

دریافت مقاله: ۱۴۰۳/۰۵/۰۹

پذیرش مقاله: ۱۴۰۳/۰۶/۲۴

نوع مقاله: پژوهشی

بهبود حداکثرسازی گسترش انتشار اطلاعات در شبکه‌های اجتماعی به کمک معیارهای شباهت کاربران

زهره سجده

گروه مهندسی کامپیوتر، دانشکده فنی و مهندسی، دانشگاه علم و فرهنگ، تهران، ایران
پست الکترونیکی: sajdehzahra@gmail.com

علیرضا رضوانیان*

استادیار، گروه مهندسی کامپیوتر، دانشکده فنی و مهندسی، دانشگاه علم و فرهنگ، تهران، ایران
پست الکترونیکی: rezvanian@usc.ac.ir

چکیده

وزن‌دهی شده و سپس معیارهای انتخاب بر روی گراف وزن‌دار اعمال می‌شود تا تاثیرگذارترین گره‌ها شناسایی شوند. کارایی الگوریتم پیشنهادی از طریق شبیه‌سازی توسط چندین آزمایش بر روی چند مجموعه داده استاندارد شبکه‌های اجتماعی ارزیابی شده و نتایج شبیه‌سازی‌ها نشان‌دهنده عملکرد مطلوب الگوریتم است؛ به طوری که از لحاظ شاخص نرخ پوشش، الگوریتم پیشنهادی براساس معیار شباهت جاکارد، نسبت به بهترین نتیجه رقیب در هر سناریو براساس نتایج ۶ مجموعه داده مختلف به طور متوسط ۶/۸۷ درصد بهبود داشته است. واژه‌های کلیدی: شبکه‌های اجتماعی، حداکثرسازی انتشار اطلاعات، معیارهای شباهت.

۱. مقدمه

در سال‌های اخیر، استفاده از شبکه‌های اجتماعی برخط به طور گسترده‌ای در میان افراد جامعه رواج یافته است. به طور روزمره و لحظه‌ای صدها میلیون

در سال‌های اخیر، استفاده از شبکه‌های اجتماعی به طور گسترده‌ای افزایش یافته و این شبکه‌ها به بستری محبوب برای تبلیغ محصولات و انتشار اطلاعات تبدیل شده‌اند. مسئله حداکثرسازی گسترش انتشار اطلاعات، که یکی از مسائل الگوریتمی کلیدی در زمینه انتشار اطلاعات است، به موضوعی داغ در تحلیل شبکه‌های اجتماعی مبدل شده است. حداکثرسازی گسترش انتشار اطلاعات در یک شبکه اجتماعی به معنای شناسایی مجموعه‌ای از چند گره محدود در یک شبکه اجتماعی است به طوری که با فعال‌سازی این مجموعه محدود، گره‌ها بتوانند براساس یک مدل انتشار مشخص، بیشترین میزان انتشار اطلاعات را ممکن سازند. در این مقاله، یک الگوریتم جدید برای یافتن گره‌های تاثیرگذار با هدف حداکثرسازی گسترش انتشار اطلاعات ارائه شده که از معیارهای شباهت کاربران در شبکه‌های اجتماعی بهره می‌برد. در الگوریتم پیشنهادی، پس از محاسبه شباهت میان گره‌های شبکه، یال‌ها

* نویسنده مسئول

ویژه‌ای برخوردار است. این موضوع، پژوهشگران را به سمت ارائه الگوریتم‌های نوآورانه متنوعی در این حوزه سوق داده است [۲].

حداکثرسازی گسترش انتشار اطلاعات، یک مسئله الگوریتمی کلیدی در زمینه انتشار اطلاعات و تحلیل تأثیرات اجتماعی است که برای اولین بار توسط کمپه و همکارانش در سال ۲۰۰۳ معرفی و مدل‌سازی شد. حداکثرسازی انتشار در یک شبکه اجتماعی به معنای یافتن مجموعه محدود چند عضوی از گره‌های اولیه انتشاردهنده اطلاعات است که بتوانند براساس یک مدل انتشار مشخص، بیشترین میزان انتشار را در پایان فرآیند انتشار در سطح شبکه ایجاد کنند. چند گره محدود انتخابی در واقع همان تعداد گره‌های اولیه تأثیرگذار هستند که انتظار می‌رود بتوانند انتشار اطلاعات را به حداکثر برسانند [۳].

با توجه به اهمیت روزافزون شبکه‌های اجتماعی و نیاز به بهینه‌سازی فرآیند گسترش انتشار اطلاعات، پژوهش در زمینه بهبود الگوریتم‌های حداکثرسازی گسترش انتشار اطلاعات امری ضروری است.

در این مقاله، به منظور بهبود عملکرد حداکثرسازی گسترش انتشار اطلاعات، یک رویکرد جدید مبتنی بر معیارهای شباهت کاربران معرفی شده است. در الگوریتم پیشنهادی، ابتدا از معیارهای شباهت برای وزندهی به گراف استفاده می‌شود. سپس، معیارهای انتخاب بر روی این گراف وزن‌دار اعمال می‌شود. به طور خاص، برای هر گره در شبکه، مجموع وزن یال‌های متصل به آن محاسبه می‌شود. در نهایت، چند گره‌ای که بالاترین امتیاز وزنی را دارند، به عنوان تأثیرگذارترین گره‌ها شناسایی و انتخاب می‌شوند.

در واقع مهم‌ترین نوآوری‌های این مقاله به شرح زیر است:

- (۱) وزن‌دار کردن گراف توسط معیارهای شباهت، و
- (۲) استفاده از مجموع وزنی گره به عنوان ابتکاری برای پیدا کردن گره‌های اولیه تأثیرگذار.

کاربر از طریق این شبکه‌ها با یکدیگر ارتباط برقرار کرده و به راحتی می‌توانند محتوای اطلاعاتی تولید کرده و به محتوای تولید شده توسط دیگران دسترسی پیدا کنند. این شبکه‌ها به کاربران امکان می‌دهند تا ایده‌ها، افکار، اخبار، شایعات و محتوای چندرسانه‌ای را در مدت زمان کوتاهی منتشر کنند. به همین دلیل، شبکه‌های اجتماعی برخط مانند فیسبوک، ایکس (توییتر سابق) و سایر سکوها شبکه‌های اجتماعی به بستری مناسب و محبوب برای تبلیغ محصولات و انتشار اطلاعات تبدیل شده‌اند. این بسترها نه تنها تعاملات اجتماعی را تسهیل می‌کنند، بلکه به شرکت‌ها و سازمان‌های مختلف نیز امکان می‌دهند تا با مخاطبان خود به‌طور مستقیم و مؤثر ارتباط برقرار کنند. از این‌رو، نقش شبکه‌های اجتماعی در بازاریابی اینترنتی و انتشار اطلاعات به طور چشمگیری افزایش یافته و اهمیت آن‌ها در دنیای امروزی غیرقابل انکار است [۱].

با توجه به رشد روزافزون شبکه‌های اجتماعی برخط و کاربردهای گسترده‌ای که در زمینه‌های مختلف دارند، اهمیت تحلیل این شبکه‌ها بیشتر از گذشته مشهود است. یکی از مهم‌ترین زمینه‌های تحقیقاتی در تحلیل شبکه‌های اجتماعی، بررسی و تحلیل گسترش انتشار اطلاعات است. این پدیده به یک ابزار قدرتمند برای بسیاری از کاربردها تبدیل شده و تلاش‌های تحقیقاتی گسترده‌ای را در حوزه‌های متعددی همچون علوم کامپیوتر، علوم اجتماعی، فیزیک و همه‌گیرشناسی به خود جذب کرده است. تحلیل انتشار اطلاعات در شبکه‌های اجتماعی می‌تواند به درک بهتر الگوهای ارتباطی، شناسایی گره‌های کلیدی و تأثیرگذار و بهینه‌سازی راهبردهای تبلیغاتی و اطلاع‌رسانی کمک کند. همچنین، این تحلیل می‌تواند در پیش‌بینی و مدیریت بحران‌های اجتماعی و بهداشتی، مانند شیوع بیماری‌ها، نقشی حیاتی ایفا کند. با توجه به پیچیدگی و پویایی شبکه‌های اجتماعی، توسعه مدل‌ها و الگوریتم‌های مؤثر برای شناسایی گره‌های تأثیرگذار به منظور حداکثرسازی گسترش انتشار اطلاعات از اهمیت

پایه‌گذار تحقیقات بعدی در زمینه تحلیل شبکه‌های اجتماعی و بازاریابی ویروسی بوده است [۵].

در سال ۲۰۰۳ کمپه و همکارانش مسئله مطرح شده توسط دومینگوس و ریچاردسون را برای تعدادی از مدل‌هایی که بیشتر در تحلیل شبکه‌های اجتماعی بررسی شده است، در نظر گرفتند. آنها یک الگوریتم تقریبی با تضمین بهینگی ارائه دادند و براساس یک چارچوب مبتنی بر توابع زیرپیمانی نشان دادند که راهبرد حریصانه، یک راه حل با بهینگی ۶۳ درصد است. این روش حریصانه در چند مرحله، گرهی را که بیشترین سود حاشیه‌ای را دارد برای اضافه کردن به مجموعه‌ی اولیه، انتخاب می‌کند، یعنی در هر تکرار آن گرهی انتخاب می‌شود که بیشترین سود را نسبت به فرایند گسترش تاثیر دارد. چارچوب پیشنهاد شده یک رویکرد کلی است که کارایی الگوریتم‌ها را برای مسئله حداکثرسازی گسترش انتشار اطلاعات در شبکه‌های اجتماعی تضمین می‌کند [۳]، [۶].

الگوریتم حریصانه‌ای که توسط کمپه و همکارانش ارائه شد، علیرغم ارائه تخمینی مناسب از گسترش تاثیر، به دلیل پیچیدگی محاسبه دقیق تاثیر مجموعه اولیه تحت مدل‌های انتشار آستانه خطی و آبخاری مستقل، از نوع ان-پی سخت است. به همین دلیل، برای به دست آوردن تخمین‌های دقیق‌تر، آنها از شبیه‌سازی مونت‌کارلو استفاده می‌کنند که شبیه‌سازی مونت‌کارلو به تعداد زیادی اجرا، معمولاً حدود ۱۰,۰۰۰ مرتبه اجرا، نیاز دارد که منجر به زمان محاسباتی طولانی و به ویژه برای شبکه‌های بزرگ، بسیار زمان‌بر است [۶].

به منظور کاهش تعداد شبیه‌سازی‌های مونت‌کارلو و ارتقاء مقیاس‌پذیری، لسکووک و همکارانش در سال ۲۰۰۷ الگوریتمی به نام CELF^۲ را معرفی کردند. این الگوریتم بر اساس این ایده طراحی شده است که سود حاشیه‌ای برای هر گره در تکرار فعلی نمی‌تواند بیشتر از سود حاشیه‌ای در تکرارهای قبلی باشد. به عبارت دیگر، اگر یک گره در مرحله‌ای از الگوریتم سود حاشیه‌ای کمتری نسبت به

در ادامه این مقاله، بخش دوم به طور مختصر به بررسی و تحلیل کارهای پیشین در زمینه حداکثرسازی گسترش انتشار اطلاعات پرداخته است. در بخش سوم، جزئیات الگوریتم پیشنهادی به طور کامل شرح داده شده است. بخش چهارم به ارزیابی کارایی الگوریتم براساس شبیه‌سازی‌های انجام شده بر روی داده‌های استاندارد شبکه‌های اجتماعی اختصاص یافته است. در نهایت، بخش پنجم به نتیجه‌گیری و ارائه پیشنهاداتی برای پژوهش‌های آینده می‌پردازد.

۲. کارهای پیشین

شناسایی یک مجموعه محدود چند عضوی از گره‌های یک شبکه اجتماعی که منجر به بیشترین میزان انتشار در شبکه شود، یک مسئله ان-پی سخت^۱ به شمار می‌رود. به همین دلیل، تاکنون روش‌های متعددی برای حل مسئله حداکثرسازی انتشار و یافتن تاثیرگذارترین گره‌ها ارائه شده است. این روش‌ها شامل الگوریتم‌های دقیق، تقریبی و ابتکاری هستند. الگوریتم‌های دقیق، هر چند که قادر به یافتن جواب بهینه هستند، به دلیل پیچیدگی محاسباتی بالا و زمان‌بر بودن، برای شبکه‌های بزرگ کاربرد عملی ندارند. به همین دلیل، الگوریتم‌های تقریبی و ابتکاری که توانایی یافتن راه‌حل‌های نزدیک به بهینه را در زمان معقول دارند، بیشتر مورد استفاده قرار می‌گیرند [۴].

در سال ۲۰۰۱، دومینگوس و ریچاردسون، یک مسئله الگوریتمی کلیدی مرتبط با بازاریابی ویروسی را در زمینه شبکه‌های اجتماعی مطرح کردند. مسئله به این صورت بود؛ که اگر بخواهیم از میان افراد، تعدادی را برای معرفی محصول جدید به دیگران انتخاب کنیم، بهترین روش برای انتخاب این افراد چگونه باید باشد، تا بتوان به بیشترین گسترش ممکن برای معرفی محصول دست یافت؟ این مطالعه به عنوان یکی از کارهای اولیه با ارائه مدل‌ها و الگوریتم‌های پایه‌ای در زمینه تحلیل شبکه‌های اجتماعی و بازاریابی ویروسی، از اهمیت زیادی برخوردار است و

2-CELF: Cost-Effective with Lazy-Forward

1-NP-hard

در [۹]، یک الگوریتم چند مرحله‌ای با تمرکز بر ساختارهای جوامع شبکه برای حداکثرسازی گسترش انتشار معرفی شده است به این صورت که براساس ساختار ارتباطی شبکه، ابتدا خوشه‌های شبکه استخراج شده و سپس براساس ویژگی‌های هر خوشه، متناسب با هر خوشه تعدادی از گره‌های اولیه از این خوشه‌ها انتخاب خواهد شد. در این روش تمرکز بر ارتباطات میان رئوس با توجه به تعاملات درون جامعه‌ای و میان جامعه‌ای است و کارایی روش صرفاً برای شبکه‌های دارای جوامع موثر است و برای شبکه‌های خلوت نیز خیلی کارا نیست.

در [۱۰]، مسئله حداکثرسازی انتشار به عنوان یک مسئله بهینه‌سازی در نظر گرفته شده است و برای تخمین تاثیرگذاری کاربران از یک معیار براساس آنتروپی استفاده شده است که برای حل آن استفاده از الگوریتم بهینه‌سازی گرگ خاکستری پیشنهاد شده است. الگوریتم گرگ خاکستری، یک الگوریتم بهینه‌سازی الهام گرفته از رفتارهای اجتماعی گرگ‌ها است، که شبیه‌سازی راهبردهای شکار گروهی گرگ‌ها به بهینه‌سازی آن برای شناسایی کاربران کلیدی در شبکه‌های اجتماعی مورد استفاده قرار گرفته است. الگوریتم مطرح شده توسط گرگ خاکستری در مقایسه با سایر الگوریتم‌های بهینه‌سازی فراابتکاری پارامتر قابل تنظیم کمتری دارد و می‌تواند با پیچیدگی زمانی کمتر و زمان همگرایی سریع‌تر فضای مسئله را جستجو کند.

در [۱۱]، الگوریتمی براساس فعالیت و اتصال گره‌ها طراحی شده است که برای شناسایی مجموعه‌ی اولیه با در نظر گرفتن دو مفهوم درجه و فعالیت گره کار می‌کند. در این پژوهش، نویسندگان این اصل را در نظر گرفته‌اند؛ که سرعتی که یک گره می‌تواند اطلاعات را گسترش دهد متناسب با اندازه تعداد اتصالات آن گره است. بنابراین، الگوریتم پیشنهادی از این فرضیه پیش‌بینی می‌کند که کاربرانی با سطح فعالیت بالا اما تعداد اتصالات کم به همان میزان باید خوب در نظر گرفته شوند اگر نامزدهای

مراحل پیشین داشته باشد، این امر نشان‌دهنده کاهش تاثیرگذاری آن در فرآیند انتشار است. با بهره‌گیری از این اصل، این الگوریتم توانسته است تعداد ارزیابی‌های مورد نیاز برای تخمین تاثیر را به طور چشمگیری کاهش دهد. این بهینه‌سازی منجر به کاهش قابل توجه زمان اجرای الگوریتم شد، به گونه‌ای که این الگوریتم می‌تواند فرآیند محاسبه را نسبت به الگوریتم حریصانه ارائه شده توسط کمپه و همکاران تا ۷۰۰ برابر سریع‌تر انجام دهد. این پیشرفت در بهینه‌سازی الگوریتم‌ها برای مسئله حداکثرسازی انتشار در شبکه‌های اجتماعی، بهبود قابل توجهی در کارایی تحلیل و تصمیم‌گیری‌های مرتبط با شبکه‌های بزرگ فراهم می‌آورد. به این ترتیب، این الگوریتم نه تنها زمان محاسبات را به حداقل می‌رساند، بلکه به تحلیل و برنامه‌ریزی بهینه‌تری در مقیاس‌های وسیع کمک می‌کند [۷].

در سال ۲۰۱۱، گویال و همکارانش الگوریتمی به نام ++CELFF معرفی کردند که نسخه بهبود یافته‌ای از الگوریتم CELFF است. ++CELFF به طور ویژه برای کاهش نیاز به شبیه‌سازی‌های غیرضروری مونت‌کارلو و بهره‌برداری از خاصیت زیرپیمانی تابع انتشار طراحی شده است. در این الگوریتم، هر گره در شبکه یک متغیر چندتایی را در خود ذخیره می‌کند که به تحلیل و محاسبه سود حاشیه‌ای کمک می‌کند. ایده اصلی ++CELFF این است که اگر یک گره در مرحله فعلی به عنوان هسته انتخاب شود، دیگر نیازی به محاسبه مجدد سود حاشیه‌ای آن گره در مراحل بعدی نیست. این ویژگی باعث کاهش زمان محاسباتی و افزایش کارایی الگوریتم می‌شود. با این بهبودها، ++CELFF توانسته است بین ۳۵٪ تا ۵۵٪ سریع‌تر از الگوریتم CELFF عمل کند. با این حال، حتی با این پیشرفت‌های قابل توجه در سرعت و کارایی فرآیند محاسباتی، ++CELFF نسبت به نسخه‌های قبلی همچنان به محاسبات فراوان مونت‌کارلو نیاز دارد که چالش‌های مربوط به مقیاس‌پذیری و زمان محاسبات همچنان باقی مانده است [۸].

بهتری در مقایسه با کاربران خفته با اتصالات بیشتر وجود نداشته باشد.

در [۱۲]، رویکردی مبتنی بر برنامه‌ریزی خطی ارائه شده است. این رویکرد شامل ۳ گام اصلی است. در گام اول گراف تاثیر از گراف شبکه اجتماعی با استفاده از ضرب ماتریس خلوت ساخته می‌شود. در گام دوم مسئله حداکثرسازی گسترش انتشار اطلاعات تحت مدل آستانه خطی به‌عنوان یک برنامه خطی با استفاده از گراف تاثیر ساخته شده در گام اول مدل‌سازی می‌شود. در گام سوم مجموعه اولیه با استفاده از نتایج برنامه خطی و روش گردکردن تصادفی انتخاب می‌شود. مزیت اصلی این روش زمان اجرای مناسب است که به اندازه مجموعه اولیه وابسته نیست.

در [۱۳]، به معرفی یک رویکرد نوآورانه برای حداکثرسازی تاثیر در شبکه‌های اجتماعی می‌پردازد. این مطالعه با تمرکز بر جنبه‌های عدالت اجتماعی و بهینه‌سازی چندهدفه، مدل جدیدی را برای انتخاب کاربران کلیدی در شبکه‌های اجتماعی پیشنهاد می‌دهد که به طور همزمان تاثیر و عدالت اجتماعی را در نظر می‌گیرد. روش پیشنهادی شامل استفاده از اطلاعات پیشین برای افزایش دقت و کاهش هزینه‌های محاسباتی مدل است و با رویکرد بهینه‌سازی چند هدفه سعی در ایجاد توازن میان گسترش موثر تاثیر و توزیع عادلانه آن در سراسر شبکه دارد.

در [۱۴]، یک الگوریتم سریع برای حداکثرسازی تاثیر در شبکه‌های اجتماعی با تمرکز بر ساختار جوامع همپوشان ارائه شده است. از آنجا که؛ در ساختارهای اجتماعی، گروه‌هایی از گره‌ها هستند که به طور جزئی با یکدیگر همپوشانی دارند، الگوریتم پیشنهادی با بهره‌گیری از این ویژگی و با استفاده معیاری با عنوان ضریب احتمال انتشار سراسری، بهینه‌سازی فرآیند انتخاب گره‌های کلیدی را در جوامع همپوشان شبکه‌های اجتماعی هدف قرار می‌دهد. این روش با ترکیب تحلیل ساختار جامعه و تکنیک‌های احتمال، قادر است به طور مؤثری گره‌های

تأثیرگذار را شناسایی کند و در عین حال زمان محاسبات را کاهش دهد.

در [۱۵]، یک چارچوب تکاملی چندتبدیلی برای حداکثرسازی گسترش تاثیر در شبکه‌های اجتماعی ارائه شده است. در این مطالعه، به رویکردهای جایگزینی برای کاهش محاسبات سنگین شبیه‌سازی های مونت کارلو برای ارزیابی گسترش انتشار و ضعف‌های آنها در فرآیند جستجو براساس اطلاعات پیشین شبکه اشاره شده است. در چارچوب پیشنهادی، از مجموعه‌ای از تبدیل‌ها و الگوریتم‌های تکاملی برای بهبود فرآیند شناسایی گره‌های تاثیرگذار شبکه‌های اجتماعی در سه مرحله استفاده شده است. نوآوری اصلی این مقاله در استفاده از راهبردهای چندتبدیلی برای بهبود کارایی و دقت الگوریتم‌های حداکثرسازی تاثیر است.

اکثر روش‌های ارائه شده برای حداکثرسازی انتشار اطلاعات که برخی از مهمترین و جدیدترین آنها در این بخش معرفی شد، دارای مشکل پیچیدگی زمانی هستند. در سال‌های اخیر با توجه به زمانبر بودن روش‌های حریصانه، استفاده از الگوریتم‌های ابتکاری بیشتر مورد توجه قرار گرفته است. از طرفی در اکثر پژوهش‌های انجام شده، نقش کلیه کاربران در تاثیرگذاری یکسان در نظر گرفته شده است به طوری که از گذشته صرفاً تعدد ارتباطات یک کاربر بدون در نظر گرفتن ساختار، شباهت و جایگاه آن نشاندهنده اهمیت یک کاربر است. بنابراین ارائه رویکردهایی که بتواند در مدل‌سازی مفروضات نزدیک به دنیای واقعی را بیشتر در محاسبات دخیل کند، مورد اهمیت است.

۳. الگوریتم پیشنهادی مبتنی بر معیارهای شباهت

ظهور شبکه‌های اجتماعی برخط به طور قابل توجهی نحوه ارتباط و تعامل افراد در سراسر جهان را تغییر داده و آنها را به جزئی جدایی‌ناپذیر از زندگی روزمره تبدیل

شباهت، از یکی از سه معیار شباهت جاکارد^۳، سالتون^۴ یا سورنسون^۵ به منظور وزن دار کردن گراف با هدف ارزیابی میزان موثر بودن ارتباط‌های به وجود آمده میان گره‌ها برای شناسایی کاربران تاثیرگذار در شبکه، استفاده می‌شود. که نحوه معیارهای شباهت جاکارد، سالتون و سورنسون به ترتیب برای دو گره دلخواه a و b به شرح روابط (۱) تا (۳) است:

$$S_{Jaccard}(a,b) = \frac{|\Gamma(a) \cap \Gamma(b)|}{|\Gamma(a) \cup \Gamma(b)|} \quad (1)$$

$$S_{Salton}(a,b) = \frac{|\Gamma(a) \cap \Gamma(b)|}{\sqrt{|\Gamma(a)| * |\Gamma(b)|}} \quad (2)$$

$$S_{Sorenson}(a,b) = \frac{2|\Gamma(a) \cap \Gamma(b)|}{|\Gamma(a)| + |\Gamma(b)|} \quad (3)$$

به طوری که در روابط (۱) تا (۳)، توابع $\Gamma(a)$ و $\Gamma(b)$ نشان‌دهنده تعداد مجموعه گره‌های همسایگی برای گره‌های a و b هستند.

۱.۳. محاسبه میزان تاثیرگذاری گره‌ها

بعد از محاسبه میزان شباهت میان گره‌ها، برای هر یک از یال‌های موجود یک وزن شباهت به دست آمده است که لازم است از آن برای محاسبه میزان تاثیرگذاری گره‌ها بهره برد. به منظور محاسبه میزان تاثیرگذاری گره‌ها و انتخاب گره‌های تاثیرگذار در مرحله بعدی، استفاده از وزن شباهت یال‌های خروجی هر گره پیشنهاد شده است، یعنی وزن هر گره برابر با مجموع وزنی شباهت یال‌های خروجی متصل به هر گره مطابق رابطه (۴) خواهد بود.

$$W_i = \sum_{(i,j) \in E} S(i,j) \quad (4)$$

به طوری که، $S(i,j)$ نشان‌دهنده وزن شباهت میان گره‌های i و j است و (i,j) معرف یال‌های خروجی متصل به گره i است.

۲.۳. انتخاب k گره تاثیرگذار

بعد از محاسبه مجموع وزن یال‌های متصل به هر گره

3-Jaccard
4-Salton
5-Sorenson

کرده است. این شبکه‌ها به اعضای خود امکان می‌دهند تا به طور مداوم با یکدیگر در ارتباط باشند و اطلاعات را از طریق بستر شبکه به اشتراک بگذارند [۳]. با توجه به آنکه، نظرات، افکار، احساسات و رفتار افراد می‌تواند تحت تأثیر تعاملات درون شبکه تغییر کند، معمولاً پذیرش پیشنهادها یا انجام اقدامات از سوی افراد به‌طور قابل توجهی به تأثیر نزدیکان، دوستان و همفکران وی مربوط می‌شود. در واقع، افراد تمایل دارند با کسانی که دارای سلايق و علايق مشابه هستند در ارتباط باشند؛ بنابراین، هرچه شباهت میان افراد بیشتر باشد، احتمال تأثیرپذیری نیز افزایش می‌یابد. این پدیده به‌ویژه در زمینه‌هایی مانند انطباق‌پذیری اجتماعی، فشار همسالان، و همچنین در زمینه‌های بازاریابی و فروش به‌وضوح قابل مشاهده است [۱۶]. بدین منظور، در این مقاله، الگوریتمی مبتنی بر شباهت کاربران برای شناسایی و انتخاب تاثیرگذارترین کاربران در شبکه، ارائه شده است. در الگوریتم پیشنهادی برای انتخاب مجموعه اولیه استفاده از یک معیار شباهت برای ارزیابی میزان شباهت کاربران پیشنهاد شده است. هدف از این کار، بررسی موثر بودن ارتباط به وجود آمده میان کاربران است، زیرا همه ارتباط‌های به وجود آمده در شبکه کارا نیستند و ممکن است میان دو کاربر ارتباطی حاصل شده باشد اما این ارتباط ضعیف بوده و موثر نباشد.

در الگوریتم پیشنهادی، پس از دریافت گراف ورودی و مقداردهی اولیه، سه مرحله اصلی شامل: (۱) وزن دار کردن گراف با انتساب وزن برای یال‌ها براساس معیار شباهت مابین گره‌ها، (۲) محاسبه‌ی میزان تاثیرگذاری هر گره براساس مجموع وزن یال‌های متصل به آن و (۳) انتخاب k گره تاثیرگذار براساس مرتب‌سازی میزان تاثیرگذاری گره به دست آمده به صورت نزولی انجام می‌پذیرد. در ادامه جزئیات هر مرحله ذکر شده است.

۱.۳. وزن دار کردن گراف با استفاده از معیارهای شباهت

در این مرحله؛ انتساب وزن به یال‌های گراف براساس معیارهای شباهت پیشنهاد شده است. برای محاسبه

مجموعه داده مشهور و پرکاربرد است که در سال ۱۹۷۷ توسط شخصی به نام وین زاکاری از اعضای یک باشگاه کاراته دانشگاهی جمع‌آوری شد. هرگره در گراف این مجموعه داده نماینده یک عضو از این باشگاه و هر یال نشان‌دهنده یک ارتباط بین دو عضو این باشگاه است. این مجموعه داده حاوی ۲۴ گره و ۷۸ یال با میانگین درجه ۴/۵۸ است.

- دلفین^۸ [۱۸]: مجموعه داده دلفین بین سال‌های ۱۹۹۴-۲۰۰۱ از نژاد خاصی از دلفین‌ها که در نیوزلند زندگی می‌کنند، جمع‌آوری شده است. این مجموعه حاوی ۶۲ گره و ۱۵۹ یال با میانگین درجه ۵/۱۲ است.
- فوتبال^۹ [۱۹]: مجموعه داده فوتبال آمریکایی، شامل بازی‌های فوتبال آمریکایی بین دانشکده‌های بخش ۱ در پاییز سال ۲۰۰۰ است. این مجموعه حاوی ۱۱۵ گره و ۶۱۳ یال با میانگین درجه ۱۰/۶۶ است.
- جاز^{۱۰} [۲۰]: مجموعه داده جاز از همکاری بین نوازندگان جاز در سال ۲۰۰۳ جمع‌آوری شده است. در گراف این مجموعه داده، هر گره نماینده یک نوازنده جاز و هر یال نشان‌دهنده دو نوازنده‌ای است که در یک گروه با هم می‌نوازند. این مجموعه حاوی ۱۹۸ گره و ۲۷۴۲ یال با میانگین درجه ۲۷/۶۹ است.
- ایمیل^{۱۱} [۲۱]: این مجموعه داده یک شبکه ارتباطی از طریق ایمیل متعلق به دانشگاهی در اسپانیا است که گره‌ها نشان‌دهنده کاربران و یال‌ها نشان‌دهنده حداقل یک ایمیل ارسال شده است. این مجموعه حاوی ۱۱۳۳ گره و ۵۴۵۱ یال با میانگین درجه ۹/۶۲ است.
- بلاگ^{۱۲} [۲۲]: مجموعه داده وب‌نوشت‌های سیاسی، شامل پیوندها بین وب‌نوشت‌های سیاسی درباره سیاست ایالات متحده آمریکا است. این مجموعه حاوی ۱۲۲۴ گره و ۳۳۴۳۰ یال با میانگین درجه ۵۴/۶۲ است.

گراف، براساس وزن به‌دست آمده برای گره، یک لیست مرتب از وزن گره‌ها به صورت نزولی ایجاد می‌شود و در نهایت k گره بالای لیست به عنوان k گره تاثیرگذار انتخاب می‌شوند. پس از استخراج k گره تاثیرگذار، ارزیابی گره‌های انتخابی بر روی مدل آستانه خطی به عنوان یک مدل انتشار استاندارد اعمال شده و خروجی‌های به‌دست آمده گزارش می‌شوند.

مدل آستانه خطی^۶ یکی از مدل‌های پرکاربردی است که برای بررسی گسترش تاثیر مجموعه گره‌های اولیه به کار برده می‌شود. در این مدل هر گره می‌تواند توسط هر یک از همسایگان متصل ورودی تحت تاثیر قرار گیرد. به هر گره یک مقدار آستانه تاثیر اختصاص داده می‌شود که این مقدار نشان‌دهنده آستانه تاثیرگذاری هر گره است. در هر لحظه هر گره فعال می‌تواند براساس میزان تاثیرگذاری مشخص شده برای یال بر روی گره‌های مجاور خود تاثیر بگذارد، در صورتی که مجموع تاثیرات دریافتی از همسایگان یک گره برابر یا بیشتر از مقدار آستانه مشخص شده باشد، آن گره نیز فعال شده و امکان تاثیرگذاری بر روی دیگر همسایگان را نیز خواهد داشت. این فرآیند تا توقف تعداد گره‌های فعال ادامه می‌یابد.

در نهایت k گره انتخاب شده بر روی مدل انتشار اعمال شده و خروجی به‌دست آمده ارزیابی خواهد شد.

۴. شبیه‌سازی

به منظور شبیه‌سازی ابتدا مجموعه داده معرفی شده و سپس شرایط شبیه‌سازی و نحوه مقایسه ذکر شده است.

۱.۴. مجموعه داده

آزمایش‌ها بر روی شش مجموعه داده کاراته، دلفین، فوتبال، جاز، ایمیل و بلاگ انجام شده است. در ادامه جزئیات معرفی شش مجموعه داده استفاده شده ذکر شده است:

- کاراته^۷ [۱۷]: مجموعه داده باشگاه کاراته زاکاری یک

8-Dolphins network
9-American Football network
10-Jazz Musicians network
11-Email network
12-Political blogs network

6-LTM: Linear Threshold Model
7-Zachary Karate Club network

۲.۴. پیکره‌بندی و تنظیمات

با توجه به آن که در الگوریتم پیشنهادی، برای محاسبه شباهت می‌توان از ۳ معیار مختلف جاکارد، سالتون و سورنسون استفاده کرد، به منظور اختصار در نامگذاری، برای نسخه‌ای از الگوریتم که از معیار جاکارد، سالتون و سورنسون بهره می‌برد، به اختصار و به ترتیب به از عنوان JSIM، SaSIM و SoSIM استفاده شده است.

برای مقایسه ارزیابی عملکرد الگوریتم‌های پیشنهادی، نتایج به دست آمده با نتایج چهار الگوریتم پایه به شرح زیر استفاده شده است.

- انتخاب تصادفی^{۱۳}: در این روش که به طور اختصاری Rnd نامگذاری شده است، به طور تصادفی با توزیع احتمال یکنواخت، k گره از میان گره‌های گراف انتخاب می‌شود.
- انتخاب بالاترین درجه^{۱۴}: این روش بر تعداد ارتباطات مستقیم گره تمرکز دارد و مجموع تعداد یال‌های متصل به هر گره به عنوان درجه آن گره مطابق رابطه (۵) در نظر گرفته می‌شود

$$d_i = \sum_j A_{ij} \quad (5)$$

به طوریکه d_i نشان‌دهنده درجه گره i است و A_{ij} به ازای وجود یال مقدار یک و به ازای فقدان یال مقدار ۰ اختیار می‌کند. بنابراین، در انتخاب بالاترین درجه، k گره دارای بالاترین درجه انتخاب می‌شود که به طور اختصاری Deg نامگذاری شده است.

- انتخاب کاهش درجه^{۱۵}: این روش، مشابه انتخاب بالاترین درجه است با این تفاوت که انتخاب k گره به صورت تدریجی یکی یکی انجام شده و پس از انتخاب هر گره، یال‌های متصل به گره‌هایی که قبلاً انتخاب شده برای گره‌های نامزد انتخاب نشده منظور نمی‌شود. برای سادگی این روش به طور اختصاری Ddc نامگذاری شده است.

- انتخاب بالاترین بینابینی^{۱۶}: این روش براساس معیار بینابینی بر روی انتخاب گره‌های واسطه‌ای که نقش

ارتباطی مهمی مابین بخش‌های مختلف گراف را دارند، تاکید دارد. در واقع، بینابینی یک گره مطابق رابطه (۶) محاسبه می‌شود.

$$b_i = \frac{\sigma_{st}(i)}{\sigma_{st}} \quad (6)$$

به طوری که، b_i نشان‌دهنده مقدار بینابینی گره i ، σ_{st} شامل کلیه کوتاهترین مسیرهایی است که مابین گره‌های s و t وجود دارد و $\sigma_{st}(i)$ شامل آن دسته از کوتاهترین مسیرهایی است که از گره i عبور می‌کند، بنابراین در این روش، k گره دارای بالاترین مقدار بینابینی انتخاب می‌شود که به طور اختصاری به صورت Btn نامگذاری شده است.

برای پیاده‌سازی، از کتابخانه‌های NetworkX، pandas و NDlib در پایتون استفاده شده است. گراف‌های ورودی مورد استفاده بدون وزن در نظر گرفته شده است. در کلیه آزمایش‌ها، از روش شبیه‌سازی مونت کارلو با تعداد اجرای ۱۰۰۰ مرتبه و مدل انتشار آستانه خطی مطابق مراجع [۲۴]، [۲۵] و [۲۶] استفاده شده است به طوری که به هر گره مقداری تصادفی آستانه در بازه $[0, 1]$ و به هر یال وزنی تصادفی برابر با یک تقسیم بر درجه هر گره اختصاص داده شده است. نتایج گزارش شده در ادامه، حاصل میانگین ۳۰ بار اجرا برای هر الگوریتم است. در ادامه کارایی الگوریتم‌های پیشنهادی با روش‌های Rnd، Deg، Ddc و Btn به عنوان الگوریتم‌های پایه و چهار الگوریتم CIM^{۱۷} [۲۳]، PIDS^{۱۸} [۲۴]، CFIN^{۱۹} [۲۵] و BPr^{۲۰} [۲۶] به عنوان الگوریتم‌های جدید مقایسه شده است. نتایج در بخشی بعدی ذکر شده است.

۴.۳. نتایج آزمایش‌ها

در این بخش سه آزمایش مختلف به منظور ارزیابی عملکرد الگوریتم پیشنهادی طراحی شده است. در آزمایش اول، نتایج گسترش انتشار اطلاعات توسط تعداد گره‌های اولیه مختلف فعال شده با الگوریتم دیگر مورد بررسی

17-CIM: Community-based Influence Maximization

18-PIDS: Positive Influence Dominating Set

19-CFIN: Community Finding Influential Node

20-BPr: Between-based Pruning

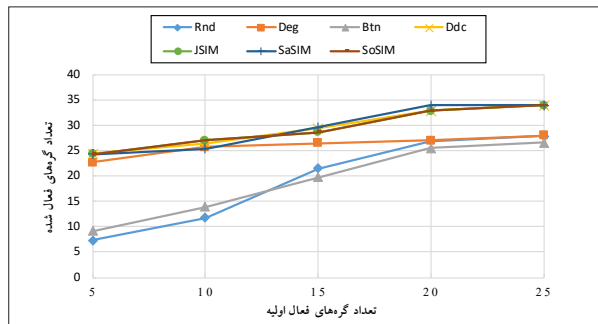
13-Random Selection

14-High Degree Centrality Selection

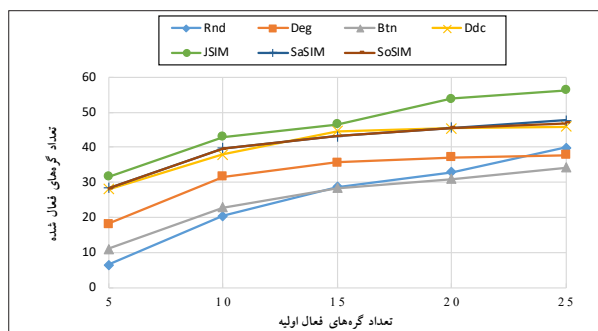
15-Degree Discount Centrality Selection

16-High Betweenness Centrality Selection

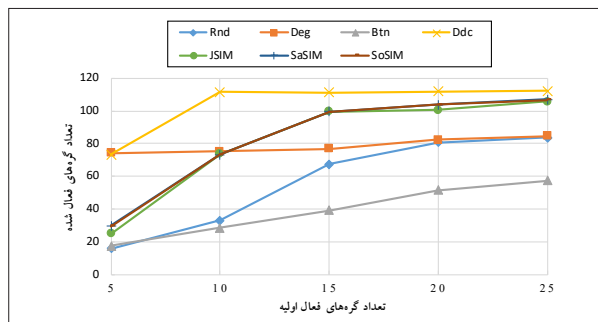
روش پیشنهادی با بهترین نتیجه از میان سایر روش‌ها برای هر شبکه در هر سناریو مورد مقایسه قرار گرفته است و متوسط بهبود برابر ۴/۹۷ درصد است.



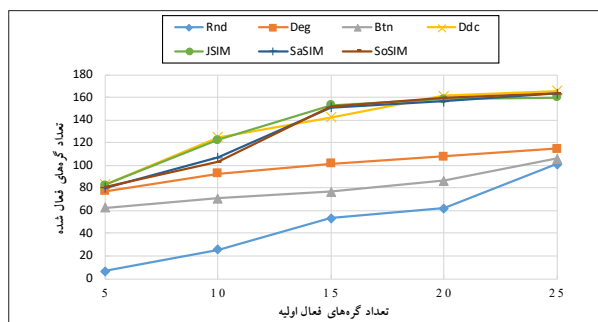
شکل (۱): میانگین گره‌های فعال شده نهایی به ازای تعداد گره‌های فعال اولیه مختلف برای مجموعه داده کاراته



شکل (۲): میانگین گره‌های فعال شده نهایی به ازای تعداد گره‌های فعال اولیه مختلف برای مجموعه داده دلفین



شکل (۳): میانگین گره‌های فعال شده نهایی به ازای تعداد گره‌های فعال اولیه مختلف برای مجموعه داده فوتبال



قرار گرفته است. در آزمایش دوم، درصد مختلف گره‌های فعال نتیجه براساس شاخص نرخ پوشش در مقایسه با الگوریتم‌های اخیر مقایسه شده است و در نهایت در آزمایش سوم، کارایی الگوریتم‌ها از لحاظ زمان اجرا مقایسه شده است.

۴.۳.۱. آزمایش اول: بررسی نتایج تعداد گره‌های فعال شده

در این آزمایش خروجی میانگین تعداد گره‌های فعال شده در پایان فرآیند انتشار تحت مدل انتشار به ازای تعداد گره‌های فعال اولیه مختلف به ازای انتخاب از ۵ گره تا ۲۵ گره مورد بررسی قرار گرفته است. نتایج به دست آمده برای این آزمایش در مقایسه با روش‌های Rnd, Btn, Deg, Ddc در شکل‌های ۱ تا ۶ به ترتیب برای مجموعه داده‌های کاراته، دلفین، فوتبال، جاز، ایمیل و بلاگ به ازای تعداد گره‌های اولیه از ۵ گره تا ۲۵ گره برای تعداد گره‌های فعال شده در پایان فرآیند انتشار نمایش داده شده است.

با توجه به نتایج به دست آمده در کلیه شکل‌ها، به ازای افزایش تعداد گره‌های فعال شده انتخابی، روند افزایشی برای تعداد گره‌های فعال شده در پایان فرآیند انتشار تحت مدل انتشار مشاهده می‌شود. براساس نتایج به دست آمده، کمترین بازدهی برای روش‌های انتخاب تصادفی^{۲۱} و سپس بینابینی (Btn) به دست آمده است و الگوریتم‌های پیشنهادی در ۳ حالت مختلف با عنوان‌های: الگوریتم پیشنهادی براساس معیار شباهت جاکارد^{۲۲}، الگوریتم پیشنهادی براساس معیار شباهت سالتون^{۲۳} و الگوریتم پیشنهادی براساس معیار شباهت سورنسون^{۲۴} در ۴ مجموعه داده از نتایج بهتری برخوردارند. همچنین در شبکه‌های خلوت با توجه به آن که اطلاعات ساختاری کمک زیادی به محاسبه شباهت نمی‌کند، تمایز بالایی بین گره‌های مختلف وجود ندارد و عملاً کارایی روش کاهش یافته است. جهت بررسی میزان بهبود عملکرد روش پیشنهادی بهترین نتیجه هر سه

21-Rnd

22-JSIM: Jaccard Similarity-based Influence Maximization

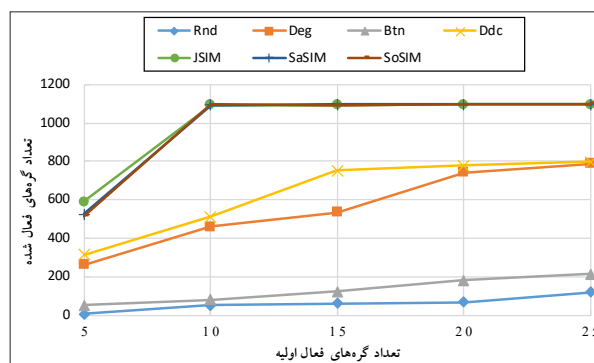
23-SaSIM: Salton Similarity-based Influence Maximization

24-SoSIM: Sorenson Similarity-based Influence Maximization

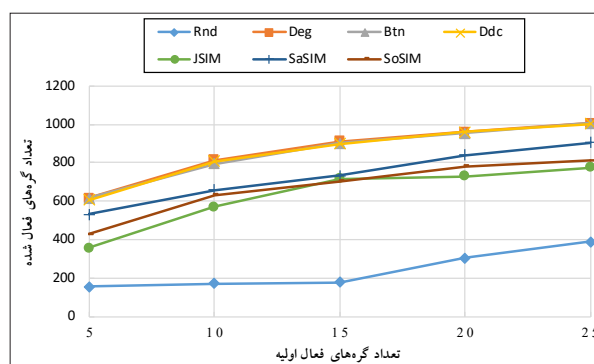
گرفته و نتایج شاخص نرخ پوشش برای الگوریتم‌های پیشنهادی (JSIM، SaSIM، SoSIM) و سه الگوریتم CIM [۲۳]، PIDS [۲۴] و CFIN [۲۵] در شکل ۷ برای کاراته، در شکل ۸ برای دلفین، در شکل ۹ برای فوتبال، در شکل ۱۰ برای جاز، در شکل ۱۱ برای ایمیل، در شکل ۱۲ برای بلاگ نمایش داده شده است. در شکل ۷، به عنوان مثال به ازای فعال شدن ۵ درصد (۲ گره) از گره‌های اولیه مجموعه داده کاراته توسط الگوریتم CIM، ۲۳/۵۳ درصد از گره‌های شبکه در پایان فرآیند انتشار تحت مدل انتشار فعال خواهند شد و این مقدار برای ۲۵ درصد گره اولیه در نهایت ۶۴/۷۱ درصد از کل گره‌های شبکه است.

با توجه به نتایج ارائه شده در شکل‌های ۷ تا ۱۲، مشاهده می‌شود که به طور کلی با افزایش درصد گره‌های فعال اولیه، روند افزایشی برای گسترش انتشار تحت مدل انتشار در پایان فرآیند انتشار وجود دارد. در رابطه با خروجی به‌دست آمده در میان مجموعه داده های مختلف، در مجموعه داده کاراته، روش PIDS نتایج بهتری برای شاخص نرخ پوشش برخوردار بوده است. برای مجموعه داده دلفین، الگوریتم SaSIM نتایج مناسبی دارد، در مجموعه داده فوتبال، روش SoSIM بهترین نتایج را دارد، در مجموعه داده جاز، روش JSIM نتایج مناسبی دارد، در مجموعه داده ایمیل، روش PIDS بهترین نتایج را تولید کرده است و در مجموعه داده بلاگ، روش JSIM بالاترین نتایج را به‌دست آورده است. بنابراین، روش پیشنهادی متناسب با ساختار گراف با روش PIDS قابل رقابت است. در رابطه با مقایسه بین ۳ روش پیشنهادی تقریباً رفتار مشابهی میان آنها مشاهده می‌شود، اما در میان همین روش‌ها، روش SoSIM در مجموعه داده مختلف بهترین نتایج را نسبت به JSIM و SaSIM تولید کرده است. بنابراین، از لحاظ شاخص نرخ پوشش، روش JSIM، نسبت به بهترین نتیجه رقیب در هر سناریو براساس نتایج ۶ مجموعه داده مختلف به

شکل (۴): میانگین گره‌های فعال شده نهایی به ازای تعداد گره‌های فعال اولیه مختلف برای مجموعه داده کاراته جاز



شکل (۵): میانگین گره‌های فعال شده نهایی به ازای تعداد گره‌های فعال اولیه مختلف برای مجموعه داده کاراته ایمیل



شکل (۶): میانگین گره‌های فعال شده نهایی به ازای تعداد گره‌های فعال اولیه مختلف برای مجموعه داده بلاگ

۴.۳.۲. آزمایش دوم: بررسی نتایج شاخص نرخ پوشش

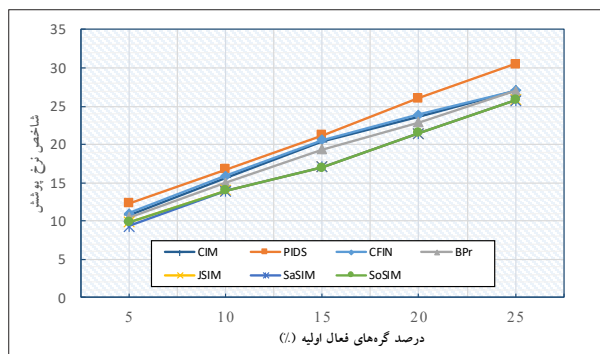
در این آزمایش از شاخص نرخ پوشش [۲۷] برای مقایسه کارایی روش‌های پیشنهادی با سایر روش‌های اخیر استفاده شده است. شاخص نرخ پوشش (CR) به‌صورت رابطه ۷ تعریف می‌شود:

$$CR = \frac{F(S)}{|V|} \quad (7)$$

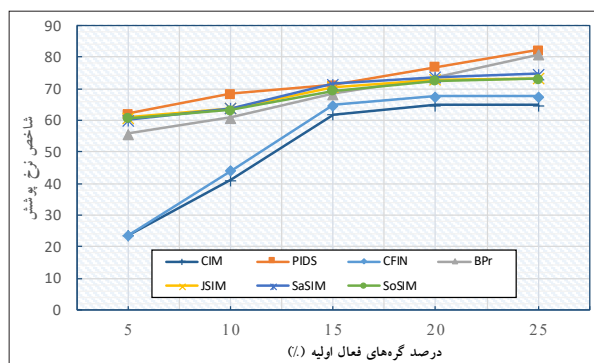
به‌طوری که، $F(S)$ تعداد گره‌های فعال شده در پایان فرآیند انتشار توسط مجموعه گره‌های اولیه S و $|V|$ تعداد گره‌های گراف است.

در این آزمایش، نتیجه نهایی برحسب درصد از ۵ درصد تا ۲۵ درصد گره فعال اولیه مورد بررسی قرار

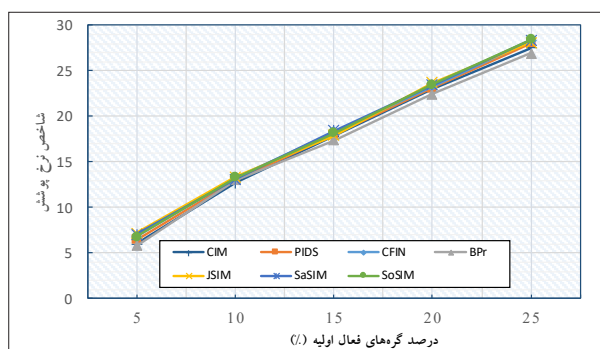
شکل (۱۰): نتایج شاخص نرخ پوشش به ازای درصد گره‌های فعال اولیه مختلف برای مجموعه داده جاز



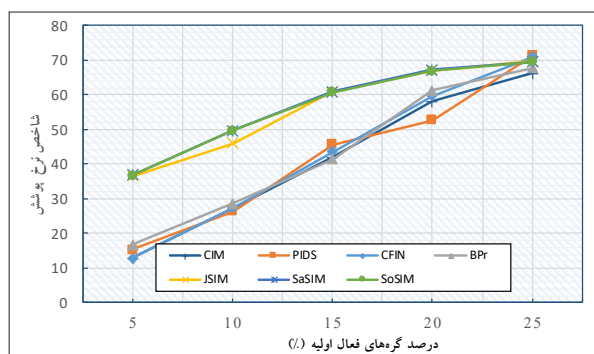
طور متوسط ۷/۸۶ درصد بهبود داشته است.



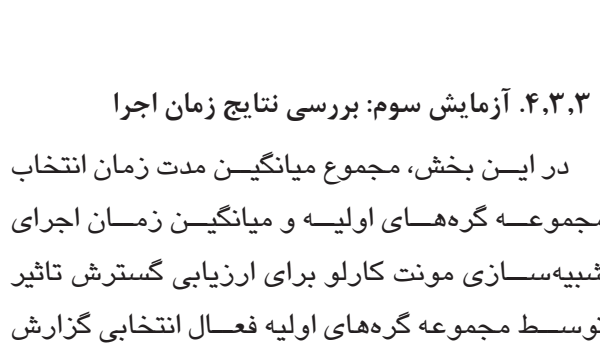
شکل (۱۱): نتایج شاخص نرخ پوشش به ازای درصد گره‌های فعال اولیه مختلف برای مجموعه داده ایمیل



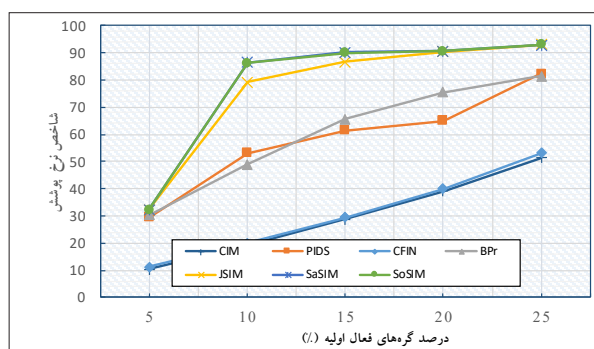
شکل (۷): نتایج شاخص نرخ پوشش به ازای درصد گره‌های فعال اولیه مختلف برای مجموعه داده کاراته



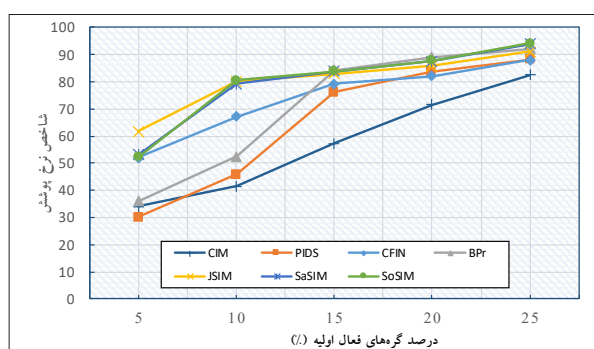
شکل (۱۲): نتایج شاخص نرخ پوشش به ازای درصد گره‌های فعال اولیه مختلف برای مجموعه داده بلاگ



شکل (۸): نتایج شاخص نرخ پوشش به ازای درصد گره‌های فعال اولیه مختلف برای مجموعه داده کاراته دلفین



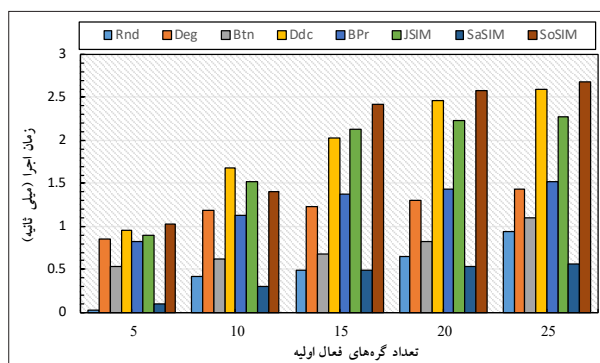
شکل (۹): نتایج شاخص نرخ پوشش به ازای درصد گره‌های فعال اولیه مختلف برای مجموعه داده فوتبال



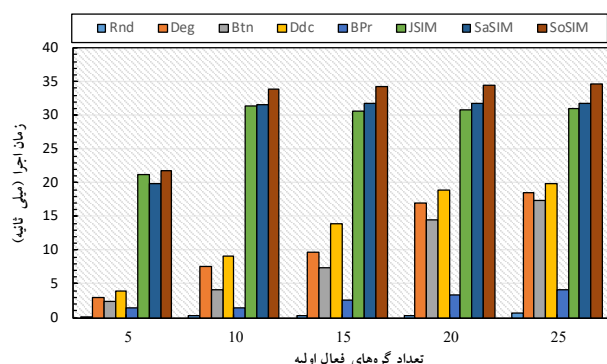
۴,۳,۳. آزمایش سوم: بررسی نتایج زمان اجرا

در این بخش، مجموع میانگین مدت زمان انتخاب مجموعه گره‌های اولیه و میانگین زمان اجرای شبیه‌سازی مونت کارلو برای ارزیابی گسترش تاثیر توسط مجموعه گره‌های اولیه فعال انتخابی گزارش شده است. در شکل‌های ۱۳ تا ۱۸، زمان اجرای الگوریتم به ازای تعداد گره‌های اولیه مختلف از ۵ گره تا ۲۵ گره گزارش شده است. با توجه به نتایج به دست آمده، بدون در نظر گرفتن روش انتخاب تصادفی، روش انتخاب مبتنی بر بینابینی^{۲۰} و درجه زمان اجرای کمتری را دارا هستند و در میان ۳ روش پیشنهادی، روش JSIM در ۴ مجموعه داده دارای کمترین زمان اجرا و روش SoSIM در ۳ مجموعه داده دارای بیشترین زمان اجرا است. با

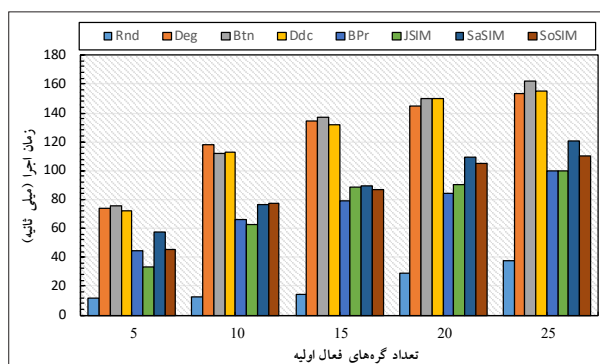
برای مجموعه داده فوتبال



شکل (۱۶): نتایج زمان اجرا به ازای تعداد گره‌های فعال اولیه مختلف برای مجموعه داده جاز



شکل (۱۷): نتایج زمان اجرا به ازای تعداد گره‌های فعال اولیه مختلف برای مجموعه داده ایمیل

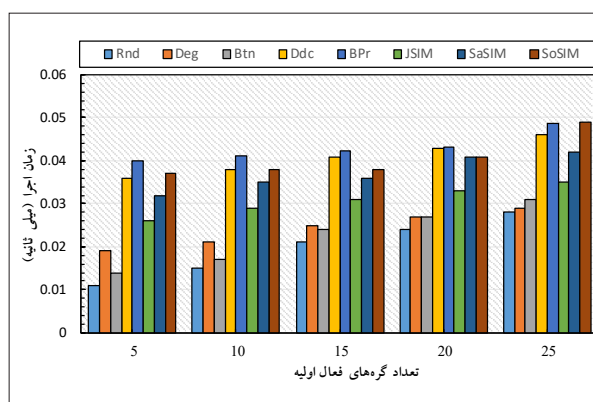


شکل (۱۸): نتایج زمان اجرا به ازای تعداد گره‌های فعال اولیه مختلف برای مجموعه داده بلاگ

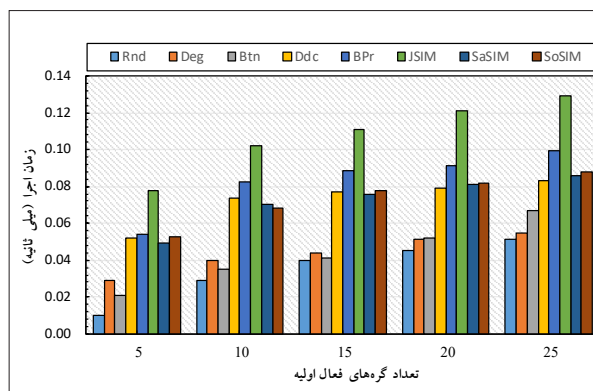
۵. نتیجه‌گیری

در سال‌های اخیر شبکه‌های اجتماعی برخط مانند فیس‌بوک، توئیتر و دیگر شبکه‌های اجتماعی به بستری

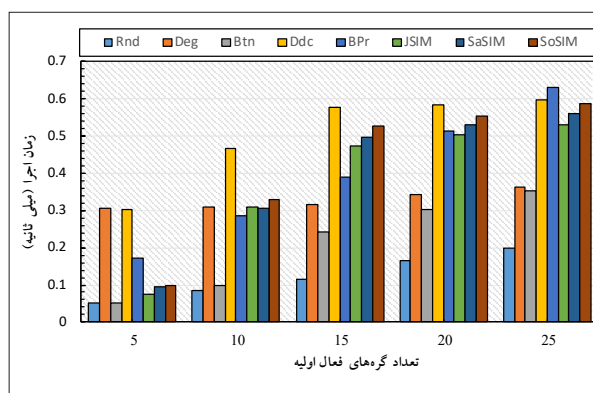
توجه به نتایجی که در آزمایش‌های دوم و سوم برای میزان انتشار مشاهده شد و زمان اجرای به‌دست آمده، می‌توان چنین تحلیل کرد که روش‌هایی که گسترش انتشار بیشتری را به وجود آورده به زمان اجرای بیشتری نیز نیاز داشته‌اند.



شکل (۱۹): نتایج زمان اجرا به ازای تعداد گره‌های فعال اولیه مختلف برای مجموعه داده کاراته



شکل (۲۰): نتایج زمان اجرا به ازای تعداد گره‌های فعال اولیه مختلف برای مجموعه داده دلفین



شکل (۲۱): نتایج زمان اجرا به ازای تعداد گره‌های فعال اولیه مختلف

- er International Publishing, 2019, pp. 315–329.
3. D. Kempe, J. Kleinberg, and É. Tardos, “Maximizing the spread of influence through a social network,” in Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York, New York, USA: ACM Press, 2003, pp. 137–146.
 4. A. Zareie and R. Sakellariou, “Influence maximization in social networks: a survey of behaviour-aware methods,” *Social Network Analysis and Mining*, vol. 13, no. 1, p. 78, 2023.
 5. P. Domingos and M. Richardson, “Mining the network value of customers,” in Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining, ACM, 2001, pp. 57–66.
 6. D. Kempe, J. Kleinberg, and É. Tardos, “Influential nodes in a diffusion model for social networks,” in *Automata, Languages and Programming*, vol. 3580, in Lecture Notes in Computer Science, vol. 3580, 2005, pp. 1127–1138.
 7. 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, ACM, 2007, pp. 420–429.
 8. A. Goyal, W. Lu, and L. V. Lakshmanan, “CELFF++: optimizing the greedy algorithm for influence maximization in social networks,” in Proceedings of the 20th international conference companion on World Wide Web, ACM, 2011, pp. 47–48.
 9. J. Guo and W. Wu, “Influence Maximization: Seeding Based on Community Structure,” *ACM Transactions on Knowledge Discovery from Data*, vol. 14, no. 6, pp. 1–22, 2020..
 10. A. Zareie, A. Sheikhamadi, and M. Jalili, “Identification of influential users in social network using gray wolf optimization algorithm,” *Expert Systems with Applications*, vol. 142, p. 112971, 2020.
 11. B. Saxena and P. Kumar, “A node activity and connectivity-based model for influence maximization in social networks,” *Social Network Analysis and Mining*, vol. 9, no. 1, p. 40, 2019.
 12. F. Ghayour-Baghbani, M. Asadpour, and H. Faili, “MLPR: Efficient influence maximization in linear threshold propagation model using linear programming,” *Social Network Analysis and Mining*, vol. 11, no. 1, p. 3, 2021.
 13. H. Gong and C. Guo, “Influence maximization considering fairness: A multi-objective optimization approach with prior knowledge,” *Expert Systems with Applications*, vol. 214, p. 119138, 2023.
 14. A. Bouyer, H. A. Beni, B. Arašteh, Z. Aghaee, and R. Ghanbarzadeh, “FIP: A fast overlapping community-based Influence Maximization Algorithm using probability coefficient of global diffusion in social networks,” *Expert systems with applications*, vol. 213, p. 118869, 2023.
 15. C. Wang, J. Zhao, L. Li, L. Jiao, J. Liu, and K. Wu, “A مناسب و محبوب برای تبلیغ محصولات و گسترش اطلاعات تبدیل شده‌اند و یافتن افراد تاثیرگذار در شبکه اجتماعی از موضوعات مهم در تحلیل شبکه‌های اجتماعی محسوب می‌شود. در این مقاله یک الگوریتم ابتکاری برای حل مسئله حداکثرسازی گسترش انتشار اطلاعات با هدف یافتن تاثیرگذارترین کاربران با استفاده از محاسبه شباهت میان گره‌های شبکه پیشنهاد شد. در الگوریتم پیشنهادی، یال‌های گراف با استفاده از یک معیار شباهت مانند جاکارد، سالتون یا سورنسون وزن‌دار شده، سپس برای انتخاب مجموعه اولیه، وزن یال‌های متصل به هر گره برای آن گره محاسبه شده و k گره با بیشترین وزن به عنوان مجموعه گره‌های اولیه برای فعال‌سازی در نظر گرفته شدند. برای ارزیابی کارایی الگوریتم پیشنهادی، آزمایش‌هایی بر روی چند مجموعه داده استاندارد شبکه‌های اجتماعی انجام شد که نتایج شبیه‌سازی حاکی از مطلوبیت نتایج الگوریتم پیشنهادی از نظر انتشار و زمان اجرا دارد. در کارهای آینده با توجه به حجم بالای محاسبات شبیه‌سازی مونت کارلو برای محاسبه تاثیر هر گره و همچنین تعداد بالای محاسبات شباهت، به نظر می‌رسد بتوان از الگوریتم‌های یادگیری ماشین به عنوان یک کاندید برای تخمین مقادیر تاثیرگذاری و شباهت گره‌ها بهره برد و تا جایی که امکان‌پذیر است، محاسبات غیرضروری را کاهش داد.
- می‌توان از نقش یادگیری ماشین و یادگیری عمیق برای تخمین ارزیابی انتشار اشاره کرد تا بتوان کارایی الگوریتم‌های مختلف در مقیاس بزرگ را بهبود بخشید.

مراجع

1. S. S. Singh, D. Srivastava, M. Verma, and J. Singh, “Influence maximization frameworks, performance, challenges and directions on social network: A theoretical study,” *Journal of King Saud University-Computer and Information Sciences*, vol. 34, no. 9, pp. 7570–7603, 2022.
2. A. Rezvanian, B. Moradabadi, M. Ghavipour, M. M. Daliri Khomami, and M. R. Meybodi, “Social Influence Maximization,” in *Learning Automata Approach for Social Networks*, A. Rezvanian, B. Moradabadi, M. Ghavipour, M. M. Daliri Khomami, and M. R. Meybodi, Eds., in *Studies in Computational Intelligence*, Cham: Springer-

- multi-transformation evolutionary framework for influence maximization in social networks,” *IEEE Computational Intelligence Magazine*, vol. 18, no. 1, pp. 52–67, 2023.
16. A. Kumar, S. S. Singh, K. Singh, and B. Biswas, “Link prediction techniques, applications, and performance: A survey,” *Physica A: Statistical Mechanics and its Applications*, vol. 553, p. 124289, 2020.
 17. M. Girvan and M. E. J. Newman, “Community structure in social and biological networks,” *Proceedings of the National Academy of Sciences*, vol. 99, no. 12, pp. 7821–7826, 2002.
 18. D. Lusseau, “The emergent properties of a dolphin social network,” *Proceedings of the Royal Society of London. Series B: Biological Sciences*, vol. 270, no. suppl_2, 2003.
 19. J. Park and M. E. Newman, “A network-based ranking system for US college football,” *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2005, no. 10, p. P10014, 2005.
 20. P. M. Gleiser and L. Danon, “Community structure in jazz,” *Advances in complex systems*, vol. 6, no. 04, pp. 565–573, 2003.
 21. J. Leskovec, K. J. Lang, A. Dasgupta, and M. W. Mahoney, “Community structure in large networks: Natural cluster sizes and the absence of large well-defined clusters,” *Internet Mathematics*, vol. 6, no. 1, pp. 29–123, 2009.
 22. L. A. Adamic and N. Glance, “The political blogosphere and the 2004 U.S. election: divided they blog,” in *Proceedings of the 3rd international workshop on Link discovery*, Chicago Illinois: ACM, 2005, pp. 36–43.
 23. Y.-C. Chen, W.-Y. Zhu, W.-C. Peng, W.-C. Lee, and S.-Y. Lee, “CIM: community-based influence maximization in social networks,” *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 5, no. 2, p. 25, 2014.
 24. M. M. D. Khomami, A. Rezvanian, N. Bagherpour, and M. R. Meybodi, “Minimum positive influence dominating set and its application in influence maximization: a learning automata approach,” *Applied Intelligence*, vol. 48, no. 3, pp. 570–593, 2018.
 25. M. M. D. Khomami, A. Rezvanian, M. R. Meybodi, and A. Bagheri, “CFIN: A community-based algorithm for finding influential nodes in complex social networks,” *The Journal of Supercomputing*, no. 77, pp. 2207–2236, 2021.
 26. P. Soltanzadeh and A. Rezvanian, “A New Pruning Method for Identifying Influential Nodes in Social Networks,” in *8th International Conference on Web Research (ICWR2022)*, Tehran, Iran, 2022, pp. 1–6.
 27. J.-L. He, Y. Fu, and D.-B. Chen, “A novel top-k strategy for influence maximization in complex networks with community structure,” *PloS one*, vol. 10, no. 12, 2015.