیر اجتماعی یا از طریق قدرت نسبی افراد در شبکه‌های اجتماعی اندازه‌گیری می‌شود، یا از طریق درجه انتشار اطلاعات در شبکه‌های اجتماعی.

ج) مدل‌های جفتی: برخی از طرح‌ها بر اساس تأثیر جفتی هستند که این تأثیر بر روابط اجتماعی و تعاملات بین کاربران تعریف می‌شود[۱۷].

در روش پیشنهادی از مدل نادیده‌گیرنده موضوع برای ارزیابی استفاده می شود که در آن مقدار انتشار اطلاعات برای هر کاربر نشان دهنده میزان قدرت نسبی یا تأثیرگذاری کاربر در شبکه‌های اجتماعی است. کاربران تأثیرگذاری که در شبکه از طریق الگوریتم NSGA-III پیدا می‌شوند، بر اساس این مدل مورد ارزیابی قرار گرفته و با هم مقایسه می‌شوند.
ما از مجموعه داده Facebook100 [۱۸] استفاده می‌کنیم که شامل روابط اجتماعی ۱۰۰ موسسه آموزش عالی در ایالات متحده است. این یک مجموعه داده واقعی و بسیار کامل بدون هیچ‌گونه نمونه‌برداری قبلی است. این مجموعه داده حاوی اطلاعاتی در مورد کاربران شبکه‌های اجتماعی در کالج‌های مختلف در دانشگاه استنفورد است. در این مجموعه داده، یک گراف مجاورت برای هر دانشکده ایجاد شده است که ارتباط بین کاربران را نشان می‌دهد. حداکثر ۴۱۰۰۰ راس و حداکثر ۱/۶ میلیون لبه دارد. ما ابتدا عملکرد هر متریک را در چهار مجموعه داده شامل Caltech-256، Reed98، Simmons81 و Haverford76 ارزیابی می‌کنیم. سپس، تأثیر جایگشت‌های مختلف معیارها را در سناریوهای مختلف با اندازه‌های بذر محدود برای یافتن کاربران تأثیرگذار ارزیابی می‌کنیم.

۵-۲- یافتن بهترین جایگشت معیارها

در سناریوی اول، تعداد کاربران تأثیرگذار، یعنی K برابر با ۲۵ در نظر گرفته شده است. در این سناریو، جمعیت اولیه الگوریتم NSGA-III شامل ۱۰۰ کروموزوم است که به صورت آرایه‌ای با ۲۵ عنصر تنظیم شده است. مقدار هر عنصر نشان دهنده یک کاربر در شبکه اجتماعی است. کاربران در فضای جستجو بر اساس آستانه مرکزیت معادله (۱۲) انتخاب می‌شوند. به عبارت دیگر، کاربرانی هستند که تعداد ارتباط آنها از میانگین تعداد کل ارتباطات کاربران در شبکه اجتماعی بیشتر است. جمعیت اولیه فقط از کاربرانی انتخاب می‌شود که شرایط آستانه فوق را دارند.



شکل ۳: مقادیر برازندگی با تکرارهای مختلف برای معیارهای فردی (سناریو با K=25)

ارزیابی جمعیت اولیه بر اساس تابع برازندگی معادله (۱۵) انجام می‌شود. از آنجایی که هدف ما در معادله (۱۵) به حداقل رساندن است، مقادیر کوچکتر در خصوص برازندگی، مطلوب ما هستند. کروموزوم‌های اولیه بر اساس ارزش برازندگی به ترتیب صعودی مرتب می‌شوند. فرض می‌کنیم که ضریب تولید جمعیت جدید ۰/۵ باشد. بنابراین نیمی از جمعیت اولیه که ارزش برازندگی کمتری دارند برای ایجاد جمعیت نسل بعدی انتخاب می‌شوند و بقیه حذف می شوند.

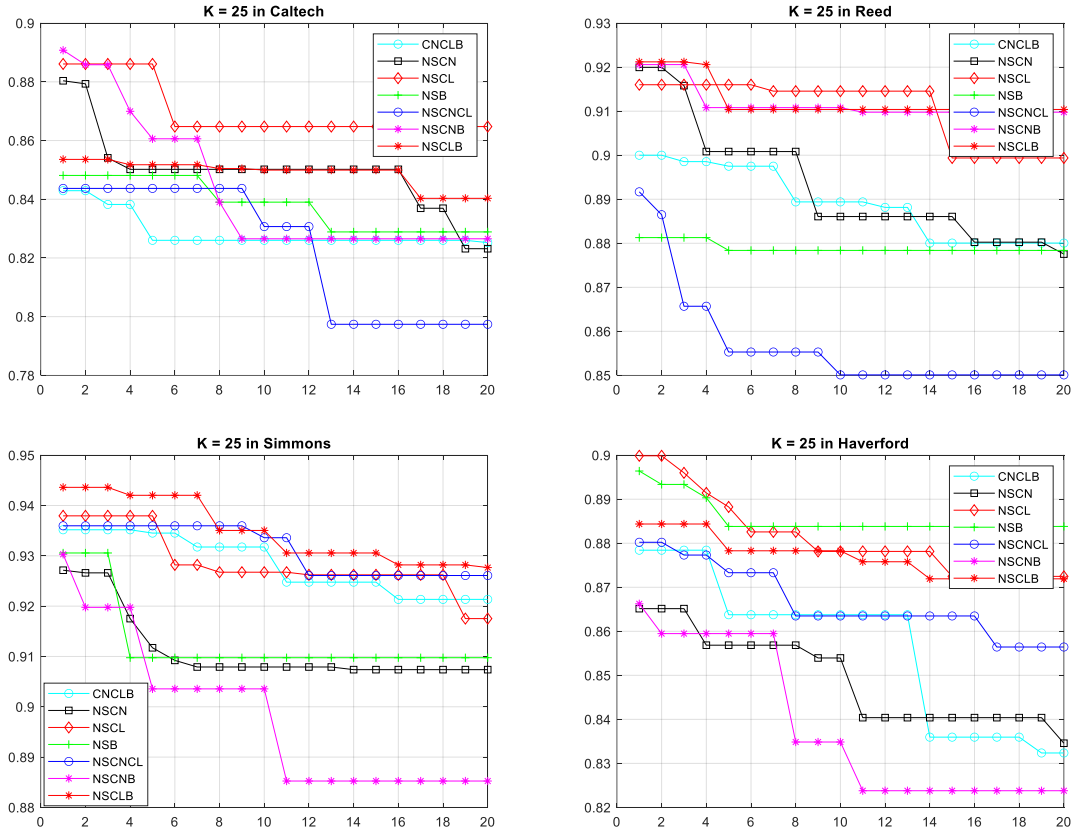
اکنون، کروموزوم‌های انتخاب شده با یکدیگر تزویج می‌شوند تا جمعیت نسل بعدی ایجاد شود. سپس عملیات جهش بر روی تعدادی از فرزندان تولید شده انجام می‌شود. در طی این عملیات، برخی از ژن‌ها ممکن است در فضای جستجوی اولیه وجود نداشته باشند. این بر اساس آستانه مرکزیت تعیین می‌شود. به عبارت دیگر، درجه مرکزیت کاربران در کروموزوم‌های جدید تولید شده ممکن است کمتر از آستانه باشد. برای ژن‌هایی با چنین مقادیر غیرقابل قبولی، یک مقدار تصادفی جدید با مقدار قبلی جایگزین می‌شود. این روند تا نیمی از کروموزوم‌های جمعیت تکرار می‌شود. سپس جمعیت جدید به اضافه جمعیت انتخاب شده قبلی، جمعیت بالقوه نسل بعدی را تشکیل می‌دهد. برای تولید نسل بعدی، نیمی از جمعیت فعلی مجدداً انتخاب شده و نیمی دیگر با توجه به ارزش برازندگی حذف می‌شوند. از بین جمعیت انتخاب شده در هر مرحله، کروموزوم با کمترین ارزش برازندگی به عنوان کروموزوم غیرغالب انتخاب شده و به لیست راه‌حل‌های بهینه اضافه می‌شود. این روند تا رسیدن به حالت توقف تکرار می‌شود. در هر مرحله، لیست راه‌حل‌های ترجیحی به روز می‌شود و راه حل بهینه سراسری در بالای لیست قرار می‌گیرد. ما حداکثر تعداد تکرارها را ۲۰ در نظر می‌گیریم. شکل ۳ مقادیر برازندگی را برای تکرارهای مختلف برای معیارهای فردی در سناریویی با K=25 نشان می‌دهد. از آنجایی که می‌خواهیم تابع هدف را در رابطه (۱۵) کمینه کنیم، مقادیر برازش کوچکتر، مطلوب ما هستند. همانطور که در شکل ۳ مشاهده می‌شود، معیار شباهت همسایگی، که ما برای اولین بار در این مقاله ابداع کردیم، تناسب بهتری را برای همه مجموعه داده‌ها در مقایسه با سه معیار دیگر ایجاد می‌کند.

شکل ۴ مقادیر برازندگی را با تکرارهای مختلف برای جایگشت‌های مختلف معیارها در سناریوی K=25 نشان می‌دهد. جایگشت‌های محاسبه شده عبارتند از CNCLB، NSCN، NSCL، NSB، NSCNCL، NSCNB و NSCLB. برای مثال، CNCLB ترکیبی از سه معیار را نشان می‌دهد: نزدیکی، مرکزیت، و بینابین بودن.

همانطور که در شکل ۴ نشان داده شده است، در مجموعه داده Caltech، معیار NSCNCL به سریع‌ترین همگرایی و کمترین ارزش برازندگی در مقایسه با سایر معیارهای ترکیبی دست می‌یابد. همچنین در مجموعه داده Reed، معیار NSLB سریعترین همگرایی و کمترین ارزش برازندگی را ایجاد می‌کند. در مجموعه داده Simmons، معیار NSCNCL به کمترین مقدار برازندگی در مقایسه با سایر معیارهای ترکیبی دست می‌یابد. در نهایت، در مجموعه داده Haverford، معیار NSCNB سریع‌ترین همگرایی و کمترین ارزش برازندگی را دارد. در همین حال، CNCLB از عملکرد ضعیف‌تری نسبت به معیارهای ترکیبی که در آنها معیار شباهت همسایگی (NS) وجود دارد برخوردار است.

جدول ۲ میانگین مقادیر برازندگی را برای معیارهای مختلف در همه سناریوها نشان می‌دهد. همانطور که در جدول مشاهده می‌شود، مقادیر برازندگی به دست آمده از معیار شباهت همسایگی کمتر از سایر معیارها است. همچنین، تمام سناریوهای ترکیبی که شامل متریک شباهت همسایگی می‌شوند، به مقادیر برازندگی بهتری منجر می‌شوند. بهترین ترکیب معیارها NSCNCL و NSCNB و CNCLB است. به عبارت دیگر، ترکیب NS با هر دو معیار بینابین بودن و نزدیکی بهترین نتایج را می‌دهد. نکته جالب این است که ترکیب سه معیار مرکزیت، نزدیکی و بینابین بودن (CNCLB) نتایج ضعیف‌تری نسبت به ترکیبات NS دارد. این نشان می‌دهد که معیار NS که در این مقاله ایجاد شده است دقیق و قوی است.

مقدار برازنگی حاصل از یک تابع کمینه‌سازی است که از ترکیب سه هدف کاهش تعداد کاربران، تأثیرگذار، کاهش هزینه محاسباتی و افزایش تاثیر در شبکه‌های اجتماعی، تشکیل یافته است. روشی که کمترین مقدار برازندگی را داشته باشد درواقع کمترین تعداد کاربران، کمترین هزینه محاسباتی و بیشترین تأثیر در شبکه‌های اجتماعی را دارد. با توجه به این که تعداد کاربران تأثیر گذار به صورت محدود در نظر گرفته شده است و هزینه محاسباتی با توجه به آستانه مرکزیت در هر روش کاهش می‌یابد، در واقع کمترین مقدار برازندگی نشان دهنده راه حلی با بیشترین مقدار نفوذ در شبکه است. به منظور اثبات این ادعا می‌توان در شکل ۴ مشاهده کرد که روش‌هایی که کمترین مقدار برازندگی را دارند نسبت سایر روش‌ها تأثیر بیشتری در شبکه اجتماعی به دست آورده‌اند.



شکل ۴: مقادیر برازندگی با تکرارهای مختلف برای معیارهای ترکیبی (سناریو با K=25)

جدول ۲: مقادیر متوسط برازندگی برای معیارهای مختلف در همه سناریوها.

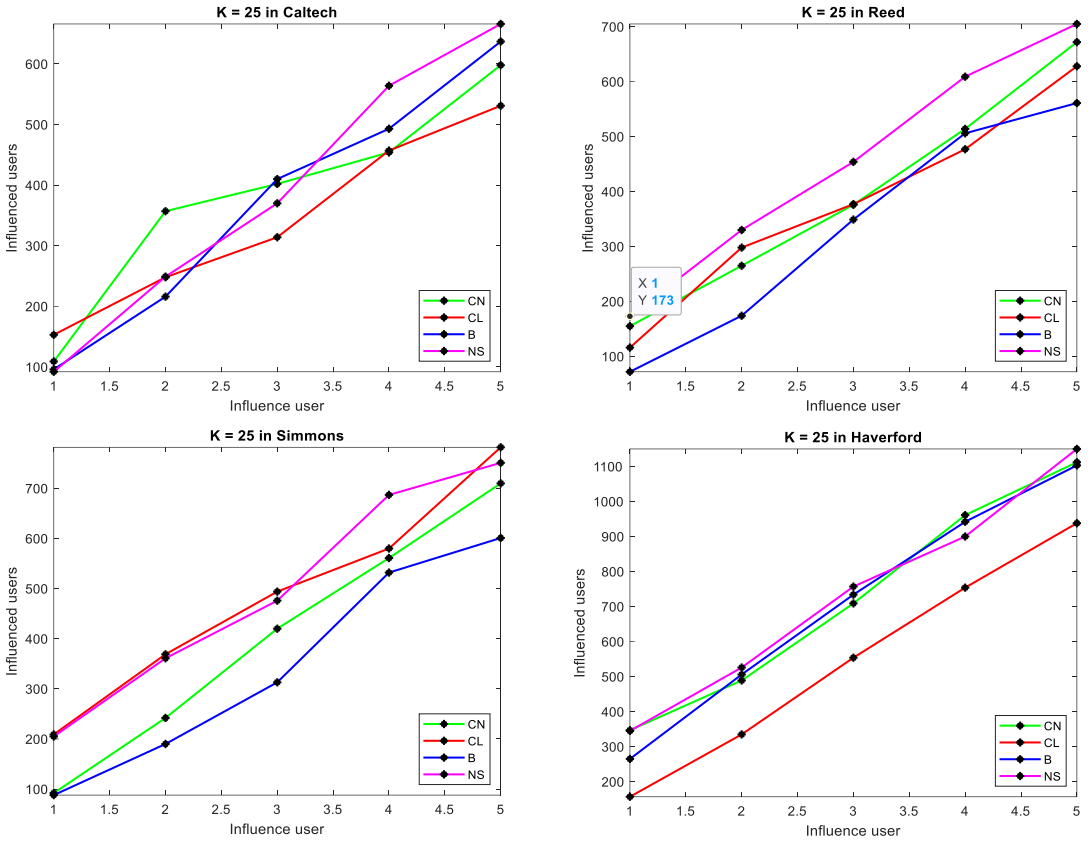
Seed size	dataset	CN	CL	B	NS	CN-CL-B	NS-CN	NS-CL	NS-B	NS-CN-CL	NS-CN-B	NS-CL-B
10	Caltech	0.728	0.736	0.757	**0.72**	0.744	**0.719**	0.727	0.74	0.759	0.759	0.722
	Reed	0.819	0.818	0.810	**0.794**	0.777	0.791	0.816	0.810	0.818	0.803	**0.771**
	Simmons	0.844	0.848	0.837	**0.826**	0.840	0.836	0.841	0.837	0.841	0.838	0.84
	Haverford	0.792	0.774	0.774	**0.751**	0.762	0.769	0.788	0.785	0.774	**0.759**	0.788
15	Caltech	0.754	0.739	0.742	**0.729**	0.761	**0.734**	0.748	0.757	0.777	**0.734**	0.748
	Reed	0.789	0.802	0.799	**0.775**	0.814	0.797	0.79	0.79	0.79	0.769	**0.765**
	Simmons	0.881	0.847	0.839	**0.835**	0.849	**0.832**	0.858	0.841	0.841	**0.832**	0.835
	Haverford	0.763	0.788	0.772	**0.76**	0.781	**0.779**	0.787	0.789	0.791	0.787	0.789
20	Caltech	0.788	0.788	0.768	**0.755**	0.798	0.793	0.758	**0.75**	**0.75**	0.758	0.769
	Reed	0.794	0.794	0.79	**0.785**	0.793	0.794	0.793	**0.79**	0.8	0.798	0.795
	Simmons	0.874	0.862	0.857	**0.85**	0.858	0.861	0.869	0.875	0.873	0.861	**0.851**
	Haverford	0.797	0.789	0.791	**0.785**	0.789	0.792	0.794	0.794	**0.784**	0.79	**0.784**
25	Caltech	0.8479	0.8659	0.8422	**0. 8363**	0.8289	0.8493	0.8701	0.8381	**0.8232**	0.8436	0.8490
	Reed	0.8721	0.8837	0.8858	**0.8692**	0.8892	0.8930	0.9104	0.8790	**0.8569**	0.9118	0.9125
	Simmons	0.9192	0.9071	0.9239	**0.9025**	9284	0.9113	0.9286	0.9129	0.9313	**0.8982**	9394
	Haverford	0.8577	0.8834	0.8579	**0.8358**	0.8566	0.8493	0.8813	**0.8857**	0.8666	0.8383	0.8769

۵-۳- اثر اندازه بذر بر کاربران تاثیرگذار

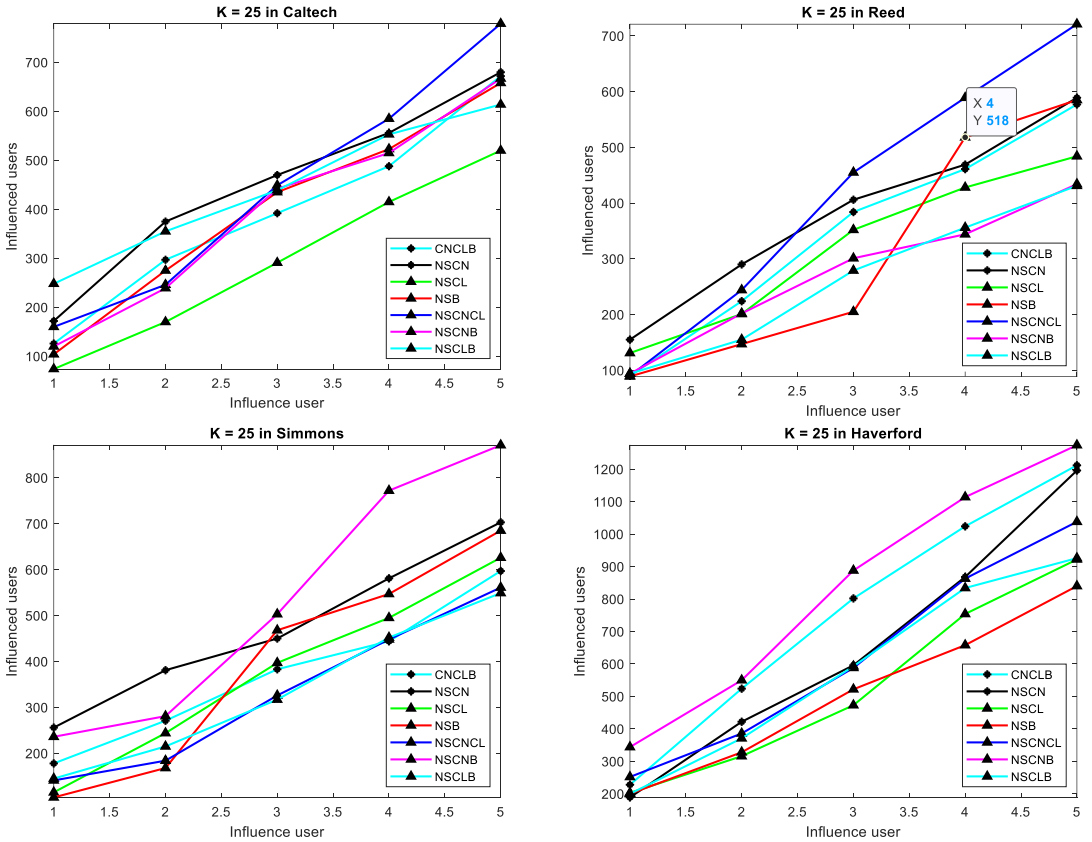
پس از اجرای روش پیشنهادی بر اساس سناریوهای مختلف، در این بخش به ارزیابی عملکرد روش پیشنهادی می‌پردازیم. یکی از شناخته‌شده‌ترین معیارهای ارزیابی عملکرد در مسئله حداکثرسازی تاثیر، تعداد کاربرانی است که تحت تأثیر کاربران تأثیرگذار اولیه قرار می‌گیرند. به عبارت دیگر، در این مساله، هدف یافتن کاربرانی است که بیشترین تأثیر را بر سایر کاربران دارند. بنابراین، هر راه حلی که بتواند تأثیرگذارترین کاربران را بیابد، می‌تواند راه حل بهینه تلقی شود. ما روش پیشنهادی را با NSGA-III بر اساس چهار معیار شرح داده شده در بالا پیاده‌سازی می‌کنیم. بر این اساس، ما چهار سناریوی مختلف شامل K=3، ۴، ۵، و ۶ کاربر تاثیرگذار را با معیارهای واحد و ترکیبی پیاده‌سازی می‌کنیم. شکل ۵ تعداد کاربرانی را که توسط کاربران تاثیرگذار اولیه در سناریویی با K=3 با استفاده از معیارهای منفرد تحت تاثیر قرار گرفته اند نشان می‌دهد. همانطور که در شکل مشاهده می‌شود، متریک شباهت همسایگی توانسته کاربران بیشتری را تحت تاثیر قرار دهد. به عبارت دیگر، کاربران تأثیرگذار که بر اساس معیارهای شباهت همسایگی یافت می‌شوند، تأثیر بیشتری بر سایر کاربران شبکه اجتماعی دارند.

شکل ۶ تعداد کاربرانی را نشان می‌دهد که تحت تأثیر کاربران تأثیرگذار اولیه در یک سناریوی K=3 با استفاده از معیارهای ترکیبی قرار گرفته‌اند. نگاهی دقیق‌تر به شکل ۶ نشان می‌دهد که معیارهای ترکیبی، هنگامی که به عنوان وزن تابع برازندگی در NSGA-III استفاده می‌شوند، می‌تواند بر بیشترین تعداد کاربران در شبکه تأثیر بگذارد. شکل نشان می‌دهد که در همه مجموعه داده‌ها، استفاده از معیارهای ترکیبی NSB، NSCLB، NSCN و NSCNB به ترتیب منجر به کسب بیشترین تعداد کاربران تحت تأثیر می‌شود.

جدول ۳ تعداد کاربران تحت تأثیر کروموزوم غیرغالب در NSGA-III را برای همه سناریوها نشان می‌دهد. همانطور که در جدول نشان داده شده است، تعداد کاربرانی که تحت تاثیر اینفلوئنسرها با استفاده از معیار تشابه همسایگی قرار گرفته‌اند، بیشتر از سایرین است. همچنین، سناریوهایی که شباهت همسایگی در ترکیب با معیارهای دیگر استفاده می‌شود، می‌تواند کاربران تحت تأثیر بیشتری را شناسایی کند.



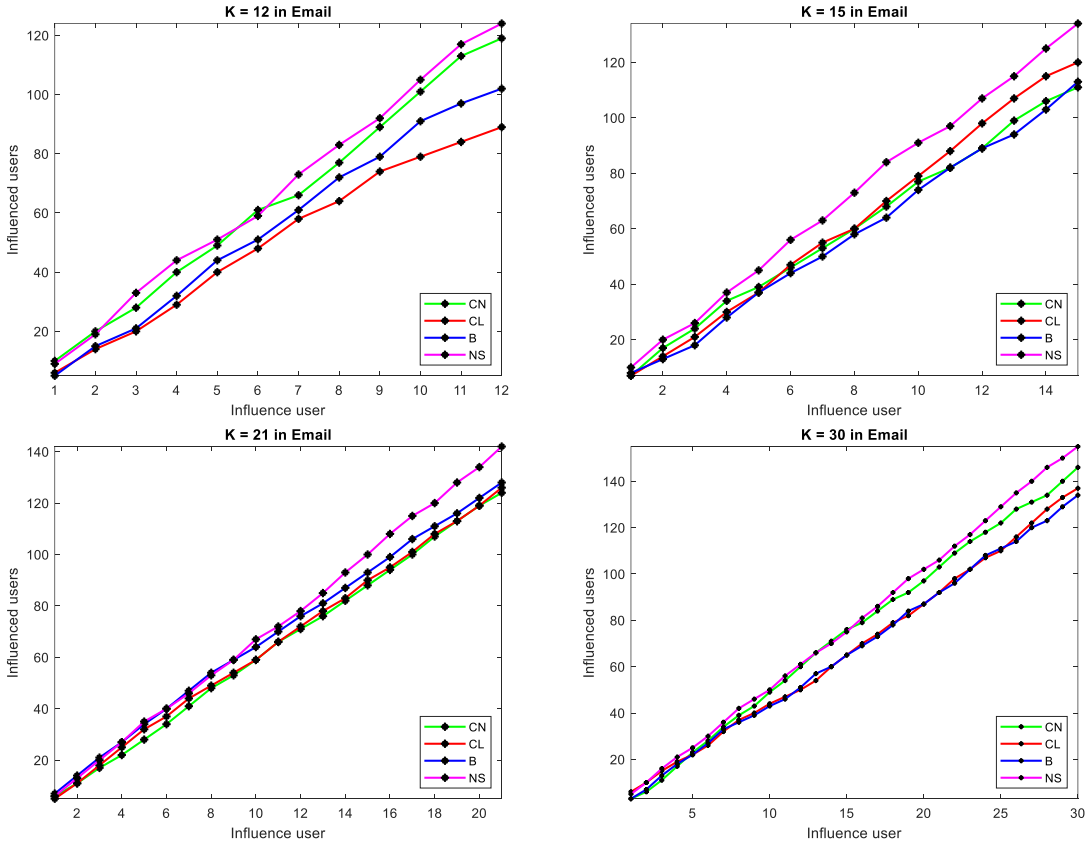
شکل ۵: تعداد کاربرانی که توسط کاربران تاثیرگذار اولیه در سناریویی با K=25 تحت تاثیر قرار می‌گیرند (با معیارهای ساده).



شکل ۶: تعداد کاربرانی که توسط کاربران تاثیرگذار اولیه در سناریویی با K=25 تحت تاثیر قرار می‌گیرند (با معیارهای ترکیبی).

۵-۴- مقایسه روش پیشنهادی

در این بخش به منظور اعتبار سنجی روش پیشنهادی، به مقایسه روش پیشنهادی با روش‌های پیشین خواهیم پرداخت. مقایسه روش پیشنهادی به منظور بررسی بهبود ایجاد شده بر اساس معیار شباهت همسایگی و کاربرد آن در الگوریتم NSGA-III در راستای افزایش تأثیر در شبکه‌های اجتماعی خواهد بود. ما روش پیشنهادی را بر روی مجموعه داده Email پیاده سازی می‌کنیم تا نتایج آن را با نتایج منتشر شده در تحقیقات پیشین شامل MTEFIM [۱۹]، MTEFIM-NK [۱۹]، MFEA [۲۰]، MFEAII [۲۱]، MFEAIM [۲۱] SBGA [۲۲]، EMEA [۲۳] و Pagerank [۲۴]، در شرایط یکسان مقایسه نماییم. تحقیقات فوق، به دلیل استفاده از معیارهای سنتی برای افزایش تأثیر در شبکه‌های اجتماعی، انتخاب شده‌اند. این تحقیقات بر روی مجموعه داده مشخصی با تعداد بذرهای معین پیاده‌سازی شده‌اند و نتایج مربوط به میزان انتشار هر روش با تنظیمات مشخص، منتشر شده است. از این رو روش پیشنهادی بر روی مجموعه داده مشخص و با تعداد بذرهای معین پیاده سازی شد تا نتایج در شرایط یکسان قابل مقایسه باشد. در شکل ۷ تعداد کاربران تحت تأثیر در روش پیشنهادی در مجموعه داده Email مختلف نشان داده شده است.

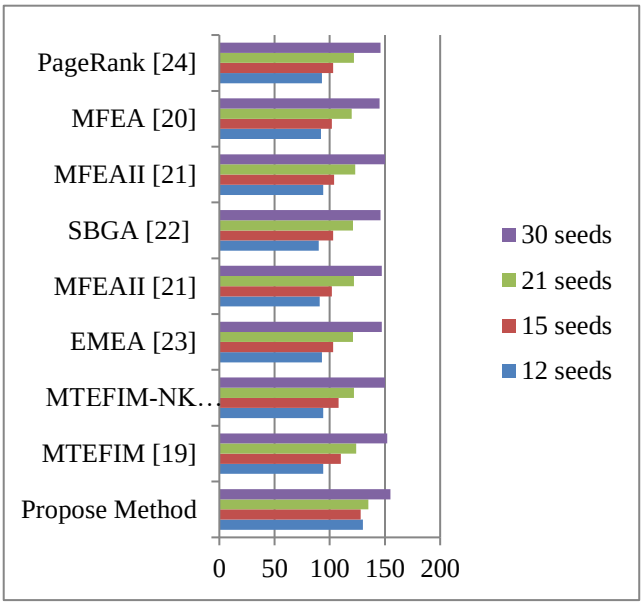


شکل ۷ تعداد کاربران تحت تأثیر در روش پیشنهادی در مجموعه داده Email

جدول ۳: تعداد کاربران تحت تأثیر ناشی از کروموزوم غیرغالب در NSGA-III برای همه سناریوها.

Seed size(K)	dataset	CN	CL	B	NS	CN-CL-B	NS-CN	NS-CL	NS-B	NS-CN-CL	NS-CN-B	NS-CL-B
10	Caltech	614	654	592	**667**	572	636	**687**	618	587	632	280
	Reed	511	594	607	**642**	548	552	615	647	587	601	**653**
	Simmons	697	715	768	**798**	728	**822**	719	755	812	**822**	761
	Haverford	990	1084	1008	**1102**	997	957	1021	1044	1011	**1104**	978
15	Caltech	764	756	806	**829**	758	**853**	801	790	706	751	822
	Reed	854	817	893	**905**	721	800	746	791	847	806	**890**
	Simmons	1067	1027	978	**1094**	900	**1139**	1024	924	1057	**1139**	932
	Haverford	1287	1347	1403	**1429**	1284	**1343**	1291	1179	1200	1297	1275
20	Caltech	904	871	972	**1022**	914	924	917	**982**	**982**	912	953
	Reed	924	849	997	**1002**	1027	987	1098	**1121**	1098	1067	974
	Simmons	957	1074	1060	**1107**	1106	1198	1201	1194	1150	1267	**1274**
	Haverford	1420	1533	1490	**1603**	1544	1597	1607	1607	**1652**	1597	1637
25	Caltech	598	531	637	**666**	672	680	520	658	**779**	667	614
	Reed	672	628	561	**700**	577	589	484	585	**721**	434	431
	Simmons	710	782	601	**791**	597	703	626	685	561	**871**	549
	Haverford	1112	938	1103	**1150**	1212	1196	922	840	1038	**1274**	962

در شکل ۸ مقایسه روش پیشنهادی با روش‌های پیشین از نظر مقدار انتشار در مجموعه داده Email بر اساس تعداد کاربران تأثیرگذار معین، نشان داده شده است.



شکل ۸. مقایسه روش پیشنهادی و روش‌های پیشین در مجموعه داده Email

همانطور که در شکل ۷ و ۸ قابل مشاهده است، روش پیشنهادی بر روی مجموعه داده Email نسبت به روش‌های پیشین نتایج بهتری در راستای انتشار اطلاعات در شبکه اجتماعی به دست آورده است. این نتایج نشان می دهد، وجود تعداد زیادی از همسایگانی با درجه مشابه و بالا می تواند موجب افزایش قدرت و نفوذ در گاربران تأثیرگذار شود. از این رو روش پیشنهادی با استفاده از معیار شباهت همیاسگی و با بهره گیری از الگوریتم NSGA-III توانسته است نتایج بهتری نسبت به سایر معیارهای اهمیت اجتماعی مانند مرکزیت، نزدیکی، بینابینی و PageRank [۱۹] و روش‌های پیشین در انتشار اطلاعات درشبکه کسب کند.

۶- نتیجه‌گیری

این مقاله به مساله بیشینه‌سازی نفوذ کاربران در شبکه‌های اجتماعی پرداخت. برای ارزیابی اهمیت کاربران در شبکه‌های اجتماعی، معیار جدیدی به نام شباهت همسایگی ارائه گردید. سپس با توجه به سختی NP، مساله را با استفاده از NSGA-III حل کردیم. هدف ما این است که به طور مشترک هزینه را به حداقل برسانیم و تعداد کاربران تحت تأثیر را به حداکثر برسانیم. ایده اصلی در معیار پیشنهادی انتخاب کاربرانی است که همسایگان مشابه بیشتری دارند. نتایج آزمایش‌ها نشان داد که روش پیشنهادی با استفاده از معیار شباهت همسایگی، قادر به شناسایی کاربران تأثیرگذارتر با هزینه کمتر است. استفاده از معیار پیشنهادی می‌تواند منجر به بهبود تعداد کاربران تحت تأثیر به ترتیب ۹/۵، ۱۶/۵ و ۶/۵ درصد در مقایسه با معیارهای مرکزیت، نزدیکی و بینابین بودن شود. ما کارایی روش پیشنهادی را در حضور معیار جدید در برابر جایگشت‌های مختلف معیارهای نزدیکی، بینابین بودن و مرکزیت ارزیابی کردیم. نتایج آزمایش‌ها نشان داد که جایگشت‌هایی که شامل معیار جدید هستند، عملکرد بهتری نسبت به سایرین دارند. همچنین متوجه شدیم کاربرانی که بیشترین تعداد کاربران مشابه را در همسایگی خود دارند باعث کارایی بیشتر در شبکه اجتماعی می‌شوند.

یافته‌های ما در این تحقیق می‌تواند کاربردهای مهمی در تشخیص

کلونی‌ها در شبکه‌های اجتماعی داشته باشد. استفاده از کاربران تاثیرگذار به عنوان یک مرکز و تشکیل جوامع با کاربران مشابه در یک سطح، یک دستاورد مهم است. برخی از کاربردها عبارتند از بازاریابی الکترونیکی، جنگ شناختی، آموزش و غیره.

مراجع

[1] S. Yang, Q. Du, G. Zhu, J. Cao, L. Chen, W. Qin, and Y. Wang, “Balanced influence maximization in social networks based on deep reinforcement learning,” *Neural Netw.*, vol. 169, pp. 334–351, Jan. 2024, doi: 10.1016/j.neunet.2023.10.030.

[2] A. Zareie and R. Sakellariou, “Fuzzy influence maximization in social networks,” *ACM Trans. Web*, vol. 18, no. 3, Art. no. 32:1–32:28, Aug. 2024, doi: 10.1145/3650179.

[3] R. Cantini, F. Marozzo, S. Mazza, D. Talia, and P. Trunfio, “A weighted artificial bee colony algorithm for influence maximization,” *Online Soc. Netw. Media*, vol. 26, Art. no. 100167, Sep. 2021, doi: 10.1016/j.osnem.2021.100167.

[4] T. K. Biswas, A. Abbasi, and R. K. Chakrabortty, “An MCDM integrated adaptive simulated annealing approach for influence maximization in social networks,” *Inf. Sci.*, vol. 556, pp. 27–48, May 2021, doi: 10.1016/j.ins.2020.12.048.

[5] Z. Aghaee, M. M. Ghasemi, H. A. Beni, A. Bouyer, and A. Fatemi, “A survey on meta-heuristic algorithms for the influence maximization problem in the social networks,” *Computing*, vol. 103, no. 11, pp. 2437–2477, Sep. 2021, doi: 10.1007/s00607-021-00945-7.

[6] A. Karczmarczyk, J. Jankowski, and J. Wątrobski, “Multi-criteria seed selection for targeted influence maximization within social networks,” in *Proc. Int. Conf. Comput. Sci. (ICCS)*, Cham, Switzerland: Springer, Jul. 2021, pp. 454–461, doi: 10.1007/978-3-030-69101-1_36.

[7] R. R. Singh, “Centrality measures: a tool to identify key actors in social networks,” in *Principles of Social Networking: The New Horizon and Emerging Challenges*, Cham, Switzerland: Springer, 2022, pp. 1–27, doi: 10.1007/978-3-030-91949-8_1.

[8] Y. D. Navaei, M. H. Rezvani, and A. M. E. Moghaddam, “A novel neighborhood-based importance measure for social network influence maximization using NSGA-III,” in *Proc. 10th Int. Conf. Artif. Intell. Robot. (QICAR)*, Tehran, Iran, Feb. 2024, pp. 113–118, doi: 10.1109/QICAR57417.2024.1234567.

[9] H. Ishibuchi, R. Imada, Y. Setoguchi, and Y. Nojima, "Performance comparison of NSGA-II and NSGA-III on various many-objective test problems," in *Proc. IEEE Congr. Evol. Comput. (CEC)*, Vancouver, BC, Canada, Jul. 2016, pp. 3045–3052, doi: 10.1109/CEC.2016.7744174.

[10] J. J. Lotf, M. A. Abdollahi Azgomi, and M. R. Ebrahimi Dishabi, "An improved influence maximization method for social networks based on genetic algorithm," *Physica A*, vol. 586, Art. no. 126480, Jan. 2022, doi: 10.1016/j.physa.2021.126480.

[11] E. A. Abbas and H. N. Nawaf, "Influence maximization based on a non-dominated sorting genetic algorithm," *Karbala Int. J. Mod. Sci.*, vol. 7, no. 2, Art. no. 5, Jun. 2021, doi: 10.33640/2405-609X.2891.

[12] T. K. Biswas, A. Abbasi, and R. K. Chakrabortty, "An improved clustering-based multi-objective evolutionary algorithm for influence maximization under variable-length solutions," *Knowl.-Based Syst.*, vol. 256, Art. no. 109856, Sep. 2022, doi: 10.1016/j.knosys.2022.109856.

[13] B. Chatterjee, T. Bhattacharyya, K. K. Ghosh, A. Chatterjee, and R. Sarkar, "A novel meta-heuristic approach for influence maximization in social networks," *Expert Systems*, vol. 40, no. 4, p. e12676, 2023, doi: 10.1111/exsy.12676.

[14] H. Li, R. Zhang, Z. Zhao, X. Liu, and Y. Yuan, "Identification of top-k influential nodes based on discrete crow search algorithm optimization for influence maximization," *Applied Intelligence*, vol. 51, no. 11, pp. 7749–7765, 2021, doi: 10.1007/s10489-021-02283-9.

[15] L. Han, K.-C. Li, A. Castiglione, J. Tang, H. Huang, and Q. Zhou, "A clique-based discrete bat algorithm for influence maximization in identifying top-k influential nodes of social networks," *Soft Computing*, vol. 25, pp. 8223–8240, 2021, doi: 10.1007/s00500-021-05749-7.

[16] D. Borkin, A. Némethová, G. Michaľčonok, and K. Maiorov, "Impact of data normalization on classification model accuracy," *Research Papers Faculty of Materials Science and Technology Slovak University of Technology*, vol. 27, no. 45, pp. 79–84, 2019, doi: 10.2478/rput-2019-0029.

[17] 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, Jan. 2018, doi: 10.1016/j.jnca.2018.01.005.

[18] A. L. Traud, E. D. Kelsic, P. J. Mucha, and M. A. Porter, "Comparing community structure to characteristics in online collegiate social networks," *SIAM Review*, vol. 53, no. 3, pp. 526–543, Sep. 2011, doi: 10.1137/080734315.

[19] C. Wang, J. Zhao, L. Li, L. Jiao, J. Liu, and K. Wu, "A multi-transformation evolutionary framework for influence maximization in social networks," *IEEE Computational Intelligence Magazine*, vol. 18, no. 1, pp. 52–67, Jan. 2023, doi: 10.1109/MCI.2022.3222050.

[20] A. Gupta, Y.-S. Ong, and L. Feng, "Multifactorial evolution: Toward evolutionary multitasking," *IEEE Transactions on Evolutionary Computation*, vol. 20, no. 3, pp. 343–357, Jun. 2016, doi: 10.1109/TEVC.2015.2458037.

[21] K. K. Bali, Y.-S. Ong, A. Gupta, and P. S. Tan, "Multifactorial evolutionary algorithm with online transfer parameter estimation: MFEA-II," *IEEE Trans. Evol. Comput.*, vol. 24, no. 1, pp. 69–83, Feb. 2020, doi: 10.1109/TEVC.2019.2906927.

[22] R.-T. Liaw and C.-K. Ting, "Evolutionary manytasking optimization based on symbiosis in biocoenosis," in *Proc. AAAI Conf. Artif. Intell.*, vol. 33, no. 1, pp. 4295–4303, Jan. 2019, doi: 10.1609/aaai.v33i01.33014295.

[23] L. Feng, L. Zhou, J. Zhong, A. Gupta, Y.-S. Ong, K. C. Tan, and A. K. Qin, "Evolutionary multitasking via explicit autoencoding," *IEEE Trans. Cybern.*, vol. 49, no. 9, pp. 3457–3470, Sep. 2019, doi: 10.1109/TCYB.2019.2902865.

[24] L. Page, S. Brin, R. Motwani, and T. Winograd, "The PageRank citation ranking: Bringing order to the web," *Stanford InfoLab*, Tech. Rep. SIDL-WP-1999-0120, Jan. 1998. [Online]. Available: https://ilpubs.stanford.edu/422/

پاورقی‌ها:

[1] Influence Maximization
[2] Scalable
[3] Metaheuristic
[4] Exploration
[5] Exploitation
[6] Non-dominated Sorting Genetic Algorithm
[7] Neighborhood Similarity
[8] Centrality
[9] Closeness
[10] Betweenness
[11] Permutations
[12] Seed
[13] Multi-criteria Decision-making
[14] Discrete Bat Algorithm (DBA)
[15] Seed Users