دو فصلنامه علمی
فناوری اطلاعات و ارتباطات ایران

سال چهاردهم، شماره‌های ۵۳ و ۵۴، پاییز زمستان
۱۴۰۱
صفحات:۲۶۷ تا ۲۹۲

A Comprehensive Review on the Influence Maximization Problem in Social Networks with Focusing on the Community Detection Methods

Mohsen Taheri Nia [*] ،**Mehdi Ismaili**[*] ،**Behrouz Minai Bidgoli**[**]
[*] Faculty of Computer Science, Islamic Azad University, Kashan branch, Kashan, Iran.
[**] Faculty of Computer Science, Islamic Azad University, Kashan branch, Kashan, Iran.

Abstract:
With the incredible development of online social networks, many marketers have exploited the opportunity to find influential people within online social networks to influence other people. The finding of influential users is known as the Influence Maximization problem, which is a key algorithmic problem in influence analysis. This problem aims to select a limited set of most influential users to maximize the number of influenced users in an online social network. This problem has been studied in recent years extensively due to its significant role in various applications, such as viral marketing, advertising, recommender systems, social media analysis, etc. The non-stop growth of social networks has intensified the time efficiency (scalability) and effectiveness challenges of this NP-Hard problem. Researchers have been controlled these challenges by exploiting the communities' structure as a useful feature of social networks. Inspired by these points, this paper provides a comprehensive review of the state-of-the-art algorithms in the influence maximization problem, focusing on the community-based models. At first, the outstanding greedy and heuristics algorithms are surveyed in addressing these challenges. Then, existing community-based algorithms especially are reviewed and compared. In the end, several directions are suggested for future researches in the influence maximization problem.

Keywords: Influence Maximization, Social Networks Analysts, Influencer, Social Networks, Community Detection, Diffusion Models.

مروری جامع بر مسئله‌ی بیشینه‌سازی تأثیر در شبکه‌های اجتماعی با تمرکز بر روش‌های مبتنی بر شناسایی اجتماعات

محسن‌طاهری‌نیا*، مهدی‌اسماعیلی* و بهروز مینایی‌بیدگلی**

* دانشکده کامپیوتر، دانشگاه آزاد اسلامی واحد کاشان، کاشان، ایران.

** دانشکده مهندسی کامپیوتر، دانشگاه علم و صنعت ایران، تهران، ایران.

تاریخ دریافت:۱۴۰۰/۰۵/۲۲ تاریخ پذیرش:۱۴۰۰/۰۸/۲۳

نوع مقاله : پژوهشی

چکیده

با تو سعه روزافزون شبکه‌های اجتماعی، ب سیاری از بازاریابان از این فر صت ا ستفاده کرده و سعی در یافتن افراد تأثیرگذار در شبکه‌های اجتماعی آنلاین دارند. این م سئله که به عنوان م سئله‌ی بی شینه سازی تأثیر شناخته می شود، یک م سئله الگوریتمی در تجزیه و تحلیل شبکه‌های اجتماعی ا ست. هدف از م سئله‌ی بی شینه سازی تأثیر، یافتن تعداد محدودی از کاربران تأثیرگذار در یک شبکه اجتماعی ا ست به‌طوری‌که بیشترین تعداد ممکن از کاربران دیگر را تحت تأثیر خود قرار دهند. این مسئله در زمینه‌های مختلف مانند بازاریابی ویروسی، تبلیغات، سیستم‌های پیشنهاددهنده، تجزیه و تحلیل ر سانه‌های اجتماعی در سال‌های اخیر به طور گسترده مورد مطالعه قرار گرفته ا ست. بنابراین، نیاز به یک مروری سازمان‌یافته و جامع در رابطه با این موضوع وجود دارد.

کارایی زمانی و اثربخ شی دو معیار مهم در تولید الگوریتم‌های برج سته در حوزه م سئله‌ی بی شینه سازی تأثیر مح سوب می شوند. برخی از محققان با بهره‌گیری از ساختار اجتماعات به‌عنوان ویژگی بسیار مفید شبکه‌های اجتماعی، این دو موضوع را به‌طور مشهودی بهبود بخشیده‌اند. اکثر مقالات مروری موجود، مطالعاتی را که تأثیر اجتماعی را بر اساس ساختار اجتماعات تحلیل کرده‌اند، نادیده گرفته‌اند. هدف این مقاله برر سی جامع الگوریتم‌های برج سته پی شنهاد شده در حوزه م سئله‌ی بی شینه سازی تأثیر در شبکه‌های اجتماعی با تأکید ویژه بر رویکردهای مبتنی بر شناسایی اجتماعات است.

واژه‌گان کلیدی: بیشینه‌سازی تأثیر، تجزیه‌وتحلیل شبکه‌های اجتماعی، افراد تأثیرگذار، شبکه‌های اجتماعی، شناسایی اجتماعات، مدل‌های انتشار.

۱. مقدمه

در دو دهه گذشته، طیف وسیعی از شبکه‌های اجتماعی ظهور کرده‌اند که در آن ها حجم ز یادی از اطلاعات و ا یده ها از طریق تعاملات بین مردم منتشر می‌شود. تأثیر اجتماعی پدیده‌ای است که در آن افراد یک شبکه اجتماعی از طریق پیروی از دوستان، آشنایان، همکاران خود تحت تأثیر آن ها قرار می‌گیرند. یکی از مهم‌ترین موضوعات مطرح در تحلیل شبکه‌های اجتماعی، مسئله‌ی بیشینه‌سازی تأثیر است. هدف از مسئله‌ی بیشینه‌سازی تأثیر، شنا سایی مجموعه‌ای از تأثیرگذارترین افراد در شبکه‌های اجتماعی ا ست به‌طوری‌که اگر فرآیند انت شار اطلاعات (تأثیر) از این مجموعه آغاز شود، در نهایت بیشترین تعداد افر[1]اد شبکه تحت تأثیر قرار بگیرند [۱]. این مسئله در بسیاری از زمینه‌های محبوب مانند بازاریابی، تبلیغات، سیستم‌های پیشنهاددهنده و غیره کاربردی شده است [۲]. بسیاری از بازاریاب‌ها با بهره‌گیری از گسترش شبکه‌های اجتماعی آنلاین، به‌جای اینکه استراتژی‌های بازاریابی سنتی را دن بال کن ند، سعی می‌کن ند از طریق اثر «د هان به د هان» یا رویکردهای بازاریابی ویروسی، بر کاربران رسانه‌های اجتماعی آنلاین تأثیر بگذارند. شایستگی قابل توجه اثر «دهان‌به‌دهان» و رویکردهای بازاریابی ویروسی این است که آن‌ها سهم شرکت‌های کوچک با

[1] نویسنده مسئول:مهدی اسماعیلی m.esmaeili@iaukashan.ac.ir

منابع محدود را از بازار افزایش می‌دهند. برای توضیح بیشتر، شرکتی را در نظر بگیرید که قصد دارد محصول جدیدی را با کمترین بودجه به‌صورت آنلاین در بازار تبلیغ کند. به دلیل منابع محدود، شرکت باید تعداد محدودی از کاربران تأثیرگذار را در داخل شبکه اجتماعی برای تجربه (رایگان یا با اندکی تخفیف) این محصول انتخاب کند. پیش‌بینی می‌شود که این کاربران تأثیرگذار در مکانیسمی آبشاری، محصول جدید را به سایر کاربران متصل به خود توصیه کنند (تبلیغ کنند). در نهایت، این محصول به تعداد زیادی از کاربران در دسترس معرفی شده و پذیرفته خواهد شد؛ بنابراین، به شرطی که کاربران تأثیرگذار اولیه به‌درستی انتخاب شوند، میزان تأثیرگذاری در شبکه حداکثر خواهد بود.

مسئله‌ی بیشینه سازی تأثیر ابتدا توسط دومینگوس و همکاران در بازاریابی ویروسی معرفی شد [۳]. سپس کمپ و همکاران این مسئله را به عنوان یک مسئله بهینه سازی گسسته فرموله کردند و دو مدل انتشار را برای شبیه‌سازی فرآیند انتشار اطلاعات در شبکه پیشنهاد کردند: مدل آستانه خطی[۲] و مدل آبشاری مستقل[۳] [۴]. آن‌ها ثابت کردند که یافتن راه حل بهینه تحت این دو مدل انتشار NP-hard است و یک الگوریتم حریصانه[۴] با تقریب ۶۳٪ از راه حل بهینه ارائه دادند. افزایش اندازه شبکه‌های اجتماعی باعث کاهش کارایی زمانی این الگوریتم حریصانه شد [۵]. برای مقابله با این چالش کارایی زمانی، طیف وسیعی از الگوریتم‌ها در سال‌های اخیر ارائه شده است.

مسئله بیشینه سازی تأثیر در شبکه‌های اجتماعی، از مسائل مهم کاربردی در بسیاری از زمینه ها مثل بازار یابی، تبلیغات و سیستم‌های پیشنهاددهنده است. جذابیت و فراگیری روزافزون شبکه های اجتماعی باعث شده که محققان زیادی تحقیقات متنوعی را پیرامون مسئله بیشینه‌سازی تأثیر در شبکه‌های اجتماعی انجام دهند. از آنجایی که این حوزه تحقیقاتی، هنوز یک موضوع چالش‌برانگیز به‌حساب می‌آید، همواره روش‌ها و الگوریتم های متنوعی در حال معرفی و بهبود کارایی هستند. بر این اساس، مقاله حاضر در تلاش است که با مرور روش‌های موجود برای شناسایی گره‌های تاثیرگذار در شبکه‌های‌اجتماعی دید جامعی در رابطه با این موضوع پدید آورد.

یکی از نقاط قوت برجسته این مقاله‌ی مروری، تمرکز بر الگوریتم های مبتنی بر شناسایی اجتماعات[۵] در مسئله‌ی بیشینه‌سازی تأثیر است. ساختار اجتماعات یکی از حیاتی‌ترین ویژگی‌های شبکه‌های اجتماعی است. در دنیای واقعی، مردم تمایل دارند با افرادی که از نظر اندیشه، رفتار و علاقه مشابه هستند تعامل داشته باشند. این موضوع منجر به تشکیل اجتماعات می شود. در مسئله‌ی بیشینه‌سازی تأثیر که هدف آن افزایش میزان تأثیرگذاری در شبکه است، استفاده از ساختار اجتماعات منجر به افزایش اثربخشی خواهد شد. از سوی دیگر، رشد نمایی اندازه شبکه‌های اجتماعی چالشی بزرگ در کارایی زمانی و مقیاس‌پذیری[۶] مسئله‌ی بیشینه‌سازی تأثیر است. یک راه حل بسیار مفید برای غیال بر این چالش، تبدیل یک گراف (شبکه اجتماعی) عظیم به زیرگراف‌های کوچک‌تر (اجتماعات) به عنوان یک استراتژی تقسیم و غیال است. از این‌رو محققان به‌طور گسترده از ساختار اجتماعات برای بهبود کارایی زمانی و اثربخشی الگوریتم‌های خود در مسئله‌ی بیشینه سازی تأثیر استفاده کرده‌اند. ازآنجاکه هیچ مقاله مروری که الگوریتم‌های مبتنی بر شناسایی اجتماعات را پوشش دهد وجود نداشت[1,2,5–11]، این مقاله به بررسی جامع الگوریتم‌های برجسته پیشنهادشده در حوزه مسئله‌ی بیشینه‌سازی تأثیر با تمرکز ویژه بر رویکردهای مبتنی بر شناسایی اجتماعات می‌پردازد. ادامه این مقاله به شرح زیر است: در بخش ۲ مسئله‌ی بیشینه‌سازی تأثیر، مدل‌های انتشار اطلاعات و فرآیند شناسایی اجتماعات شرح داده می‌شود. بخش ۳ مرور ادبیات جامع از الگوریتم‌های افزایش تأثیر را ارائه می‌دهد. سرانجام، بخش ۴ به جمع‌بندی و ارائه مسیری برای تحقیقات آتی می‌پردازد. در شکل ۱ چارچوب کلی مقاله نشان داده شده است.

شکل ۱. قالب کلی مقاله

۲. مسئله‌ی بیشینه‌سازی تأثیر

جدول ۱ نمادهای پرتکرار این مقاله را نشان می‌دهد.

[2] Linear Threshold
[3] Independent Cascade
[4] Greedy algorithm
[5] Community Detection
[6] Scalability

جدول ۱ . نمادهای پرتکرار

شرح	نماد
گراف G با مجموعه گره V و مجموعه یال E	$G(V, E)$
به ترتیب برابر تعداد گره‌ها و یال‌ها در گراف G	m, n
مجموعه سبد گره‌های فعال	S
تعداد عناصر مجموعه S	k
تعداد شبیه‌سازی مونت‌کارلو	R
مدل انتشار	M
تعداد گره‌های تحت تأثیرقرارگرفته توسط مجموعه S تحت مدل انتشار M در گراف G	$\delta_{G,M}(S)$
حد آستانه فعال‌سازی گره v در مدل انتشار آستانه خطی	θ_v
احتمال فعال‌سازی گره v در مدل انتشار آبشاری مستقل	$P_{u,v}$
وزن تأثیریال جهت‌دار (v,u) از گره v به سمت گره u	$b_{v,u}$
مجموعه همسایگان ورودی گره v	$nbr^{in}(v)$
مجموعه همسایگان خروجی گره v	$nbr^{out}(v)$
درجه گره v	$deg(v)$
درجه ورودی گره v	$deg^{in}(v)$
درجه خروجی گره v	$deg^{out}(v)$

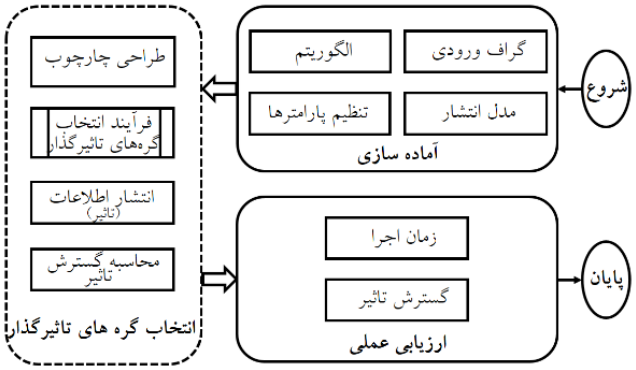
۱.۲. بیشینه‌سازی تأثیر

به‌طورکلی، تأثیر اجتماعی به توانایی تلقین یا تغییر ایده‌ها، باورها، افکار یا اعمال یک فرد یا یک اجتماع اشاره دارد. بر اساس موارد ذکرشده، مسئله‌ی بیشینه‌سازی تأثیر، مسئله یافتن مجموعه کوچکی از افراد شبکه اجتماعی است که میزان تأثیر اجتماعی را در شبکه‌های اجتماعی حداکثر کنند [۵]. مسئله‌ی بیشینه‌سازی تأثیر به شرح زیر فرموله می‌شود:

فرض کنید که گراف $G(V, E)$ داده شده است که در آن هر گره $v \in V$ نشان‌دهنده‌ی یک کاربر و هر یال (u, v) نشان‌دهنده رابطه بین کاربر u و v است. همچنین عدد صحیح مثبت $k \leq |V|$ به عنوان تعداد افراد مجموعه سبد[۷] S داده شده باشد. در ضمن مدل M یک مدل انتشار تصادفی است که فرآیند پخش اطلاعات را از افراد مجموعه سبد S در گراف G شبیه‌سازی می‌کند. میزان گسترش تأثیر مجموعه سبد S با k گره برابر با تعداد گره‌های تحت تأثیرقرارگرفته توسط مجموعه S تحت مدل انتشار M است که با $\delta_{G,M}(S)$ نشان داده می‌شود. هدف از مسئله‌ی بیشینه‌سازی تأثیر، یافتن زیرمجموعه‌ی $S \subseteq V$ با k عنصر است که S^* نامیده می شود، به‌طوری‌که میزان تأثیر S^* بر گراف G با توجه به قوانین احتمالی مشخص شده توسط مدل M، بیشینه باشد [۴]، یعنی،

$$\delta_{G,M}(S^*) = \text{argmax}_{|S|\leq k, S\subseteq V}\delta_{G,M}(S) \quad (۱)$$

لازم به ذکر است که میزان گسترش تأثیر به‌طور مستقیم به مدل انتشار احتمالی M بستگی دارد. در شکل ۲ چارچوب کلی مسئله بیشینه‌سازی تاثیر نشان داده شده است.



شکل ۲. چارچوب کلی مسئله بیشینه‌سازی تاثیر

۲.۲. مدل‌های انتشار[۸]

مدل‌های انتشار، مدل‌های تصادفی هستند که فرآیند انتشار اطلاعات را در شبکه‌های اجتماعی شبیه سازی می‌کنند [۴]. هدف همه مدل‌های انتشار، انعکاس ارتباطات بین افراد در شبکه‌های اجتماعی واقعی است. در مدل‌های انتشار، هر گره $v \in V$ دارای وضعیت فعال یا غیرفعال است. وضعیت فعال گره v بدین معنی است که گره v تحت تأثیرقرار گرفته است و همچنین قابلیت فعال‌سازی و تأثیرگذاری روی همسایگان خود را دارد. وضعیت غیرفعال به معنای آن است که تحت تأثیرقرار نگرفته است و قابلیت فعال‌سازی همسایگان خود را ندارد. مدل آستانه خطی و مدل آبشاری مستقل از پرکاربردترین مدل‌های انتشار در این مسئله هستند. در این مدل‌ها، وضعیت یک گره می‌تواند از حالت غیرفعال به وضعیت فعال تبدیل شود اما نمی‌توان وضعیت گره را از حالت فعال به غیرفعال تغییر داد.

وضعیت همه گره‌ها در ابتدا غیرفعال است. پس از انتخاب k گره اولیه و ساخت مجموعه سبد $(S \subseteq V)$، وضعیت گره‌های مجموعه سبد S به وضعیت فعال تغییر می‌یابد. روند انتشار اطلاعات از گره‌های فعال سبد S آغاز می شود. آن‌ها می‌توانند هم‌سایگان خود را در گام‌های زمانی جداگانه فعال کنند. گره‌هایی که اخیراً فعال شده‌اند می‌توانند همسایگان خود را فعال کنند و این روال همین‌طور ادامه پیدا می‌کند. این فرآیند انتشار هنگامی‌که گره جدیدی فعال نشده باشد، به پایان خواهد رسید. مدل‌های مختلف انتشار دارای استراتژی‌های متنوعی برای تبدیل حالت گره از غیرفعال به فعال هستند. مدل‌های آستانه خطی و آبشاری مستقل به‌عنوان متداول‌ترین مدل‌های انتشار در شبکه‌های اجتماعی دارای استراتژی‌های مختلف برای شبیه‌سازی فرآیند انتشار اطلاعات هستند [۴].

[۷] Seed set

[۸] Diffusion Models

۱.۲.۲. مدل آستانه خطی

فعال شدن گره v در مدل آستانه خطی بستگی به فعال‌بودن همسایگان آن دارد. در این مدل، یک عدد تصادفی θ_v از توزیع یکنواخت در بازه $[0,1]$ به هر گره v به‌عنوان آستانه فعال‌سازی آن گره اختصاص داده شده است. همچنین به هر یال جهت‌دار از گره u به گره v، یک وزن $w_{u,v}$ به‌عنوان مقدار تأثیر گره u روی گره v اختصاص داده می‌شود، به‌طوری‌که شرط

$$\sum_{u \in neighbor^{in}(v)} w_{u,v} \leq 1 \qquad (۲)$$

برقرار باشد [۴]. گره v در گام زمانی بعدی فعال خواهد شد، اگر مجموع وزن همسایگان فعال گره v حداقل برابر θ_v باشد [۴]، به‌عبارت‌دیگر

$$\sum_{u \in neighbor^{in}(v) \cap u \in S} w_{u,v} \geq \theta_v \qquad (۳)$$

در نهایت هنگامی که هیچ گره جدیدی فعال نشده باشد، فرآیند انتشار خاتمه می‌یابد.

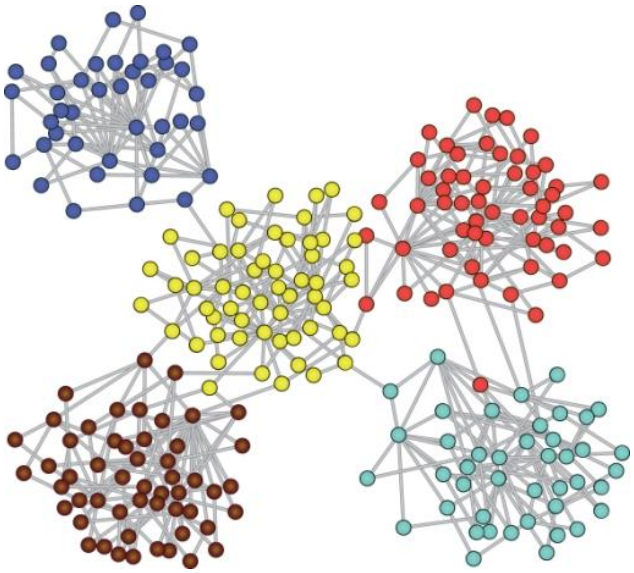
۲.۲.۲. مدل آبشاری مستقل

استراتژی فعال‌سازی در مدل آبشاری مستقل به شرح زیر است. فرض کنید که مجموعه S_{t-1} شامل گره‌های فعال‌شده در گام زمانی $t-1$ باشد. در گام زمانی t، هر گره $u \in S_{t-1}$ فقط یک شانس برای فعال کردن هر همسایه غیرفعال v با احتمال فعال‌سازی $P_{u,v}$ دارد. اگر گره u موفق شود، آنگاه گره v در گام زمانی $t+1$ فعال شده و در تمام گام‌های زمانی بعدی فعال باقی می‌ماند، در غیر این صورت غیرفعال خواهد ماند. صرف‌نظر از نتیجه فعال‌سازی گره u روی همسایگان خود، گره u هرگز نمی‌تواند گره‌ای را در گام زمانی بعدی فعال کند. این روند گسترش تا زمانی که امکان فعال‌سازی بیشتر وجود نداشته باشد، ادامه می‌یابد. توجه داشته باشید که در مواردی که چندین گره فعال سعی در فعال‌سازی یک گره غیرفعال دارند، تلاش‌ها مستقل از یکدیگر خواهند بود. مدل آبشاری وزن‌دار[9] توسعه‌ای از مدل آبشاری مستقل است که تابع احتمالی فعال‌سازی آن بر اساس درجه هر گره است [۴] :

$$P_{u,v} = 1/degree^{in}(v) \qquad (۴)$$

۳.۲. شناسایی اجتماعات

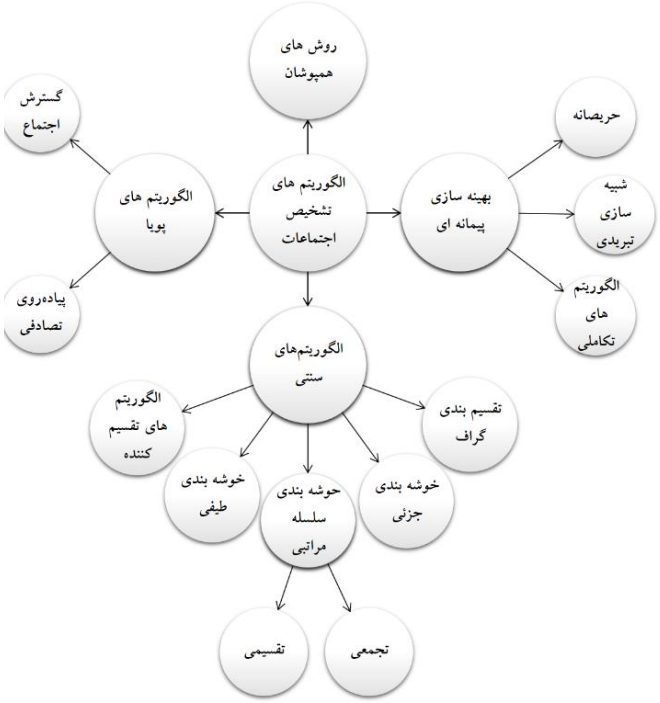
اجتماعات ساختارهای پنهان و یکی از ویژگی‌های مفید شبکه‌های اجتماعی هستند [۱۲]. در سال‌های اخیر موضوع شناسایی اجتماعات توجه زیادی را در تجزیه‌وتحلیل شبکه‌های اجتماعی به خود جلب کرده است [۱۳]. از روش شناسایی اجتماعات می‌توان برای استخراج خواص مفیدی از شبکه‌ها استفاده کرد. اجتماع به گروهی از گره‌ها گفته می شود که دارای بیشترین تراکم یال‌ها بین گره‌های گروهشان و کمترین تراکم یال‌ها با گره‌های گروه‌های دیگر هستند. در شکل ۳ یک شبکه اجتماعی که دارای ۵ اجتماع ناهمپوشان[10] است، رسم شده است.



شکل ۳. یک گراف با پنج اجتماع ناهمپوشان

۱.۳.۲. روش‌های تشخیص اجتماع

تاکنون، الگوریتم‌های زیادی برای یافتن اجتماعات در شبکه‌های پیچیده پیشنهاد شده است. در این بخش، یک مرور کلی بر چهار گروه مهم از الگوریتم‌های تشخیص اجتماعات ارائه شده‌است. شکل ۴ نیز روش‌های کلی تشخیص اجتماعات را به تصویر کشیده است.



شکل ۴. الگوریتم‌های شناسایی اجتماعات

۱.۱.۳.۲. الگوریتم‌های سنتی[11]

الف. تقسیم بندی گراف[12]

این روش گراف را به k خوشه با اندازه‌های ازپیش‌تعریف شده

[9] Weighted Cascade
[10] Nonoverlapping (Disjoint) commiuty
[11] Traditional Techniques
[12] Graph partitioning

تقسیم می‌کند، به طوری که تعداد یال‌های داخل یک خوشه از تعداد یال‌های بین خوشه‌ها متراکم‌تر باشد. الگوریتم‌های مشهور تکنیک های تقسیم بندی گراف عبارتند از روش Spectral Bisection [۱۴] و الگوریتم Kernighan-Lin [۱۵].

ب. خوشه‌بندی سلسله مراتبی[۱۳]

گراف‌ها ممکن است دارای ساختار سلسله مراتبی باشند، یعنی هر اجتماع ممکن است مجموعه‌ای از خوشه‌های کوچک در سطوح مختلف باشد. در مدل سلسله مراتبی نیاز به دانستن ماتریس شباهت و یا ماتریس الگو خواهیم داشت. هر چند این روش ما را از دانستن تعداد جوامع به عنوان ورودی بی نیاز می‌کند اما در عین حال تصمیم گیری برای اینکه درخت ایجاد شده در چه عمقی بهترین تقسیم شبکه را دارد قطعی نیست. ایجاد دسته‌های کوچک خوشه‌ها که معمولاً اطلاعات مهمی را نمی‌دهد از سایر معایب این روش به حساب می‌آید. الگوریتم‌های خوشه‌بندی سلسله مراتبی را می توان به دو دسته تقسیم کرد:

• الگوریتم‌های تجمعی: این یک تکنیک از پایین به بالا است زیرا در ابتدا هر گره را به عنوان یک خوشه جداگانه در نظر گرفته و به طور مکرر آنها را بر اساس بالاترین شباهت ادغام می‌کند و به اجتماع منحصر به فرد ختم می‌شود.

• الگوریتم‌های تقسیم کننده: این یک تکنیک از بالا به پایین است زیرا در ابتدا کل شبکه را به عنوان یک خوشه واحد در نظر می گیرد و با حذف یال‌های بین گره‌های کم شباهت، خو شه را به طور مکرر تقسیم می‌کند و به جوامع منحصر به فرد ختم می‌شود.

ج. خوشه‌بندی جزئی[۱۴]

خوشه‌بندی جزئی یک گراف را به خوشه‌های غیر همپوشان از پیش تعیین شده تقسیم می‌کند. هدف این است که گره‌ها را به k خوشه تقسیم کنیم تا تابع هزینه را بر اساس اندازه‌گیری عدم شباهت بین گره‌ها حداکثر/حداقل کنیم. الگوریتم‌های k-mean [۱۶] و fuzzy k-mean [۱۶] نمونه‌هایی از تکنیک‌های خوشه‌بندی جزئی هستند.

د. خوشه‌بندی طیفی[۱۵]

روش های طیفی نیز حجم زیادی از تحقیقات در این زمینه را شامل می شوند، چرا که اصولاً نتایج این دسته از الگوریتم‌ها در مقایسه با روش‌های مبتنی بر شباهت از کیفیت بالاتری برخوردار است. این روش‌های به عنوان پایه تئوریک برای طراحی الگوریتم‌های شناسایی اجتماع معرفی شده‌اند . در روش‌های تشخیص اجتماع مبتنی بر آنالیز طیف، یک الگوریتم خوشه‌بندی طیف نیمه نظارت شده با بررسی ماتریسی گره‌ها در شبکه و خوشه‌بندی طیف توسعه داده می‌شود. هر چند کیفیت این روش از روش‌های خوشه‌بندی محض بیشتر می باشد ولی غالبا سربار محاسباتی بالایی نیز دارد. الگوریتم Laplacian spectral partitioning [۱۷] یک نمونه از تکنیک‌های خوشه‌بندی طیفی است.

ه. الگوریتم‌های تقسیم‌کننده[۱۶]

این روش‌ها یال‌های بین‌خوشه‌ای را، بر اساس کمترین شباهت حذف می‌کند تا جوامع را از یکدیگر جدا کند. نمونه‌های اصلی این دسته شامل الگوریتم Girvan-Newman [۱۸] است که در آن یال‌ها به صورت تکراری بر اساس نمره میانگی یال[۱۷] حذف می شوند.

۲.۱.۳.۲. تکنیک‌های مبتنی بر بهینه‌سازی پیمانه‌ای[۱۸]

پیمانگی[۱۹] یک تابع کیفیت برای تقریب جوامع است. هرچه مقدار پیمانگی بیشتر باشد، تقسیم‌بندی بهتر است. گریمان و نیومن معیار پیمانگی Q را بصورت زیر تعریف کردند [۱۹]:

$$Q = \frac{1}{2m}\sum_{u,v\epsilon V\ ,u\neq v}\left(A_{uv} - \frac{k_u \cdot k_v}{2m}\right)\cdot \sigma(C_u, C_v) \qquad (۵)$$

که در آن m برابرتعداد یال های گراف، A ماتریس مجاورتی گراف k_u برابر درجه گره u است. همچنین هرگاه دو گره v و u در یک اجتماع قرار داشته باشند، تابع $\sigma(C_u, C_v)$ مقدار ۱ و درغیر اینصورت مقدار ۰ را برمی‌گرداند. این دسته شامل سه تکنیک زیر است.

الف. تکنیک های حریصانه[۲۰]

الگوریتم جستجوی حریصانه نیومن [۲۰] اولین الگوریتم پیشنهادی برای بهینه‌سازی پیمانه‌ای بود. این یک تکنیک تجمعی است که در ابتدا، هر گره متعلق به یک پیمانه مجزا است، سپس آنها بر اساس افزایش مقدار پیمانگی به طور تکراری ادغام می شوند. الگوریتم CNM[۲۱] نسخه سریع الگوریتم حریص نیومن است که توسط ساختار داده های کارآمد اجرا می‌شود [۲۱]. Louvain یک الگوریتم حریص مبتنی بر بهینه سازی پیمانه‌ای برای کشف جوامع در گراف‌های وزنی پیچیده است [۲۲]. این روش اجتماعات متفاوتی را به هر گره اختصاص می‌دهد (هر گره یک اجتماع). سپس به طور مکرر گره‌ها را بر اساس افزایش مقدار پیمانگی ادغام می‌کند. در صورت عدم افزایش، گره در اجتماع خود باقی می‌ماند. این روش تا زمانی که امکان بهبود بیشتر وجود نداشته باشد، تکرار می‌شود.

ب. شبیه‌سازی تبریدی[۲۲]

این یک رویکرد تصادفی گسسته است که از بهینه سازی پیمانه‌ای برای بهینه‌سازی جهانی[۲۳] تابع هدف استفاده می‌کند. در ابتدا، شبکه را به بخش‌های تصادفی تجزیه می‌کند. بهینه‌سازی بر اساس حرکات محلی[۲۴] و جهانی است. حرکتهای محلی، یک گره را به طور

[13] Hierarchical clustering
[14] Partitional clustering
[15] Spectral clustering
[16] Divisive algorithms
[17] edge-betweenness score
[18] Modularity Optimization Based Techniques
[19] Modularity
[20] Greedy techniques
[21] Clauset-Newman-Moore
[22] Simulated Annealing
[23] Global
[24] Local

تصادفی از یک اجتماع به اجتماع دیگر بر اساس افزایش مقدار پیمانگی منتقل می‌کند. حرکتهای جهانی شامل تقسیم و ادغام اجتماعات است [۲۳].

ج. الگوریتم‌های تکاملی[۲۵]

الگوریتم های تکاملی یک کلاس از الگوریتم های بهینه‌سازی فرامکاشفه‌ای[۲۶] مبتنی بر هوش مصنوعی هستند. آنها به دلیل یادگیری موثر محلی و قابلیت های جستجوی جهانی معروف شده اند. الگوریتم‌های MLAMA-Net [۲۴] و COMBO [۲۵] نمونه‌هایی از تکنیک‌های خوشه‌بندی تکاملی هستند.

۳.۱.۳.۲. تکنیک های همپوشان[۲۷]

تراکم کلیک[۲۸] معروفترین روشی است که برای شناسایی جوامع همپوشان در شبکه ها استفاده می شود. ایده اصلی است که کلیک‌ها از یال‌های چگال داخل اجتماعات ساخته می‌شوند نه از یال های بین اجتماعات که پراکنده هستند. اجتماعاتی که از k کلیک ساخته شده با شند، به زیرگراف‌های کامل با k راس ا شاره دارند. الگوریتم‌هایCliquePercolation [۲۶] ، انتشار برچسب [۲۷]، SLPA[۲۹] [۲۸] نمونه‌هایی از تکنیک‌های همپوشان هستند.

۴.۱.۳.۲. الگوریتم‌های پویا[۳۰]

این بخش بر تکنیک‌های تشخیص اجتماع در شبکه‌های پویا، مانند توییتر، فیسبوک، لینکدین و غیره تمرکز می‌کند. این تکنیک‌ها تغییرات اجتماعات اعم از ایجاد یال های جدید و حذف آنها و همچنین اضافه یا حذف شدن گره‌ها را در طول به روزرسانی‌های زمانی در شبکه کنترل می‌کنند.

الف. پیاده‌روی تصادفی[۳۱]

پیاده‌روی تصادفی [۲۹] از محبوبترین الگوریتم‌هایی است که برای شناسایی جوامع مورد استفاده قرار گرفته‌است. در یک پیاده‌روی تصادفی، عابر شروع به قدم زدن در داخل یک اجتماع از یک گره می‌کند و در هر گام به طور تصادفی و یکنواخت به سمت گره مجاور انتخاب شده حرکت می‌کند. عابر به دلیل تراکم زیاد و مسیرهای متعدد مدت زیادی را در جوامع متراکم حرکت می‌کند. الگوریتم PageRank [۳۰] و Infomap [۳۱] نمونه‌هایی از محبوب‌ترین الگوریتم‌ها بر اساس پیاده‌روی‌های تصادفی هستند.

ب. گسترش اجتماع[۳۲]

یک گسترش اجتماع در یک شبکه مجموعه‌ای از گره‌ها است که با انتشار یک ویژگی، عمل یا اطلاعات مشابه در شبکه با هم گروه بندی می شوند. نمونه‌هایی از این دسته شامل الگوریتم‌های انتشار برچسب [۲۷] و رنگ آمیزی گره پویا[۳۳] [۳۲] است.

۲.۳.۲. اجتماعات همپوشان[۳۴] و ناهمپوشان[۳۵]

به دلیل ویژگی های مفید ساختار اجتماعات در شبکه های اجتماعی، محققان در سال های اخیر به نقش اجتماعات در مسئله بیشینه‌سازی تأثیر توجه روزافزونی داشته اند. در مقایسه با روش‌های سنتی حریصانه/ابتکاری، روش‌های مبتنی بر اجتماعات می‌توانند زمان محاسبه را کاهش داده و اثربخشی را افزایش دهند [۳۳]. روش‌های تشخیص اجتماعات به دو دسته تقسیم می شوند: تشخیص اجتماعات همپوشان و تشخیص اجتماعات بدون همپوشانی.

اجتماعات بدون همپوشانی: بعضی تحقیقات مبتنی بر اجتماعات هنگام تقسیم بندی یک گراف به اجتماعات، نقش گره های همپوشان (گره‌هایی که متعلق به بیش از یک اجتماع هستند) را در نظر نمی‌گیرند، یعنی آنها معمولاً تأثیر یک گره را صرفاً در محدوده اجتماع خودش ارزیابی می‌کنند. در واقع گره های همپوشان پیوندهایی هستند که جوامع مختلف را به هم متصل کرده و مسیرهای گسترش تأثیر بین آنها را ایجاد می‌کنند. نادیده گرفتن این گره‌های همپوشان، کیفیت مجموعه گره‌های تأثیرگذار انتخابی و میزان گسترش تأثیر را کاهش می‌دهد. از طرف دیگر از آنجایی که اجتماعات بدست آمده در این روش‌ها با یکدیگر همپوشانی ندارند، در نتیجه اندازه اجتماعات خروجی کوچکتر شده و باعث افزایش سرعت اجرای الگوریتم بیشینه‌سازی تأثیر خواهد شد.

اجتماعات همپوشان: در دنیای واقعی افراد ممکن است همزمان در چند اجتماع یک شبکه اجتماعی عضویت داشته باشند. به همین دلیل الگوریتم‌هایی پیشنهاد شدند که یک شبکه اجتماعی را به اجتماعات همپوشان بخش‌بندی کنند. از آنجا که گره‌های همپوشان اجتماعات مختلف را به هم متصل می کنند، هنگام فرایند گسترش تأثیر، اثربخشی الگوریتم را بهبود می بخشند [۳۴]. ولی از آنجا که اجتماعات با یکدیگر همپوشانی دارند، ممکن است اندازه اجتماعات بدست‌آمده بزرگ شود که باعث کاهش کارایی زمانی الگوریتم انتخاب گره‌های تأثیرگذار خواهدشد.

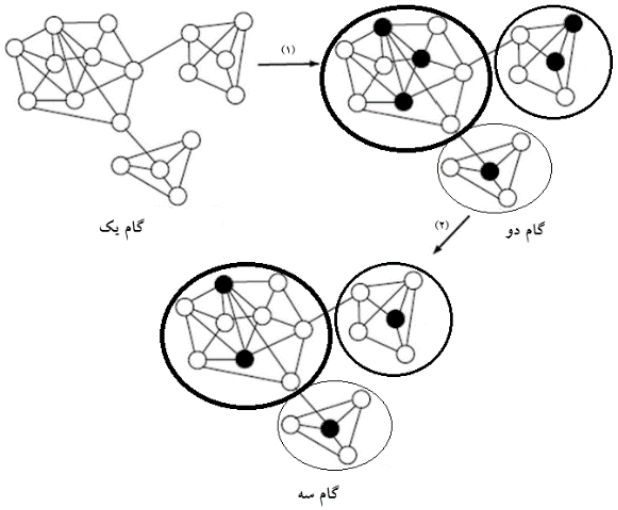
بطورکلی در رویکردهای بیشینه‌سازی تأثیر مبتنی بر ساختار اجتماعات، انتخاب بین الگوریتم تشخیص اجتماعات همپوشان و ناهمپوشان منجر به یک تبادل بین کارایی زمانی و اثر بخشی الگوریتم بیشینه‌سازی تأثیر می‌شود.

۳.۳.۲. شناسایی اجتماعات و مسئله بیشینه‌سازی تاثیر

مقوله شناسایی اجتماعات برای یافتن تأثیرگذارترین گره‌ها در شبکه‌های اجتماعی از دو جنبه حائز اهمیت است: ۱. با افزایش

[25] Evolutionary algorithms
[26] Metaheuristic
[27] Overlapping techniques
[28] Clique percolation
[29] Speaker-listener Label Propagation Algorithm
[30] Dynamic Algorithms
[31] Random walk
[32] Diffusion Community
[33] dynamic node colouring
[34] Overlapping
[35] Non- overlapping

اندازه شبکه های اجتماعی آنلاین، هزینه محاسباتی انتخاب تأثیرگذارترین گره‌ها به‌طور تصاعدی رشد می‌کند. در این شرایط، تقسیم شبکه به اجتماعات نقش اساسی در بهبود کارایی زمانی این مسئله دارد. ۲. در دنیای واقعی، اطلاعات در بین افراد داخل اجتماع سریع‌تر و مؤثرتر منتشر می‌شود؛ بنابراین، با کشف اجتماعات و سپس انتشار اطلاعات در داخل آن‌ها عملکرد بهتری از نظر میزان گسترش تأثیر حاصل خواهد شد. ازاین‌رو، تعدادی از محققان از ساختارهای اجتماعات برای افزایش کارایی زمانی و اثربخشی در مسئله‌ی بیشینه‌سازی تأثیر استفاده کرده‌اند که در بخش ۳،۳، مطالعات برجسته آن‌ها مرور خواهد شد. در شکل ۵، مثالی از انتخاب گره‌های تأثیرگذار یک شبکه ساده به کمک روش شناسایی اجتماعات نشان‌داده شده‌است. در گام ۱ شبکه به سه اجتماع تقسیم‌بندی خواهد شد. سپس در گام ۲، گره‌های تأثیرگذار هر اجتماع بدست خواهد آمد و مجموعه گره‌های نامزد را تشکیل خواهند داد. در گام آخر تأثیرگذارترین گره‌های شبکه صرفاً از میان مجموعه گره‌های نامزد انتخاب خواهند شد. در این شکل دایره های سیاه رنگ معرف گره‌های تأثیرگذار است.



شکل ۵. شناسایی اجتماعات و بیشینه‌سازی تاثیر

۴.۲. مجموعه داده‌ها[36]

در این بخش برخی از مجموعه داده‌های معروف و پرکاربرد که برای تحقیق در حوزه بیشینه سازی تاثیر در شبکه‌های اجتماعی مناسب هستند معرفی شده و ویژگیهای آنها در جدول ۲ لیست شده است.

NetHEPT یک شبکه از نوع همکاری در گروه آرشیو نظریه فیزیک انرژی بالا[37] است. در این شبکه کوچک، اگر نویسنده‌ای حداقل در یک مقاله با نویسنده دیگر همکاری داشته باشد، بین آنها یک یال ایجاد می‌شود.

Epinions یک مجموعه داده جهت‌دار با اندازه متوسط از یک شبکه اجتماعی آنلاین است که در آن شخصی به شخص دیگر اعتماد[38] می‌کند. در وب سایت این مجموعه داده، همه اعضا می‌توانند نظرات خود را در مورد محصولات ارسال کنند و سایر کاربران می‌توانند انتخاب کنند که به نظرات اعتماد کنند یا نه.

Digg-friend یک گراف دوستی جهت‌دار است. در این شبکه اجتماعی، کاربران با یکدیگر رابطه دوستی دارند.

Amazon این شبکه کالاهایی در آمازون است که بصورت زیر توسط آمازون ذکر شده است "مشتری که کالای A را خریداری کرده است کالای B را نیز خریداری کرده است". گره‌ها نمایانگر محصولات هستند و اگر محصول A اغلب با محصول B خریداری شود، گراف شامل یک یال جهت‌دار از محصول A به محصول B است.

Web-Google این مجموعه داده از گراف وب سایت شرکت گوگل گرفته شده است. در این مجموعه داده بزرگ، گره‌ها نشان‌دهنده صفحات هستند و یال‌های جهت‌دار نشان‌دهنده پیوندهای بین آنها است.

Youtube یک وبسایت اشتراک ویدیو هست که شامل شبکه اجتماعی کاربران یوتیوب و ارتباطات دوستی آنهاست.

Wikitopcats یک گراف از صفحات وب سایت ویکی‌پدیا است. این شبکه ابتدا با در اختیار گرفتن بزرگترین مولفه قویاً همبند از وب سایت ویکی‌پدیا ایجاد می‌شود. سپس به صفحات گروههای برتر این مولفه قویاً همبند (آنهایی که دارای حداقل ۱۰۰ صفحه هستند) محدود می شود. در نهایت این مجموعه داده شامل بزرگترین مولفه قویاً همبند از این گراف محدود شده خواهد بود.

Pokec یک شبکه اجتماعی دوستی است. این شبکه یکی از مجموعه داده‌های بسیار بزرگ است که شامل ۱،۶ میلیون گره و ۳۰،۱ میلیون یال جهت دار است.

Flickr شبکه اجتماعی کاربران فلیکر و ارتباطات دوستی آنهاست. این شبکه با ایجاد یال‌هایی بین تصاویر به اشتراک گذاشته شده ساخته شده است.

LiveJournal یک شبکه اجتماعی رایگان نمونه‌گیری شده است. گره‌ها کاربران این شبکه هستند و یال‌های جهتدار نشان دهنده‌ی رابطه دوستی میان کاربرانند. این مجموعه داده شامل حدود ۴،۸ میلیون گره، با ۳۴،۵ میلیون جفت روابط دوستی (۶۸،۹ میلیون یال جهتدار) است.

۳. سازمان‌دهی الگوریتم‌های بیشینه‌سازی تأثیر

مسئله‌ی بیشینه‌سازی تأثیر ابتدا توسط دومینگوس و ریچاردسون در سال ۲۰۰۱ به‌عنوان یک مسئله احتمالی مطرح [۳] و سپس در سال ۲۰۰۳ توسط کمپ و همکاران [۴] به‌عنوان یک مسئله بهینه‌سازی گسسته فرموله شد. بعد از آن‌ها، نویسندگان به بررسی

[36] Datasets
[37] Arxiv High Energy Physics Theory
[38] who-trust-whom

جدول۲. مجموعه داده‌های پرکاربرد در حوزه بیشینه‌سازی تاثیر در شبکه‌های اجتماعی

مجموعه داده	تعداد گره‌ها	تعداد یال‌ها	تعداد مثلثی‌ها	میانگین درجه	قطر	ضریب جینی	ضریب خوشه‌بندی	نوع شبکه
NetHept	۱۵ هزار	۶۳ هزار	۲۶۳ هزار	۲۵,۴	۱۵	۰,۵۷	۰,۱۲	همکاری
Epinions	۷۶ هزار	۵۰۹ هزار	۱,۶ میلیون	۱۳,۴	۱۵	۰,۸۱	۰,۰۷	شبکه اجتماعی
Digg-friend	۲۸۰ هزار	۱,۷ میلیون	۱۴,۲ میلیون	۱۲,۴	۱۸	۰,۸۵	۰,۰۶	شبکه دوستی
Amazon	۴۰۴ هزار	۳,۴ میلیون	۳,۴ میلیون	۱۶,۸	۲۵	۰,۳۳	۰,۱۷	فروش محصول
Web-Google	۸۷۶ هزار	۵,۱ میلیون	۱۳,۴ میلیون	۱۱,۷	۲۴	۰,۶	۰,۰۶	صفحات وب
Youtube	۳,۲ میلیون	۹,۴ میلیون	۱۲ میلیون	۱۱,۶	۳۱	۰,۷۳	۰,۰۱	شبکه اجتماعی
Wikitopcats	۱,۸ میلیون	۲۸,۵ میلیون	۵۲,۱ میلیون	۲۱,۴	۹	۰,۷۳	۰,۲۷	صفحات وب
Pokec	۱,۶ میلیون	۳۰,۶ میلیون	۳۲,۶ میلیون	۳۷,۵	۱۴	۰,۶۲	۰,۰۵	شبکه دوستی
Flickr	۲,۳ میلیون	۳۳,۲ میلیون	۸۳۸ میلیون	۲۸,۸	۲۳	۰,۸۸	۰,۱۱	شبکه دوستی
LiveJournal	۴,۸ میلیون	۶۹,۱ میلیون	۲۸۶ میلیون	۲۸,۳	۲۰	۰,۷۲	۰,۱۲	شبکه دوستی

جنبه سختی مسئله و ارائه راه‌حل‌هایی پرداختند. ازآنجاکه یافتن راه‌حل بهینه در مسائل NP-Hard بسیار بغرنج است، رویکردهای تقریبی جایگزینی‌های بهتری هستند. به‌طورکلی در مسئله‌ی بیشینه‌سازی تأثیر، انتخاب افراد تأثیرگذار به‌وسیله روش‌های رتبه‌بندی مرکزیت درجه[۳۹]، حریصانه[۴۰] و مکاشفه‌ای[۴۱] انجام می‌شود [۱۰]. روش رتبه‌بندی درجه، افراد تأثیرگذار را بر اساس درجه آن‌ها انتخاب می‌کند. به‌عبارت‌دیگر، همه گره‌های شبکه از بالاترین درجه به پایین‌ترین درجه به ترتیب نزولی رتبه‌بندی می‌شوند و افراد تأثیرگذار از بالای لیست انتخاب می‌شوند. الگوریتم حریصانه با انتخاب بهینه محلی در هر مرحله، افراد تأثیرگذار را انتخاب می‌کند. تکنیک‌های مکاشفه‌ای هر روش به‌کارگرفته شده برای یادگیری، کشف یا حل مسائل هستند. در این نوع روش‌ها علی‌رغم حرکت به سمت راه‌حل بهینه، هیچ تضمینی برای کیفیت پاسخ نهایی وجود ندارد.

در این مقاله الگوریتم‌های ارائه شده در مسئله‌ی بیشینه سازی تأثیر در سه دسته طبقه‌بندی می‌شوند: حریصانه، مکاشفه‌ای و مبتنی بر شناسایی اجتماعات (شکل ۶)

الگوریتم‌های حریصانه معمولاً از چارچوب حریصانه برای حل مسائل بهینه‌سازی استفاده می‌کنند. ازاین‌رو، این الگوریتم‌ها با در نظر گرفتن شرایط مسئله همیشه بهترین راه‌حل را انتخاب می‌کنند به امید آنکه ادامه این روند منجر به بهینه‌سازی شود. در برخی مسائل سخت که الگوریتم‌های حریصانه و کلاسیک زمان اجرای بسیار طولانی دارند، الگوریتم‌های مکاشفه‌ای مورد استفاده قرار می‌گیرند. الگوریتم‌های مکاشفه‌ای دسته‌ای از الگوریتم‌های تقریبی هستند که می‌توانند در مدت‌زمان کوتاهی راه‌حل‌های تقریبی نزدیک به راه‌حل بهینه پیدا کنند. در الگوریتم های مبتنی بر اجتماعات که نوعی استراتژی تقسیم و غیال هستند، ابتدا کل شبکه اجتماعی به اجتماعات کوچک‌تر تقسیم می شود و سپس یکی از الگوریتم‌های حریصانه یا مکاشفه‌ای یا ترکیبی از آن‌ها برای یافتن تأثیرگذارترین افراد مورد استفاده قرار خواهند گرفت. در بخش های بعدی، رویکردهای موجود در مسئله‌ی بیشینه‌سازی تأثیر بررسی می‌شوند.

شکل ۶. طبقه بندی الگوریتم‌های ارائه‌شده در بیشینه‌سازی تأثیر

۱.۳. رویکردهای حریصانه

کمپ و همکاران در سال ۲۰۰۳ مسئله‌ی بیشینه‌سازی تأثیر را به‌عنوان یک مسئله بهینه‌سازی ترکیبی تحت دو مدل انتشار متداول آبشاری مستقل و آستانه‌ی خطی فرموله کردند. آن‌ها ثابت کردند که مسئله‌ی بیشینه‌سازی تأثیر تحت این دو مدل NP-Hard است و یک الگوریتم حریصانه با تقریب ۶۳٪ از پاسخ بهینه ارائه دادند [۴]. در هر تکرار از اجرای الگوریتم، یک گره با بیشترین سود حاشیه‌ای[۴۲] به مجموعه سبد S اضافه می شود تا اینکه تعداد گره‌های مجموعه S تأمین شود. محاسبه دقیق میزان گسترش تأثیر تحت این دو مدل NP-Hard# است؛ بنابراین این الگوریتم حریصانه مجبور است که شبیه‌سازی‌های مونت‌کارلو[۴۳] را حدود ۱۰ هزار بار تکرار کند تا به یک تقریب بهتر برسد. این سربار محاسباتی زیاد، باعث کاهش کارایی زمان در شبکه‌های اجتماعی بزرگ خواهد

[39] Node Centrality Ranking
[40] Greedy
[41] Heuristic
[42] Marginal gain
[43] Monte Carlo Simulations

شد [۹]. تلاش‌های زیادی برای مقابله با چالش کارایی زمان این الگوریتم حریصانه انجام شده است.

لسکوویچ و همکاران در سال ۲۰۰۷ از خاصیت زیرپیمانه‌ای‌بودن تابع گسترش تأثیر استفاده کردند و الگوریتمی کارآمد به نام CELF[۴۴] پیشنهاد دادند [۳۵]. ایده آن‌ها این است که سود حاشیه‌ای هر گره در دور فعلی نمی‌تواند از سود حاشیه‌ای‌اش در دورهای قبلی بیشتر باشد. الگوریتم بدین صورت است که در تکرار اول، گسترش تأثیر همه گره‌ها محاسبه و سپس گره‌ها به ترتیب نزولی میزان گسترش تأثیرشان در یک صف مرتب می‌شوند. اولین گره صف برداشته شده و به مجموعه سبد S اضافه می‌شود. در بقیه تکرارها، مقدار گسترش تأثیر فقط برای گره جلویی صف محاسبه می‌شود. پس از مرتب کردن مجدد صف، اگر اولین گره در صف همچنان در جلوی صف باقی بماند، آنگاه آن گره دارای بیشترین سود حاشیه‌ای است که باید از صف حذف شده و به مجموعه سبد S اضافه شود. الگوریتم CELF با اجتناب از محاسبه میزان تأثیر در بسیاری از گره‌ها، پیچیدگی زمانی الگوریتم حریصانه‌ی کمپ را کاهش می‌دهد و سرعت آن را ۷۰۰ برابر بهبود می‌بخشد. زمان بالای اجرای دور اول الگوریتم چالش اصلی CELF است؛ زیرا مانند الگوریتم حریصانه‌ی کمپ، باید تأثیر همه گره‌ها در دور اول محاسبه شود.

الگوریتم CELF++ که توسط گویال و همکاران در سال ۲۰۱۱ معرفی شد [۳۶]، با اجتناب از شبیه‌سازی‌های غیرضروری مونت‌کارلو به ۳۰-۵۰٪ بهبود در زمان اجرا نسبت به الگوریتم CELF دست می‌یابد. در عمل و در مواجهه با شبکه های بزرگتر، افزایش سرعت قابل توجهی نسبت به CELF مشاهده نشد [۳۷].

ژو و همکاران در سال ۲۰۱۵ روشی به نام UBLF[۴۵] را برای به دست آوردن کران بالای تابع گسترش تأثیر با استفاده از ماتریس‌ها پیشنهاد کند [۳۸]. نتایج گزارش شده حاکی از کاهش بیش از ۹۵ درصدی شبیه‌سازی مونت‌کارلو و البته بهبود کارایی زمانی ۲ تا ۱۰ برابر سریع‌تر از CELF بود. این روش همانند الگوریتم‌های CELF و CELF++ علیرغم تولید مجموعه سبد با کیفیت بالا، پیچیدگی بدترین زمان اجرا را بهبود نمی‌دهد. همچنین این الگوریتم فقط برای دو مدل انتشار آبشاری مستقل و آستانه خطی طراحی شده است.

الگوریتم NewGreedy نسخه بهبودیافته الگوریتم حریصانه‌ی کمپ است که توسط چن و همکاران در سال ۲۰۰۹ ارائه شد [۳۹]. ایده اصلی ایجاد گراف کوچک‌تر با حذف یال‌هایی است که نقشی در روند انتشار اطلاعات ندارند. با توجه به این ایده، فضای جستجو کاهش می‌یابد و سرعت افزایش می‌یابد. یکی از معضلات الگوریتم NewGreedy زمان طولانی ایجاد گراف کوچکتر در هر مرحله از تکرار اصلی الگوریتم حریصانه است که منجر به ناکارآمدی در گراف های بزرگ می‌شود. برای دستیابی به نتایج بهتر، نویسندگان الگوریتم MixedGreedy را پیشنهاد کردند که الگوریتم NewGreedy را در اولین تکرار و الگوریتم CELF را در بقیه تکرارها اجرا می‌کند. در عمل این الگوریتم نیز نسبت به NewGreedy افزایش سرعت قابل توجهی نداشت [۳۷].

در الگوریتم staticGreedy که توسط چنگ و همکاران در سال ۲۰۱۳ معرفی شد [۴۰]، در ابتدا چند تصویر لحظه‌ای[۴۶] از شبیه‌سازی مونت‌کارلو گرفته می‌شود. در مرحله بعد، k گره با بیشترین سود حاشیه‌ای در بین همه تصاویر نمونه‌برداری شده، به‌عنوان گره‌های مجموعه سبد انتخاب می‌شوند. به دلیل پیچیدگی زمانی اجتناب‌ناپذیر staticGreedy در بدترین حالت $O\left(\varepsilon^{-2}kmn^2\ log\binom{n}{k}\right)$، از یک روش هرس برای کاهش زمان اجرای آن در الگوریتم staticGreedyDU استفاده می‌شود.

الگوریتم Pruned-MC[۴۷] که توسط اوزاکا و همکارانش در سال ۲۰۱۴ معرفی شد، یک روش نمونه‌گیری مبتنی بر تصویر لحظه‌ای است [۴۱]. با استفاده از یک ساختار شاخص‌گذاری روی تصاویر لحظه‌ای به همراه تکنیک هرس کردن، کارایی الگوریتم staticGreedyDU بهبود یافته و زمان اجرا کاهش می‌یابد.

علیرغم بهبود اثربخشی و استخراج گره‌های تاثیرگذار با کیفیت بالا در هر سه الگوریتم staticGreedy، staticGreedyDU و Pruned-MC، پیچیدگی زمانی بدترین حالت آنها بسیار پرهزینه است $O\left(\varepsilon^{-2}kmn^2\ log\binom{n}{k}\right)$ به‌خصوص در مواجهه با گراف‌هایی با میلیون‌ها گره و میلیاردها یال.

تانگ و همکاران در سال ۲۰۱۴ الگوریتم دو فاز TIM[۴۸] را در حوزه مسئله‌ی بیشینه‌سازی تأثیر معرفی کردند [۴۲]. در فاز اول به‌منظور تخمین پارامتر φ، یک کران پایین روی بیشترین مقدار گسترش تأثیر به دست می‌آید. در مرحله دوم، تعداد φ عدد دستیابی‌معکوس[۴۹] از گراف انتخاب می‌شود. سرانجام، k گره که بیشترین تعداد دستیابی‌معکوس را پوشش می‌دهد به‌عنوان گره‌های نهایی مجموعه سبد انتخاب می‌شوند. الگوریتم TIM+ نسخه بهبودیافته TIM است که برای سرعت بخشیدن به روش تخمین پارامتر φ پیشنهاد شده است. TIM دو برابر سریع‌تر از CELF++ و TIM+ دو برابر سریع‌تر از TIM است.

الگوریتم IMM[۵۰] که توسط تانگ و همکاران در سال ۲۰۱۵ معرفی شد [۴۳]، یک روش دوفازی مشابه الگوریتم TIM است که از تکنیک‌های پیشرفته تخمین مانند روش مارتینگل[۵۱] استفاده می‌کند. مصرف زیاد حافظه یکی از مهم‌ترین اشکالات TIM،

[44] Cost Effective Lazy Forward
[45] Upper Bound based Lazy Forward
[46] Snapshot
[47] Pruned Monte Carlo
[48] Two phases Influence Maximization
[49] Rreverse Reachability (RR)
[50] Influence Maximization via Martingales
51 Martingales

+TIM و IMM است. زیرا اولاً آنها برای اطمینان از کران پایین مجبور به تولید تعداد زیادی نمونه (دستیابی‌معکوس) هستند. ثانیاً می‌بایست همه‌ی نمونه‌ها را بصورت همزمان در حافظه نگهداری کنند تا روش حریصانه k گره تاثیرگذار را انتخاب کند.

بیشتر الگوریتم‌های ذکرشده در این بخش، باوجود ارائه ضمانت تئوری نمی‌توانند شبکه‌های اجتماعی در مقیاس بزرگ را کنترل کنند.

۲.۳. رویکردهای مکاشفه‌ای

با توجه به پیچیدگی زمانی بالای الگوریتم‌های حریصانه فوق و ناکارآمد بودن آن‌ها در مواجهه با شبکه‌های اجتماعی بسیار بزرگ، طیف گسترده‌ای از الگوریتم‌های مکاشفه‌ای برای حل معضل کارایی در مسئله‌ی بیشینه‌سازی تأثیر ارائه شده است. یکی از متداول‌ترین ایده‌های مکاشفه‌ای، انتخاب گره‌های تأثیرگذار بر اساس درجه آن‌ها است. مرکزیت درجه[۵۲]، مرکزیت نزدیکی[۵۳]، مرکزیت بینابینی[۵۴]، مرکزیت بردار ویژه[۵۵]، رتبه‌بندی صفحات[۵۶] و HITS[۵۷] از معروف ترین رویکردهای مکاشفه‌ای بر اساس درجه‌ی گره‌ها هستند. دو عیب اصلی الگوریتم‌های مبتنی بر درجه‌ی گره به شرح زیر است:

۱- درجه‌ی گره و انتشار اطلاعات دو مفهوم متفاوت و مجزا هستند.

۲. ازآنجاکه این دسته از الگوریتم‌های مکاشفه‌ای، همپوشانی تأثیر بین سبدهای مختلف را در نظر نمی‌گیرند، پدیده تخمین بیش‌ازحد[۵۸] رخ می‌دهد.

چن و همکاران در سال ۲۰۰۹ الگوریتم DegreeDiscount را برای به‌حساب آوردن همپوشانی تأثیر گره‌ها پیشنهاد دادند [۳۹]. در این الگوریتم پس از انتخاب یک گره با بیشترین درجه به‌عنوان گره تأثیرگذار، نمرات تأثیر همسایگان آن گره یک واحد کاهش می‌یابد. عیب اصلی این رویکرد این است که فقط تأثیر همسایگان مستقیم را در نظر می‌گیرد و تأثیر مسیرها و همسایگان غیرمستقیم را نادیده می‌گیرد.

ازآنجاکه محاسبه میزان تأثیر در گراف جهت دار بدون دور (DAG[۵۹]) دقیق و کارآمد است، چن و همکارانش در سال ۲۰۱۰ مدل LDAG[۶۰] را پیشنهاد کردند [۴۴]. در این مدل، برای هر گره DAG های محلی ساخته می‌شود. سپس، الگوریتم حریصانه‌ی کمپ برای یافتن گره های تأثیر گذار بر روی هر DAG اعمال می‌شود. در شبکه‌های اجتماعی بزرگ، روند ساخت DAG ها ازنظر مصرف حافظه و زمان اجرا بسیار ناکارآمد است.

SP1M نام الگوریتمی است که توسط کیمورا و همکاران در سال ۲۰۱۰ معرفی شده است [۴۵]. ایده اصلی آن در نظر گرفتن کوتاه‌ترین مسیر از گره u به گره v برای فعال سازی است. الگوریتم SP1M[۶۱] نیازی به اجرای سنگین شبیه‌سازی‌های مونت‌کارلو ندارد. البته از آنجایی که SP1M برای محاسبه میزان گسترش تأثیر صرفاً طول مسیر را درنظر گرفته و احتمالات را نادیده می‌گیرد قادر به تولید جواب تقریبی خوبی در همه‌ی گراف‌ها نیست.

با الهام از ایده SP1M، در سال ۲۰۱۰ چن و همکاران مدل‌های MIA[۶۲] و PMIA[۶۳] را پیشنهاد دادند [۴۶]. مدل MIA دامنه پخش اطلاعات را از کل گراف به یک ساختار درختی محلی، محدود می‌کند. ازآنجاکه محاسبه تأثیر در یک درخت با دقت و کارایی بیشتری انجام می شود، میزان گسترش تأثیر سریع‌تر تخمین زده می‌شود؛ بنابراین از شبیه‌سازی گران‌قیمت مونت‌کارلو اجتناب خواهد شد. به‌منظور اینکه مجموعه سبد فعلی تأثیر سبدهای بعدی را مسدود نکند، مدل PMIA تأثیر درون ساختار درخت را پس از افزودن گره به مجموعه سبد اصلاح می‌کند. اشکال اصلی این الگوریتم ها معضل مقیاس پذیری در گراف های متراکم است. همچنین در جایی که اندازه ساختارهای درختان محلی خیلی بزرگ شوند، دقت تخمین میزان گسترش تأثیر کم خواهد شد.

ایده الگوریتم SimplePath که توسط گویال و همکاران در سال ۲۰۱۱ پیشنهاد شده است، محاسبه تأثیر مجموعه سبد تحت مدل آستانه خطی است [۴۷]. این کار با شمارش تمام مسیرهای ساده از هر گره در مجموعه سبد انجام می‌شود. ازآنجاکه شمارش تمام مسیرهای ساده در گراف #P-Hard است، از تکنیک هرس‌کردن استفاده می‌شود. این الگوریتم، از بهینه‌سازی CELF[35] برای کارایی بیشتر بهره می برد. این الگوریتم بر خلاف الگوریتم LDAG[44]، میزان گسترش تأثیر را بدون شمارش همه‌ی مسیرهای ساده روی گراف اصلی تخمین می‌زند. از این رو زمان و حافظه بهتری نسبت به الگوریتم LDAG دارد [۴۸]. نقطه ضعف الگوریتم SimplePath استفاده از خصوصیات مدل انتشار آستانه خطی برای تخمین میزان تأثیر است، از این رو قابل تعمیم به سایر مدل‌های انتشار نیست.

در کل روش‌های مبتنی‌بر شمارش مسیر مانند LDAG، SP1M، MIA، PMIA و SimPath در مواردی که تعداد مسیرهای تأثیر گره‌ها یا محدوده تأثیر گره‌ها بزرگ باشد، نمی‌توانند توازنی بین کارایی و دقت برقرار کنند.

نارایانام و همکاران در سال ۲۰۱۱ الگوریتم SPIN[۶۴] را معرفی کردند که در آن گره‌ها نقش بازیکنان در بازی‌های ائتلافی را ایفا

52 Degree Centrality
53 Closeness Centrality
54 Betweenness Centrality
55 Eigenvector Centrality
56 PageRank
57 Hyperlink-Induced Topic Search
58 overestimation
59 Directed Acyclic Graph
60 Local Directed Acyclic Graph
61 shortest-path based heuristic algorithm
62 Maximum Influence Arborescence
63 Prefix excluding MIA
64 ShaPley value based Influential Nodes

می‌کنند [۴۹]. فرآیند انتشار اطلاعات در SPIN مشابه آنچه که در اتحاد ائتلاف اتفاق می‌افتد، انجام می‌شود. در این مدل، گره‌ها با توجه به مقدار شپلی آن‌ها رتبه‌بندی می‌شوند و سپس k تا از بالاترین آن‌ها به‌عنوان گره‌های مجموعه سبد انتخاب می شوند. این الگوریتم علیرغم داشتن زمان اجرای خوب نمی‌تواند هیچ ضمانتی روی کران بدترین حالت فراهم کند. همچنین از آنجا که این الگوریتم صرفاً برای مدل انتشار آستانه‌ی خطی طراحی شده است، برای سایر مدل های انتشار قابل استفاده نیست.

الگوریتم IRIE[۶۵] تعمیم الگوریتم PageRank[30] است که توسط یونگ و همکاران در سال ۲۰۱۲ پیشنهاد شده است [۵۰]. الگوریتم IRIE از دو عملیات اصلی تشکیل شده است: رتبه‌بندی تأثیر(IR) و تخمین تأثیر(IE). روش کار بدین‌صورت است که یک دستگاه معادلاتی شامل n معادله خطی با n متغیر به‌منظور به دست آوردن لیست رتبه‌بندی تأثیرات گره‌ها ساخته و حل می‌شود. اگر k گره بالای لیست رتبه‌بندی تأثیر به‌دست‌آمده به‌عنوان مجموعه سبد نهایی انتخاب شوند، پدیده همپوشانی تأثیر[۶۶] رخ خواهد داد. برای جلوگیری از این امر، نویسندگان تکنیک تخمین تأثیر را برای محاسبه و تفریق تأثیر اضافی مجموعه سبد بر گره‌های دیگر معرفی کردند. برمبنای پیشنهاد نویسندگان، استفاده از الگوریتم‌های دیگر مانند Greedy یا PMIA به منظور محاسبه تأثیر اضافی مجموعه سبد باعث افزایش زمان اجرای این الگوریتم خواهد شد.

لئو و همکاران در سال ۲۰۱۴ یک چارچوب حریصانه به نام Group-PR ارائه کردند که بسطی از الگوریتم PageRank است [۵۱]. این الگوریتم به‌جای استفاده از یک گره، از مجموعه‌ای از گره‌ها برای محاسبه میزان گسترش تأثیر استفاده می‌کند. در مرحله اول، امتیاز PageRank هر گره محاسبه می شود. در مرحله دوم، گرهای با بیشترین سود حاشیه‌ای به مجموعه سبد اضافه می شود. این فرآیند با اضافه شدن k گره به مجموعه سبد خاتمه می‌یابد.

چارچوب IMRank[۶۷] یک استراتژی مبتنی بر رتبه‌بندی است که توسط چنگ و همکاران در سال ۲۰۱۴ ارائه شده است [۵۲]. ایده این چارچوب تکراری، کشف رتبه‌بندی خود-سازگار[۶۸] از هر رتبه‌بندی اولیه است. پس از کشف لیست رتبه‌بندی خود- سازگار، k گره بالایی آن لیست به‌عنوان اعضای تأثیرگذارتر به مجموعه سبد اضافه خواهند شد.

روش های مبتنی بر رتبه‌بندی مانند IMRank، PageRank، IRIE و Group-PR سربار محاسباتی را بسیار کاهش می‌دهند. اما از آنجا که نحوه سازوکار این الگوریتم ها با ماهیت مسئله‌ی بیشینه سازی‌تأثیر متفاوت است، گسترش تأثیر تخمین زده شده می‌تواند با گسترش تأثیر واقعی گره‌ها تحت مدل‌های انتشار اختلاف جدی داشته باشد. بعلاوه خصوصیات مدل‌های انتشار در این دسته از الگوریتم‌ها نادیده گرفته شده‌است.

IPA[۶۹] نام الگوریتم مبتنی بر مسیر است که توسط کیم و همکاران در سال ۲۰۱۳ ارائه شده است. فرض این الگوریتم بر این است که مسیرهای انتشار از یکدیگر مستقل هستند؛ بنابراین برای محاسبه گسترش تأثیر، از یک روش موازی سازی با کارایی بیشتر استفاده می‌کند. اشکالات الگوریتم PMIA [۴۶] به این الگوریتم نیز وارد است.

گالوترا و همکاران در سال ۲۰۱۶ مدلی به نام EaSyIM ارائه دادند [۵۳]. ایده اصلی آن شمارش مسیرهای ساده برای تخمین تأثیر هر گره بود. این مدل از روش IRIE[50] برای تخمین تأثیر سراسری گره‌ها و دستیابی به عملکرد بهتر استفاده می‌کند. نتایج تجربی گزارش شده در این تحقیق نشان‌دهنده گسترش تأثیر بهتر و زمان اجرای کمتر در مقایسه با الگوریتم‌های دیگر مانند CELF++[36] و TIM+[42] است.

گورسوی و همکاران در سال ۲۰۱۸ مسئله‌ی بیشینه‌سازی تاثیرهدفمند و بودجه دار را تحت مدل آستانه خطی قطعی تعریف و مدل مقیاس پذیر TABU–PG[۷۰] را طراحی کردند [۵۴]. این مدل یک الگوریتم تکراری و حریصانه بر مبنای پیشنهادات بالقوه در زمان انتخاب گره‌های سبد است. این مدل با بهره‌گیری از تکنیک‌های مقیاس پذیری سعی در کاهش زمان اجرا با محدود کردن تعداد گره های نامزد یا با انتخالب چند گره در یک مرتبه دارد. این موارد منجر به تبادلی بین زمان اجرا و میزان اثربخشی الگوریتم خواهد شد.

TIFIM[۷۱] یک چارچوب تکراری دو مرحله‌ای برای بیشینه‌سازی تأثیر است که توسط هی و همکاران در سال ۲۰۱۹ پیشنهاد شد [۵۵]. در مرحله اول به منظور حذف گره‌های کم‌اهمیت و کاهش پیچیدگی محاسبات، یک چارچوب تکراری به ترتیب نزولی برای انتخاب گره‌های نامزد پیشنهاد می‌شود. برای محاسبه سود گسترش هر گره، استراتژی تخصیص اولین و آخرین[۷۲] بر اساس نتایج آخرین تکرار و معیار اندازه گیری دو مرحله ای[۷۳] ارائه شده است. در مرحله دوم، مفهوم تسلط آپیکال[۷۴] را برای محاسبه پدیده همپوشانی گسترش در بین گره‌ها تعریف شده است. نویسندگان ادعا کردند که الگوریتم پیشنهادیشان قادر به تولید یک مجموعه سبد باکیفیت تر از الگوریتم IMM [۴۳] در زمانی نزدیک به الگوریتم PageRank هست.

65 Influence Ranking Influence Estimation
66 influence overlap
67 Influence Maximization Rank
68 Self-Consistent Ranking
69 Independent Path Algorithm
70 TArgeted and BUdgeted Potential Greedy
71 Two-stage Iterative Framework Influence Maximization
72 First-Last Allocating Strategy
73 two-hop measure
74 Apical Dominance

MLIM[۷۵] یک الگوریتم بر مبنای حداکثر احتمال ا ست که تو سط لیو و همکاران در سال ۲۰۲۰ برای مدل انتشار آبشاری مستقل پیشنهاد شده است [۵۶]. در ابتدا با استفاده از معیار حداکثر احتمال مقدار L برای هر راس محاسبه خواهد شد. سپس یک الگوریتم حریصانه پیشنهاد شده است تا گره‌ها را به ترتیب کوچترین مقدار L به مجموعه سبد ا ضافه کند. نتایج گزارش شده نشان داده است که الگوریتم پیشنهادی توانسته اثربخشی بیشتری را در حین به دست آوردن زمان کمتر در مقایسه با الگوریتم‌های نه چندان مطرح ایجاد کند.

ونگ و همکاران در سال ۲۰۲۱، مساله بیشینه‌سازی نفوذ نامتجانس (DIM)[۷۶]را مورد بررسی قرار دادند که هدف آن انتخاب K گره به گونه‌ای است که هم تعداد گره‌های فعال شده و هم تنوع گره‌های فعال شده را بتوان به حداکثر ر ساند [۵۷]. آنها با طراحی یک تابع تنوع برای مدل کردن توزیع گره های فعال در میان همه جوامع، مساله DIM را با به حداکثر ر ساندن یک ترکیب خطی پارامتری از تنوع و تعداد گره فعال مدل کردند. همچنین برای مقابله با NP-Hardness، یک الگوریتم بیشینه‌سازی نفوذ متنوع کارآمد از طریق مارتینگل را پیشنهاد کردند که جواب تقریبی $1 - 1/e$ را باز می‌گرداند.

الگوریتم‌های مکا شفه‌ای ذکر شده در این بخش زمان اجرای بهتری نسبت به الگوریتم‌های حریصانه در بخش قبلی دارند. بااین‌وجود، این الگوریتم ها نمی توانند کران بدترین حا لت را فراهم کنند و ازلحاظ نظری ضعیف هستند. جدول ۳ الگوریتم‌های حریصانه و مکاشفه‌ای مطرح و برجسته درزمینه‌ی بیشینه‌سازی تأثیر در شبکه‌های اجتماعی را نشان می‌دهد.

۳.۳. رویکردهای مبتنی بر شناسایی اجتماعات

اجتماعات ساختارهای پنهان و یکی از مهم‌ترین خ صو صیات ذاتی شبکه‌های اجتماعی هستند. همه رویکردهای حریصانه و مکاشفه‌ای ذکرشده در دو بخش قبلی عمدتاً بر ارتباطات بین گره‌ها و مکانیسم انتشار اطلاعات متمرکز هستند و نقش حیاتی ساختارهای اجتماعات در یک شبکه اجتماعی را نادیده گرفته‌اند. افزایش تصاعدی اندازه شبکه‌های اجتماعی و گسترش سریع‌تر و مؤثرتر اطلاعات در داخل اجتماعات، اصلی‌ترین عواملی هستند که توجه محققان را به استفاده از رویکردهای مبتنی بر اجتماعات جلب کرده ا ست. وا ضح ا ست که این ا ستراتژی تقسیم و غیال نقش مهمی در کارایی زمانی و اثربخشی مسئله‌ی بیشینه‌سازی تأثیر دارد.

اولین بار در سال ۲۰۱۰، وانگ و همکاران سعی کردند که مع ضل مقیاس‌پذیری در مسئله‌ی بیشینه‌سازی تأثیر را با بهره‌گیری از ساختارهای اجتماعات بهبود بخشند [۵۸]. آن‌ها الگوریتم CGA[۷۷] را پیشنهاد دادند که در ابتدا گراف را به اجتماعات کوچک‌تر تقسیم می‌کند. سپس با استفاده از برنامه‌نویسی پویا آن اجتماعی را که باید برای یافتن گره‌های تأثیرگذار بررسی شود، انتخاب می‌کند. نتایج تجربی آن‌ها نشان داد که CGA نمی‌تواند از عهده شبکه‌های اجتماعی در مقیاس بزرگ برآید.

مدل CIM[۷۸] که توسط چن و همکاران در سال ۲۰۱۴ پیشنهاد شد [۵۹]، با فرایند شناسایی اجتماعات با سیاست عدم همپوشانی اطلاعات آغاز می‌شود. سپس برای هر اجتماع، یک مجموعه سبد اولیه ساخته می شود. سرانجام گره‌های تأثیرگذار نهایی، از مجموعه سبدهای اولیه متنا سب با تعداد گره هر سبد، انتخاب خواهند شد. علیرغم کارایی زمانی خیلی‌خوب مدل CIM، فرض نادرست عدم همپو شانی اطلاعات در فرایند شنا سایی اجتماعات منجر به کاهش میزان گسترش تأثیر و دقت این مدل شده است.

با استفاده از مفهوم انطباق[۷۹]، لی و همکاران در سال ۲۰۱۵ الگوریتمی به نام CINEMA[۸۰] بر مبنای ساختار اجتماعات ابداع کردند [۶۰]. پس از مرحله شناسایی اجتماع، گره‌های تأثیرگذار نهایی بر اساس بیشترین بهره کسب کرده در داخل اجتماعات شنا سایی شده، انتخاب می شوند. CINEMA به دلیل ا ستفاده از شبیه‌سازی‌های پرهزینه مونت‌کارلو در شبکه‌های بزرگ قابل استفاده نیست. همچنین صرفاً به‌کارگیری معیار بیشترین بهره برای یافتن گره‌های تأثیرگذار در داخل اجتماعات منجر به کاهش دقت این الگوریتم شده است.

رحیم خانی و همکاران در سال ۲۰۱۵ مدل ComPath را ارائه دادند که برای شناسایی اجتماعات از الگوریتم SLPA[۸۱][۲۸] استفاده می‌کرد [۶۱]. در این مدل، تأثیرگذارترین اجتماعات و تأثیرگذارترین گره‌های هر اجتماع بر اساس معیار مرکزیت درجه شناسایی می‌شوند. در پایان، گره‌های تأثیرگذار نهایی از میان گره‌های کاندیدای اجتماعات بر اساس معیار فاصله داخلی[۸۲] انتخاب می‌شوند.

در مدل INCIM که تو سط بزرگی و همکاران در سال ۲۰۱۶ ارائه شده [۶۲]، هر اجتماع به‌عنوان یک گره جداگانه در نظر گرفته می شود؛ بنابراین یک گراف جدید ساخته می شود که گره‌های آن اجتماعات گراف پایه ه ستند. در این گراف جدید، هر گره (اجتماع) را می‌توان به‌عنوان یک ماژول انتشار مجزا در نظر گرفت. به‌طورکلی، ابتدا اجتماعات توسط SLPA [۲۸] شناسایی و سپس گراف جدید اجتماعات ساخته می شود. همچنین، الگوریتم SimplePath[۴۷] برای محاسبه گسترش تأثیر و استخراج گره‌های تأثیرگذار مورد استفاده قرار گرفته است. پیچ یدگی ز مانی بالای الگوریتم SimplePath در شبکه‌های بزرگ چالش اصلی INCIM است که

75 Maximum Likelihood Influence Maximization
76 Diversified Influence Maximization
77 Community Greedy Algorithm
78 Community-based Influence Maximization
79 Conformity
80 Conformity-aware INfluEnce MAximization
81 Speaker-listener Label Propagation Algorithm
82 intra-distance criterion

سعی شده است با بهینه‌سازی **CELF [35]** بهبود یابد.

جدول ۳. الگوریتم‌های برجسته حریصانه و مکاشفه‌ای

الگوریتم	مدل انتشار	پیچیدگی زمانی	مزیت	چالش
Greedy [4]	آبشاری مستقل، آستانه خطی	$O(krnm)$	ارائه ضمانت تئوری روی میزان اثربخشی	عدم کارایی زمانی روی گراف‌های بزرگ
CELF [23]	آبشاری مستقل، آستانه خطی	$O(krnm)$	ارائه ضمانت تئوری روی میزان اثربخشی	عدم کارایی زمانی روی گراف‌های بزرگ
SP1M [34]	آبشاری مستقل	$O(knm)$	عدم اجرای سنگین شبیه‌سازی‌های مونت‌کارلو	عدم فراهم کردن اثربخشی لازم در همه گراف‌ها (معیار طول مسیر برای محاسبه میزان تاثیر)
NewGreedy [26]	آبشاری مستقل	$O(krm)$	ارائه ضمانت تئوری روی میزان اثربخشی	زمان طولانی ایجاد گراف کوچکتر در هر مرحله از تکرار اصلی الگوریتم
Degree-Discount [26]	آبشاری مستقل، آبشاری وزندار	$O(k \log n + m)$	پیچیدگی زمانی خطی	عدم ارائه ضمانت تئوری روی میزان اثربخشی
MIA/PMIA [35]	آبشاری مستقل	$O(nt_{i\theta} + kn_{o\theta}\, n_{i\theta}\, (n_{i\theta} + \log n))$	ارائه ضمانت تئوری روی میزان اثربخشی	عدم مقیاس‌پذیری در گراف‌های متراکم- کاهش اثربخشی روی درختان محلی خیلی بزرگ
LDAG [33]	آستانه خطی	$O(nt_{\theta} + kn_{\theta}\, m_{\theta}\, (m_{\theta} + \log n))$	دقیق بودن محاسبه میزان تأثیر در گراف جهت دار بدون دور	ناکارمدی روند ساخت DAG ها ازنظر مصرف حافظه و زمان- عدم توازن بین کارایی زمانی و اثربخشی
CELF++ [24]	آبشاری مستقل، آستانه خطی	$O(krnm)$	ارائه ضمانت تئوری روی میزان اثربخشی	عدم کارایی زمانی روی گراف‌های بزرگ
SimplePath [36]	آستانه خطی	$O(kln\mathcal{P}_{\theta})$	استفاده از بهینه‌سازی CELF جهت افزایش اثربخشی	وابستگی به خصوصیات مدل انتشار آستانه خطی – عدم توازن بین کارایی زمانی و اثربخشی
SPIN [39]	آستانه خطی	$O(tr(n+m) + n \log n + kn)$	پیچیدگی زمانی نسبتاً خطی	عدم ضمانت روی کران بدترین حالت - وابستگی به خصوصیات مدل انتشار آستانه خطی
IRIE [40]	آبشاری مستقل	$O(k(n_{o\theta}\, k\, +\, m))$	اثربخشی قابل قبول بدلیل جلوگیری از پدیده همپوشانی تأثیر	افزایش زمان اجرای بدلیل استفاده از الگوریتم‌هایی مانند Greedy یا PMIA برای محاسبه تأثیر اضافی
StaticGreedy [27]	آبشاری مستقل	$O(\ \varepsilon^{-2}kmn^2\ log\binom{n}{k})$	ارائه ضمانت تئوری روی میزان اثربخشی	پیچیدگی زمانی بسیار بالا
IPA [43]	آبشاری مستقل	$O(\ c^{-1}nO_v\, n_{uv} + k2(\ c^{-1}O_v\, n_{uv}\ ))$	افزایش سرعت بدلیل استفاده از تکنیک های موازی‌سازی	کاهش اثربخشی روی درختان محلی خیلی بزرگ
TIM/TIM+ [29]	آبشاری مستقل، آستانه خطی	$O(\ \varepsilon^{-2}k(m+n)\ log(n))$	ارائه ضمانت تئوری روی میزان اثربخشی	مصرف زیاد حافظه
Group-PR [41]	آبشاری مستقل	$O(krm)(Linear), O(m + k^2n)(Bound)$	استفاده از مجموعه‌ای از گره‌ها برای محاسبه گسترش تأثیر	اثربخشی کم در حد الگوریتم رتبه‌بندی صفحات- عدم استفاده از خصوصیات مدل‌های انتشار
IMRank [42]	آبشاری مستقل	$O(nTd_{max}\, log\, d_{max})$	پیچیدگی زمانی نسبتاً خطی	عدم استفاده از خصوصیات مدل‌های انتشار – اختلاف زیاد بین تأثیر تخمینی با تأثیر واقعی
Pruned-MC [28]	آبشاری مستقل	$O(\ \varepsilon^{-2}kmn^2\ log\binom{n}{k})$	ارائه ضمانت تئوری روی میزان اثربخشی	پیچیدگی زمانی بسیار بالا
UBLF [25]	آبشاری مستقل، آستانه خطی	$O(krnm)$	ارائه ضمانت تئوری روی میزان اثربخشی	عدم کارایی زمانی روی گراف‌های بزرگ
IMM [30]	آبشاری مستقل، آستانه خطی	$O(\ \varepsilon^{-2}k(m+n)\ log(n))$	ارائه ضمانت تئوری روی میزان اثربخشی	مصرف زیاد حافظه
EaSyIM [44]	آبشاری مستقل، آستانه خطی	$O(k\mathcal{D}(n+m))$	گسترش تأثیر بهتر و زمان اجرای کمتر در مقایسه با +TIM	عدم ارائه ضمانت تئوری روی میزان اثربخشی
TABU–PG [45]	آستانه خطی	---	عدم اجرای سنگین شبیه‌سازی‌های مونت‌کارلو	اختلاف زیاد بین گسترش تأثیر تخمینی با گسترش تأثیر واقعی
TIFIM [46]	آستانه خطی	$O(T_l * \sum_{i=1}^{n} \lvert V_i^2 \rvert)$	سرعت اجرای قابل قبول (نزدیک به الگوریتم رتبه‌بندی صفحات)	عدم ارائه ضمانت تئوری روی میزان اثربخشی
MLIM [47]	آبشاری مستقل	---	عدم اجرای سنگین شبیه‌سازی‌های مونت‌کارلو	عدم ارائه ضمانت تئوری روی میزان اثربخشی
DIM [48]	آبشاری مستقل	$O(\ \varepsilon^{-2}k(m+n)\ log(n))$	ارائه ضمانت تئوری روی میزان اثربخشی	مصرف زیاد حافظه

علائم اختصاری: nوm به ترتیب برابر تعداد گره‌ها و یال‌های گراف- k برابر تعداد گره‌های سبد S- r برابر شبیه سازی‌های مونت‌کارلو- در PMIA نماد $t_{i\theta}$ برابر پیچیدگی زمانی ساخت MIIA و $n_{i\theta}$ و $n_{o\theta}$ برابر بزرگترین اندازه MIIA و MIOA- در LDAG نماد t_{θ} برابر پیچیدگی زمانی ساخت DAG و n_{θ} و m_{θ} برابر بیشترین تعداد گره و یال در یک LDAG است- در SimPath نماد ρ_{θ} برابر بیشترین تعداد مسیرهای ساده از یک راس و l برابر مقدار پیشرو- در SPIN، t برابر اندازه مجموعه نمونه‌گیری‌شده- در IPA، c برابر تعداد فرایندهای موازی ، O_v برابر تعداد بیشترین گره روی یک مسیر تاثیر، n_{uv} برابر بیشترین‌مسیرهای تاثیر بین دو گره- در IMRannk، T برابر تعداد تکرار، d_{max} برابر بیشترین تعداد مسیرهای منتهی به یک گره است- در EaSyIM، $\mathcal{D}$ برابر قطر گراف- در TIFIM، T_l برابر تعداد تکرار، V_i^2 مجموعه گره در میان معیار TwoHop گره V_i است.

هالاپاناوار و همکاران در سال ۲۰۱۶ مدلی را برای تسریع در فرانید اکتشاف گره‌های تأثیرگذار در شبکه‌های پیچیده با استفاده از شناسایی اجتماعات ایجاد کرند [۶۳]. تعبیه یک روش موازی‌سازی در الگوریتم Louvain [۲۲] نقطه قوت روش آن‌هاست. همچنین، در این مقاله به‌طور عملی نشان داده‌شده است که با استفاده از رویکرد شناسایی گره در اجتماعات شناسایی‌شده و در نهایت تجمیع سبدها، میزان گسترش تأثیر به‌طور قابل‌توجهی افزایش یافته است.

ژائو و همکاران در سال ۲۰۱۶ الگوریتمی معرفی کردند که در آن گام‌های شناسایی اجتماعات و محاسبه گسترش تأثیرگره‌ها با هم تجمیع شده‌اند [۶۴]. با توجه به الگوریتم ارائه شده، تأثیرگذارترین گره در اجتماع می‌تواند برچسب خود را در طی فرآیند انتشار به همه گره‌های دیگر در اجتماعش منتشر کند. همچنین، آن‌ها معیاری برای محاسبه مرکزیت گره‌ها تعریف کردند. این مدل در مقایسه با سایر الگوریتم‌ها به برتری متمایزی هم از لحاظ کارایی زمانی و هم از لحاظ اثربخشی در گراف‌هایی با اندازه متوسط دست یافت. ازآن‌جایی که دو گام شناسایی اجتماعات و یافتن گره‌های تاثیرگذار با هم تجمیع شده اند، زمان اجرای الگوریتم بسیار بهبود می یابد. از طرف دیگر از آنجا که اجتماعات بوسیله‌ی روش انتشار برچسب شکل می‌گیرند، اجتماعات حاصل با ذات مسئله‌ی بیشینه‌سازی تأثیر(سازوکار مدل‌های انتشار) بسیار سازگار و منطبق هستند. در نتیجه این الگوریتم گره‌های تاثیرگذار را با دقت بالاتری انتخاب می‌کند.

الگوریتم DIN[۸۳] یک مدل بدون پارامتر بر مبنای ترکیبی از جنبه‌های ساختاری و معنایی اجتماعات است که توسط جاادی و همکاران در سال ۲۰۱۶ پیشنهاد شده است [۶۵]. ایده‌های اصلی این الگوریتم شناسایی اجتماعات همپوشان، مدل‌سازی معنایی هر اجتماع و سپس انتخاب گره‌های تأثیرگذار است. بدون پارامتر بودن و استخراج اجتماعات همپوشان (همانند شبکه‌های اجتماعی واقعی) از نقاط قوت این الگوریتم به شمار می‌آیند. اما فرآیند مدل‌سازی معنایی هر اجتماع در جاهایی که اندازه اجتماعات بزرگ باشد منجر به کندی الگوریتم خواهد شد.

شانگ و همکاران در سال ۲۰۱۶ یک چارچوب حریصانه بر مبنای شناسایی اجتماعات به نام CoFIM[۸۴] ارائه دادند [۶۶]. در این الگوریتم فرآیند انتشار اطلاعات یک فرآیند دو فاز است. در فاز انبساط سبد، اطلاعات از گره‌های مجموعه سبد (که معمولاً در اجتماعات مختلف واقع شده‌اند) به همسایه‌هایشان انتشار می‌یابد. در فاز انتشار درون اجتماعی، روند انتشار در داخل اجتماعات مستقل از یکدیگر رخ می‌دهد. استراتژی آن‌ها برای انتخاب گره‌های تأثیرگذار بر اساس الگوریتم حریصانه‌ی ابداعی است که باعث شده این الگوریتم کارایی قابل قبولی در زمان اجرا داشته باشد. این الگوریتم نتوانست تضمینی برای اثربخشی ارائه دهد و در برخی از گراف‌ها عملکردی درحد الگوریتم‌های مبتنی‌برمرکزیت داشت. البته بعدها بسیاری از نویسندگان الگوریتم پیشنهادی خود را با الگوریتم CoFIM (به عنوان یک الگوریتم مبنا) مقایسه کرده اند [۶۷]و [۶۸] و [۳۴].

در مطالعه دیگری در سال ۲۰۱۷ حسینی پژوه و همکاران دو الگوریتم جدید در مسئله‌ی بیشینه‌سازی تأثیر تحت مدل‌های انتشار آستانه خطی، آستانه ضربی[۸۵] و آستانه حداقلی[۸۶] را با استفاده از معیار مرکزیت نزدیکی و شناسایی اجتماعات ارائه دادند [۶۹]. دو الگوریتم پیشنهادی C-SPIN[۸۷] و C-SGA[۸۸] به ترتیب از رویکردهای SPIN[۸۹] [۴۹] و SGA[۹۰] [۷۰] برای ساخت مجموعه‌ی گره‌های کاندیدا استفاده می‌کنند. الگوریتم LabelRank [۷۱] به‌منظور شناسایی اجتماعات به‌کارگرفته شده است. برای انتخاب گره‌های تأثیرگذار نهایی از معیار مرکزیت نزدیکی استفاده شده است. ارائه الگوریتم برای مدل‌های انتشار آستانه ضربی و آستانه حداقلی را می‌توان از نکات برجسته این تحقیق برشمرد. لازم به ذکر است که الگوریتم‌های پیشنهادی این تحقیق در گراف‌های بزرگ مقیاس‌پذیر نیستند. به همین دلیل نویسندگان الگوریتم‌های پیشنهادی خود را صرفاً در دیتاست‌های بسیار کوچک (با ۳۴ تا ۱۸۶۷ گره) بررسی کرده‌اند.

بزرگی و همکاران در سال ۲۰۱۷ به مطالعه مسئله رقابت را در مسئله‌ی بیشینه سازی تأثیر پرداختند و الگوریتم CI2[۹۱] را معرفی کردند [۷۲]. در مرحله اول اجتماعات توسط الگوریتم SLPA شناسایی می‌شوند. سپس به‌منظور انتخاب گره‌های تأثیرگذار، الگوریتم حریصانه‌ی کمپ [۴] روی هر اجتماع تحت مدل انتشار رقابتی DCM[۹۲] (که نسخه‌ای از مدل آستانه خطی است) اجرا می‌شود. تأثیرگذارترین گره‌های انتخاب‌شده هر اجتماع به مجموعه‌ی نامزدها اضافه می شوند. سرانجام در یک روش تکراری، گره‌هایی با بیشترین سود حاشیه‌ای به‌عنوان تأثیرگذارترین گره‌های شبکه اجتماعی انتخاب خواهند شد. از معایب این الگوریتم می‌توان به سرعت پایین اجرای الگوریتم برای گراف‌های بزرگ به‌دلیل استفاده از الگوریتم حریصانه‌ی کمپ اشاره کرد. همچنین این الگوریتم تک‌منظوره بوده و رقابت همزمان را در نظر نگرفته و صرفاً برای رقابت دو بازیگر طراحی شده است نه چند بازیگر.

الگوریتم های BCRIM[۹۳] و ICRIM[۹۴] نسخه های بهبود یافته

[83] Detecting Influential Nodes
[84] COmmunity-based Framework Influence Maximization
85 Multiplication Threshold
86 Minimum Threshold
[87] Community-aware SPIN
[88] Community-aware SGA
[89] Shapley value-based influential nodes
[90] Set-based coding Genetic Algorithm
[91] Competitive Influence Improvement
[92] Decidable Competitive Model
[93] Basic Community Robust Influence Maximization
[94] Improved Community Robust Influence Maximization

الگوریتم حریصانه‌ی کمپ هستند که توسط یی و همکاران در سال ۲۰۱۸ پیشنهاد شده‌اند [۷۳]. ایده اصلی آن تقسیم کل گراف به اجتماعات با استفاده از الگوریتم کلاسیک k-means است. در طی فرآیند شناسایی اجتماعات، اطلاعات گره‌ها و همسایگان آن‌ها با استفاده از روش LINE به منظور استخراج اجتماعات باکیفیت‌تر نگهداری می‌شود. در نهایت برای انتخاب گره های تأثیر گذار، الگوریتم حریصانه‌ی‌کمپ روی دو گروه از گره‌ها اجرا می‌شود: گره های هر اجتماع و گره های مرزی با اجتماعات همسایه. الگوریتم های BCRIM و ICRIM به نتایج قابل قبولی ازنظر سرعت و اثربخشی روی گراف‌های کوچک و متوسط دست یافتند. اما بدلیل استفاده از الگوریتم حریصانه‌ی کمپ [۴] در مواردی که اندازه اجتماعات بدست‌آمده بزرگ باشد (گراف‌های بزرگ)، دارای کارایی زمانی قابل قبولی نیستند.

در مدل GTaCB[۹۵] که توسط جلایر و همکاران در سال ۲۰۱۸ پیشنهاد شده است [۷۴]، از چند معیار مرکزیت برای یافتن تأثیرگذارترین گره‌ها تحت مدل انتشار SIR[۹۶] استفاده شده است: مرکزیت درجه، مرکزیت نزدیکی، مرکزیت بینابینی و الگوریتم رتبه‌بندی صفحات. در این تحقیق، گراف توسط الگوریتم خوشه‌بندی طیفی به چندین زیرگراف تقسیم می‌شود. سپس الگوریتم TOPSIS[۹۷] برای رتبه‌بندی گره های هر زیرگراف اجرا می‌شود. در مرحله بعدی، زیرگراف‌ها بر اساس تعداد گره‌های آن‌ها، به ترتیب نزولی مرتب می‌شوند. در نهایت بر اساس ترتیب به‌دست‌آمده اجتماعات، مهم‌ترین گره‌ی هر اجتماع (با بالاترین امتیاز TOPSIS) به مجموعه سبد S اضافه می‌شود تا اینکه k گره انتخاب شوند. اشکال این روش، انتخاب گره‌های تأثیرگذار نهایی از هر زیرگراف برمبنای معیار تعداد گره‌های زیرگراف‌ها است؛ حال اگر یک اجتماع بسیار بزرگ و یک اجتماع کوچکتر وجود داشته باشد، تعداد گره‌های انتخابی از هر دو یکسان خواهد بود. این سیاست منجر به استخراج یک مجموعه سبد با کیفیت پایین خواهد شد.

کو و همکاران در سال ۲۰۱۸ یک مدل ترکیبی به نام HybridIM[۹۸] ارائه دادند [۷۵] که با ترکیب تکنیک شناسایی اجتماعات و استراتژی مبتنی بر مسیر، یک چارچوب PB-CD[۹۹] را به وجود می‌آورد. در مرحله اول، گراف توسط یک روش مبتنی بر مسیر به اجتماعات تقسیم می‌شود. همچنین برای محاسبه گسترش تأثیر هر گره از مجموع وزن مسیرهای منتهی به گره های دیگر استفاده می‌کند. در مرحله بعدی برای هر اجتماع یک صف بر اساس ترتیب نزولی سود حاشیه‌ای گره‌ها ایجاد می‌شود. سپس الگوریتم CELF به‌طور مستقیم روی هر صف اعمال می‌شود. در انتها، k گره با بیشترین سود حاشیه‌ای به‌عنوان گره‌های تأثیرگذار نهایی انتخاب می‌شوند. اثربخشی قابل‌قبول این الگوریتم بدین دلیل است که اجتماعات بوسیله‌ی یک روش مبتنی‌برمسیر ساخته می‌شوند. در نتیجه اجتماعات حاصله با فرآیند پخش اطلاعات در مدل‌های انتشار بسیار سازگار هستند. ولی متاسفانه از آنجاکه HybridIM، الگوریتم CELF (با زمان اجرای بسیار زیاد) را روی همه‌ی اجتماعات اجرا می‌کند، در مواردی که اندازه اجتماعات بدست‌آمده بزرگ باشد دارای کارایی زمانی قابل قبولی نیست.

در سال ۲۰۱۸ لی و همکاران با استفاده از رویکرد شناسایی اجتماعات، الگوریتمی مکان-آگاه[۱۰۰] پیشنهاد کردند [۷۶]. هدف آن انتخاب یک مجموعه سبد اولیه است که میزان گسترش تأثیر را بر روی کاربران هدف که توسط یک پرسش[۱۰۱] مشخص شده‌اند، بیشینه کند. چالش‌های اصلی پیدا کردن کاربران هدف و محاسبه کارآمد اولویت‌ها و ترجیحات آن‌ها است که با یک راه‌حل مبتنی بر R - درخت[۱۰۲] حل شده‌اند. از الگوریتم MIA که نسخه بهبودیافته الگوریتم حریصانه‌ی کمپ است برای محاسبه میزان گسترش تأثیر استفاده شده است. در ابتدا، گراف توسط الگوریتم خوشه‌بندی طیفی به اجتماعات تقسیم می‌شود تا سربار محاسباتی بالا را کاهش دهد. سپس الگوریتم MIA در هر اجتماع به‌طور مستقل اجرا خواهد شد. سربار محاسباتی الگوریتم MIA در گراف‌های بزرگ و در نتیجه مقیاس‌پذیری از معظلات این الگوریتم مکان-آگاه است.

الگوریتم IMPC[۱۰۳] که یک مدل بیشینه‌سازی تأثیر مبتنی بر چند همسایه بالقوه است توسط شانگ و همکاران در سال ۲۰۱۸ پیشنهاد شد [۷۷]. در رویکرد آن‌ها، فرآیند انتشار تأثیر مانند ایده الگوریتم CoFIM [۶۶] در دو مرحله صورت می‌گیرد: ۱. گسترش سبدها بر مبنای چند همسایه بالقوه ۲. انتشار تأثیر درون اجتماعات. همچنین، یک تابع هدف برای ارزیابی تأثیر کلی به‌عنوان ترکیبی از تأثیر در طول دو مرحله، تعریف شده است. این الگوریتم همانند الگوریتم CoFIM [۶۶] نتوانست تضمینی برای اثربخشی ارائه دهد و در برخی از گراف ها عملکردی در حد الگوریتم های مبتنی‌برمرکزیت داشت.

Liqing و همکاران در سال ۲۰۱۹ یک الگوریتم سه فاز به نام PHG[۱۰۴] را ابداع کردند [۷۸] که دارای فازهای زیر است: فاز تقسیم، فاز مکاشفه‌ای و فاز حریصانه. در فاز تقسیم، با استفاده از الگوریتم Louvain اجتماعات را شناسایی می‌شوند. سپس بر اساس معیار مرکزیت، گره‌های کلیدی هر اجتماع انتخاب شده و یک مجموعه نامزد تشکیل می‌شود. در فاز مکاشفه‌ای، گره‌های تأثیرگذارتر مجموعه نامزد، به‌وسیله ترکیب وزن میزان تأثیر گره‌ها

[95] Greedy TOPSIS and Community-Based
[96] Susceptible-Infected-Recovered
[97] Technique for Order Performance by Similarity to Ideal Solution
[98] Hybrid Influencae Maximization
[99] Path Based Community Detection
[100] location aware
[101] query
[102] R-tree
[103] Influence Maximization framework based on Multi-neighbor Potential in Community
[104] Partition-Heuristic-Greedy

و تأثیر اجتماعات آن‌ها مشخص می شوند. در فاز حریصانه تو سط الگوریتم حریصانه‌ی کمپ، k گره با بیشترین سود حاشیه‌ای به‌عنوان تأثیرگذارترین گره‌ها، به مجموعه سبد اضافه می شوند. به عنوان یک نقطه قوت، استفاده از الگوریتم Louvain منجر به استخراج جوامع بهینه خواهد شد. نقطه‌ضعف این روش اندازه‌گیری میزان گسترش تأثیر گره بر اساس معیار مرکزیت درجه در گام مکاشفه‌ای و بازتاب اثرات منفی آن به گام‌های بعدی است.

در سال ۲۰۱۹ سینگ و همکاران با تمرکز بر ترجیحات کاربران به‌عنوان یک موضوع در مسئله‌ی بیشینه‌سازی تأثیر، چارچوب C2IM[۱۰۵] را پیشنهاد دادند [۷۹]. برای این منظور، مدل‌های انتشار انتشار آبشاری مستقل و آستانه‌ی خطی را توسعه دادند و به ترتیب مدل‌های CIC[۱۰۶] و CLT[۱۰۷] را معرفی کردند. در ابتدا، اجتماعات تو سط الگوریتم خو شه‌بندی سلسله مراتبی شنا سایی می شوند. سپس الگوریتم انتخاب گره‌های تأثیرگذار بر اساس درجه انتشار گره‌ها با در نظر گرفتن ترجیحات کاربران بر روی هر اجتماع اجرا می‌شود. وجود یک مرحله پیش‌پردازش برای حذف کردن گره‌های کم‌اهمیت، در کل باعث افزایش سرعت اجرای الگوریتم شد. در طرف مقابل، نادیده گرفتن تأثیر گره‌ها بر روی گره‌های اجتماعات دیگر بر روی کیفیت مجموعه سبد خروجی اثر منفی گذاشت.

هوانگ و همکاران در سال ۲۰۱۹ بیشینه‌سازی تأثیر موضوع-آگاه را در مدل CTIM[۱۰۸] مطالعه کردند [۸۰]. آن‌ها الگوریتم شناسایی اجتماع را به‌جای اجرای مستقل، با مدل انتشار تأثیر ادغام کردند. آن‌ها یک مدل متغیر نهفته جامع ساخته‌اند که موارد زیر را ضبط می‌کند: ۱. رابطه موضوع-آیتم ۲. علاقه‌مندی‌های موضوعی در سطح اجتماع ۳. توزیع عضویت در اجتماع برای هر کاربر. همچنین آن‌ها برای آموزش مدل پیشنهادی‌شان، الگوریتم نمونه‌برداری گیبس را توسعه داده‌اند. نویسندگان در نتایج گزارش شده خود [۸۰](صفحه ۲۱۴۷ پاراگراف ۲) ادعا کردند که الگوریتم پیشنهادیشان، الگوریتم حریصانه‌ی کمپ را از نظر اثربخشی شکست داده است. لازم به ذکر است که الگوریتم حریصانه‌ی کمپ [۴] تاکنون دارای بیشترین میزان اثربخشی در میان همه الگوریتم‌های مطرح‌شده در مسئله‌ی بیشینه‌سازی تأثیر است. از این رو این ادعای عجیب نویسندگان باید مورد بررسی بیشتر قرار گیرد.

ComBIM[۱۰۹] یک چارچوب مبتنی بر اجتماع برای حل مسئله‌ی بیشینه‌سازی تأثیر بودجه‌بندی شده است که توسط بانرجی و همکاران در سال ۲۰۱۹ پیشنهاد شده است [۸۱]. این چارچوب چهار مرحله اصلی دارد: ۱. شناسایی اجتماع ۲. توزیع بودجه کل بین اجتماعات ۳. انتخاب گره‌های تأثیرگذار مجموعه سبد ۴. انتقال بودجه استفاده‌نشده از یک اجتماع به اجتماع دیگر. پیچیدگی زمانی این الگوریتم $O(ln^2 + nd + l^2)$ مقیاس پذیری الگوریتم را در شبکه‌های اجتماعی بزرگ زیر سوال خواهد برد. (n تعداد گره های گراف، l تعداد اجتماعات و d بزرگترین درجه در گراف است)

سینگ و همکاران در سال ۲۰۱۹ یک مدل بیشینه‌سازی تأثیر مبتنی بر اجتماع به نام CoIM[۱۱۰] که بر حل معضل کارایی زمان متمرکز است، پیشنهاد کردند [۸۲]. این مدل گراف اصلی را با استفاده از الگوریتم خو شه‌بندی سلسله مراتبی بر اساس شباهت ژاکارد به اجتماعات تقسیم می‌کند. در مرحله بعدی، گره‌های نهایی مجموعه سبد را بر اساس تأثیر محلی گره‌ها در داخل اجتماع انتخاب می‌کند. ایراد اساسی این مدل این است که تأثیر گره‌ها بر روی گره های اجتماعات دیگر را به‌حساب نمی‌آورد و در نتیجه کاهش اثربخشی است.

اتیف و همکاران در سال ۲۰۲۰ یک چارچوب فازی مبنا برای حل مسئله‌ی بیشینه‌سازی تأثیر ارائه دادند [۸۳]. در اولین گام، با تقسیم شبکه به اجتماعات بو سیله الگوریتم CNM-Similarity، فضای جستجوی مسئله کاهش می‌یابد. اجتماعات کشف‌شده به‌عنوان ماژول هایی برای یافتن تأثیرگذارترین گره ها استفاده می‌شوند. در پایان، گره‌های نهایی سبد بر اساس منطق فازی به صورت نسبی انتخاب می شوند. همانطورکه نویسندگان پیشنهاد دادند، مقیاس‌پذیری این الگوریتم فازی مبنا می‌بایست مورد مطالعه بیشتر قرار بگیرد. لازم به ذکر است که نویسندگان کارایی و اثربخشی الگوریتم پیشنهادیشان را با هیچکدام از الگوریتم‌های برجسته در این زمینه محک نزده اند.

الگوریتم CIMA[۱۱۱] که یک مدل مبتنی بر شناسایی اجتماعات در شبکه‌های خصوصیت‌مبنا است تو سط هوانگ و همکاران در سال ۲۰۲۰ معرفی شد [۶۷]. این الگوریتم دارای سه مرحله اصلی است: ۱. شناسایی اجتماعات ۲. شناسایی اجتماعات منتخب به‌منظور کاهش فضاهای جستجو برای انتخاب سبد ۳. انتخاب گره‌های تأثیرگذار. آن‌ها همچنین برای بهبود دقت پیش‌بینی، مدلی را برای پیش‌بینی قدرت تأثیر بین گره‌ها در این نوع شبکه‌ها ارائه دادند. علیرغم استخراج گره‌های تاثیرگذار با کیفیت بالا، پیچیدگی زمانی بدترین حالت این الگوریتم بسیار پرهزینه است $O(n^3)$.

ژانگ و همکاران در سال ۲۰۲۰ مدلی را بر اساس ساختار اجتماعات و اختلاف توزیع تأثیر در مسئله‌ی بیشینه‌سازی تأثیر بررسی کردند که CBIMA[۱۱۲] نامیده می‌شود [۸۴]. در اولین گام، ساختارهای اجتماعات شناسایی می‌شوند. در مرحله بعدی، یک الگوریتم مکاشفه‌ای برای یافتن گره‌های نامزد از بین گره‌های داخلی و مرز هر اجتماع به کار گرفته می‌شود. سرانجام بر مبنای خاصیت

[105] Community Context-aware Influence Maximization
[106] Context-aware Independent Cascade model
[107] Context-aware Linear Threshold model
[108] Community-based Topic-aware Influence Maximization
[109] COMmunity-based Budgeted Influence Maximization
[110] COmmunity- based Influence Maximization
[111] Community-based Influence Maximization in Attributed networks
[112] Community Based Influence Maximization Algorithm

زیرمدولاریته بودن تابع گسترش تأثیر، گره‌های دارای بیشترین بهره حاشیه‌ای به‌عنوان گره‌های تأثیرگذار مجموعه سبد انتخاب می‌شوند. اثربخشی نسبتاً بالای CBIMA (نزدیک به الگوریتم حریصانه‌ی کمپ) به خاطر الگوریتم مکاشفه‌ای پیشنهادی به همراه معیار اختلاف توزیع گسترش تأثیر گره‌ها است. از طرف دیگر، فرآیند یافتن گره های نامزد از بین گره های داخلی و مرزی هر اجتماع باعث افت کارایی زمانی این مدل پیشنهادی شده است.

مدل TI-SC[113] یک رویکرد مبتنی بر اجتماع است که توسط بنی و همکاران در سال ۲۰۲۰ ارائه شده است [۸۵]. این مدل با بررسی روابط بین گره‌های اصلی (هسته) و توانایی امتیازدهی گره‌های دیگر، تأثیرگذارترین گره ها را پیدا می‌کند. این روش امتیازدهی باعث کاهش همپوشانی گره‌های مجموعه سبدهای متفاوت شده و یک سبد با کیفیت بالا به‌دست می‌آید. همچنین به‌منظور کاهش سربار محاسباتی اضافی، در ابتدا اجتماعات کوچک حذف شدند.

هی و همکاران در سال ۲۰۲۰ مسئله بیشینه‌سازی عقیده‌ی مبتنی بر شناسایی اجتماعات را تحت مدل پیشنهادی CAOM[114] مطالعه کردند [۶۸]. آن ها یک مدل انتشار عقیده وزن‌دار را به‌منظور محاسبه تغییر پویایی ارزش عقیده ابداع کردند. پس از مرحله شناسایی اجتماعات، به‌منظور کاهش سربار محاسباتی اضافی، اجتماعات مهم‌تر شناسایی شده تا گره‌های تأثیرگذار صرفاً از داخل آن‌ها برداشته شوند. در مرحله بعدی با استفاده از معیار یک-هاپ، گره‌های تأثیرگذار نامزد صرفاً از میان گره‌های بالقوه هر اجتماع بااهمیت‌تر انتخاب می شوند. سرانجام، گره‌های نهایی مجموعه سبد از میان گره های نامزدشده بر اساس معیار دو-هاپ و حذف همپوشانی انتخاب خواهند شد. استفاده از الگوریتم Louvain برای تشخیص اجتماعات و همچنین ابداع مدل انتشار پویا از نقاط قوت این تحقیق است. از طرف دیگر شناسایی اجتماعات بااهمیت‌تر و حذف اجتاماعات کم‌اهمیت، یک چالش بزرگ است که باعث افت کارایی و اثربخشی شده است.

الگوریتم CFIN[115] که توسط خمامی و همکاران [۸۶] در سال ۲۰۲۰ پیشنهاد شد، شامل دو مرحله اصلی است: ۱. انتخاب سبد ۲. گسترش محلی اجتماع. پس از مرحله شناسایی اجتماع، به‌منظور کاهش سربار محاسباتی اضافی، اجتماعات مهم‌تر بر اساس اندازه و تراکم‌شان انتخاب می شوند. سپس k گره با بالاترین درجه به‌عنوان گره‌های اولیه از اجتماعات انتخاب می‌شوند. گره‌های نهایی سبد به‌وسیله الگوریتم simplePath [۴۷] تأثیر محلی خودشان را در درون اجتماعات گسترش می‌دهند. حذف اجتماعات کم اهمیت از روند پیداکردن گره‌های تاثیرگذار از دو جنبه حائز اهمیت است. از نظر زمان اجرای الگوریتم، مسلماً باعث افزایش سرعت شده است. ولی ممکن است که گره‌های تاثیرگذارتر در اجتماعاتی باشند که حذف شده‌اند و بدین‌طریق در سبد نهایی قرار نخواهند گرفت.

احمد و همکاران در سال ۲۰۲۰ یک رویکرد مبتنی بر اجتماع به نام HWSMCB[116] معرفی کردند [۸۷]. این مدل یک روش جمع وزن‌دار پویا را برای رتبه‌بندی گره ها در هر اجتماع پیاده‌سازی می‌کند. همچنین، یک روش تصمیم‌گیری چندمعیاره برای در نظر گرفتن هم‌زمان ویژگی‌های توپولوژیکی گره‌ها استفاده شده است. بر اساس تحلیل نویسندگان، پیچیدگی زمانی این الگوریتم $O({n_1}^3)$ می‌باشد که n_1 برابر تعداد گره‌های بزرگترین اجتماع است. در مواردی که اندازه اجتماعات استخراج شده بزرگ باشد، کارایی زمانی این الگوریتم بسیار افت پیدا خواهد کرد.

همان‌طور که ذکر شد، سیاست الگوریتم‌های CAOM [۶۸] و CIMA [۶۷] بر این است که گره‌های تأثیرگذار نهایی را فقط از اجتماعات مهم انتخاب کنند. پیداکردن معیار و سنجه‌ای مناسب برای تعیین درجه اهمیت اجتماعات دشوار است و باعث افت کارایی می شود. وانگ و همکاران در سال ۲۰۲۱ مدل مبتنی بر اجتماعات CNCG[117] را برای دور زدن این چالش پیشنهاد دادند [۳۴]. پس از شناسایی اجتماعات همپوشان، به‌منظور محاسبه تأثیر هر گره در داخل اجتماعش، معیار مرکزیت حساس به پوشش گره تعیین شد. سپس گره‌های تأثیرگذار نهایی مستقیماً با ترکیب ساختار اجتماعات شناسایی‌شده با یک استراتژی از پیش طراحی‌شده (بدون شناسایی اجتماعات مهم‌تر) انتخاب شدند. الگوریتم CNCG مشابه الگوریتم CoFIM [۶۶] علیرغم سرعت بالا، نتوانست اثربخشی خود را در همه مجموعه داده‌ها تضمین کند. در جدول ۴ الگوریتم‌های بیشینه‌سازی تأثیر مبتنی بر اجتماعات لیست شده است.

۴. مقایسه و تحلیل

در بخش قبل، الگوریتم‌های برجسته در حوزه مسئله‌ی بیشینه‌سازی تأثیر از لحاظ دو معیار کلیدی «زمان اجرا» و «گسترش تأثیر» مورد بررسی قرار گرفتند.

۱.۴. معیارهای ارزیابی الگوریتم‌های موجود

تقریباً همه‌ی الگوریتم‌های معرفی شده در مسئله‌ی بیشینه سازی تأثیر ازلحاظ دو معیار «زمان اجرا[118]» و «گسترش تأثیر[119] (اثربخشی)» مورد ارزیابی قرار گرفته‌اند [۸]. از معیار «زمان اجرا» برای ارزیابی کارایی زمانی و مقیاس پذیری الگوریتم ها استفاده می‌شود. به عبارت دیگر زمان اجرای الگوریتم برابر با زمان سپری شده برای یافتن k گره تأثیرگذار به‌منظور تشکیل مجموعه سبد S است.

[113] Top-k Influential - Scoring Criteria
[114] Community-based Approach for Opinion Maximization
[115] Community Finding Influential Node
[116] Hybrid Weighted Sum Method Community-Based
[117] Community structure and Node Coverage Gain
[118] Runnug Time
[119] Influence Spread

جدول ۴. الگوریتم‌های بیشینه‌سازی تأثیر مبتنی بر تشخیص اجتماعات

الگوریتم	مدل انتشار	پیچیدگی زمانی	تشخیص اجتماعات	مزیت	چالش
CGA [22]	آبشاری مستقل	$O(m + rm_{\mathcal{P}}(n(Z-M) + k(M+n_{\mathcal{P}})))$	انتشار برچسب	دقت بالا بدلیل استفاده از الگوریتم انتشار برچسب	ناتوانی در مواجهه با گراف‌های بزرگ
CIM [49]	مدل انتشار حرارتی	---	سلسله مراتبی	کارایی زمانی بالا	کاهش اثربخشی بدلیل فرض عدم همپوشانی اجتماعات
CINEMA [50]	آبشاری مستقل	$O(k'm'n' + kTRm')$	ℓ-way	---	استفاده از شبیه‌سازی‌های پرهزینه مونت‌کارلو
ComPath [51]	آستانه خطی	---	SLPA	سرعت بالا بدلیل استفاده از معیار مرکزیت درجه	کاهش اثربخشی بدلیل استفاده از معیار مرکزیت درجه
INCIM [53]	آستانه خطی	---	SLPA	استخراج اجتماعات بهینه بدلیل استفاده از الگوریتم SLPA	پیچیدگی زمانی بالای الگوریتم SimplePath در شبکه‌های بزرگ
Parallel [54]	آبشاری مستقل، آستانه خطی	$O(nk^2\|V\|(\|V\|+\|E\|))$	Louvain	موازی‌سازی در الگوریتم Louvain	پیچیدگی زمانی بسیار بالا
IM-LPA [55]	آبشاری مستقل، آستانه خطی	$O(Tmd)$	انتشار برچسب	تجمیع شناسایی اجتماعات و یافتن گره‌های تاثیرگذار	عدم ارائه ضمانت تئوری روی میزان اثربخشی
DIN [56]	آستانه خطی	$O(n^2)$	DOCNet	بدون پارامتر بودن و استخراج اجتماعات همپوشان	عدم کارایی زمانی بدلیل فرآیند مدل‌سازی معنایی اجتماعات بزرگ
CoFIM [57]	آبشاری وزندار	$O(k^2 n k_{max})$	Louvain	کارایی زمانی قابل قبول	عدم ارائه ضمانت تئوری روی میزان اثربخشی
CI2 [59]	آستانه خطی رقابتی	---	SLPA	اثربخشی قابل قبول بدلیل استفاده از الگوریتم SLPA	سرعت پایین در گراف‌های بزرگ به‌دلیل استفاده از الگوریتم حریصانه
GTaCB [61]	SIR	---	طیفی	سرعت بالای الگوریتم بیشینه‌سازی تاثیر	استخراج یک مجموعه سبد با کیفیت پایین
HybridIM [62]	آبشاری وزندار	---	انتشار برچسب	استخراج گره‌های تاثیرگذار با کیفیت	عدم کارایی زمانی در گرافهای بزرگ (بدلیل استفاده از CELF)
Location-aware [63]	آبشاری وزندار، آبشاری سه گانه	---	طیفی	اثربخشی قابل قبول بدلیل استفاده از الگوریتم MIA	سربار محاسباتی الگوریتم MIA در گراف‌های بزرگ و عدم مقیاس‌پذیری
IMPC [64]	آبشاری وزندار	$O(n + m + k(m + k_{max}))$	Louvain	پیچیدگی زمانی خطی	عدم ارائه ضمانت تئوری روی میزان اثربخشی
PHG [65]	آستانه خطی	---	Louvain	استخراج جوامع بهینه	انتخاب گره‌های تاثیرگذار بر اساس معیار مرکزیت درجه
C2IM [66]	آبشاری مستقل، آستانه خطی	$O((\|V\|+\|E\|) * \log(\|V\|+\|E\|))$	انتشار برچسب	افزایش کارایی زمانی بدلیل اعمال الگوریتم انتشار برچسب	کاهش اثربخشی بدلیل نادیده گرفتن تأثیر گره‌ها بر روی اجتماعات دیگر
ComBIM [68]	آبشاری مستقل	$O(ln^2 + nd + l^2)$	Node Similarity	اثربخشی قابل قبول	عدم مقیاس‌پذیری
COIM [69]	آبشاری مستقل، آبشاری وزندار، آستانه خطی	---	سلسله مراتبی	کارایی زمانی قابل قبول	کاهش اثربخشی بدلیل نادیده گرفتن تأثیر گره‌ها بر روی اجتماعات دیگر
Fuzzy-based [70]	آبشاری مستقل	---	CNM	ارائه رویکردی بر مبنای فازی	عدم مقایسه با الگوریتم‌های شاخص
CIMA [71]	آبشاری مستقل	$O(n^3)$	Louvain	اثربخشی قابل قبول بدلیل استخراج اجتماعات بهینه	پیچیدگی زمانی بالا
CBIMA [72]	آبشاری مستقل	---	k-means	اثربخشی نسبتاً بالا به علت الگوریتم مکاشفه‌ای پیشنهادی	افت کارایی زمانی بدلیل انتخاب از بین گره‌های داخلی و مرزی اجتماع
TI-SC [73]	آبشاری مستقل	$O(n \log n + m + Rn + (k-1)Rn')$	Louvain	افزایش اثربخشی بدلیل کاهش همپوشانی گره‌های سبد	حذف اجتماعات کوچک و در نتیجه کاهش سربار محاسباتی اضافی
CAOM [76]	وزندار	---	Louvain	ابداع مدل انتشار پویا و استخراج اجتماعات بهینه	افت کارایی زمانی بدلیل فرآیند شناسایی اجتماعات مهم
CFIN [74]	آستانه خطی	$O(n^2)$	DLACD	افزایش سرعت بدلیل حذف اجتماعات بی‌اهمیت	عدم مقیاس‌پذیری
HWSMCB [75]	SIR	$O(n_1^3)$	طیفی	اثربخشی مناسب تصمیم‌گیری چندمعیاره	پیچیدگی زمانی بالا
CNCG [77]	آبشاری مستقل، آبشاری وزندار	$O((kl+n) < d >)$	NLA	استخراج اجتماعات همپوشان و افزایش اثربخشی	عدم ارائه ضمانت تئوری روی میزان اثربخشی روی گراف‌های بزرگ

علائم اختصاری: mوn به ترتیب برابر تعداد گره‌ها و یال‌های گراف- k برابر تعداد گره‌های سبد S- r برابر شبیه‌سازی‌های مونت‌کارلو- در CGA n_p و m_p برابر تعداد گره‌ها و یالها در بزرگترین اجتماع، Zو M به ترتیب تعداد اجتماعات قبل و بعد از گام ترکیب و P یک مقدار آستانه‌ای برای چگال بودن اجتماع است - در CINEMA، n' و m' برابر بیشترین تعداد گره‌ها و یال‌ها در یک زیرگراف و k' تعداد تکرار است- در IM-LPA، T برابر تعداد تکرار الگوریتم شناسایی اجتماعات، d برابر میانگین درجه گره‌هاست- در CoFIMو IMPC مقدار k_{max} برابر بیشترین درجه گره‌هاست- در ComBIM، d برابر میانگین درجه گره و l برابر تعداد تکرار الگوریتم است. درTI-SC، n' برابر بیشترین تعداد گره‌ها و یال‌ها در یک زیرگراف است- در HWSMCB، l برابر تعداد گره‌ها در بزرگترین اجتماع است. در CNCG نیز l برابر تعداد تکرار الگوریتم و d برابر میانگین درجه گره است

کارایی زمانی یک معضل اساسی در مجموعه داده‌های بزرگ محسوب می‌شود. همچنین، معیار «گسترش تأثیر (اثربخشی)» برای ارزیابی اثربخشی (دقت) الگوریتم ها مورد استفاده قرار می‌گیرد. به‌عبارت‌دیگر، کیفیت مجموعه سبد S برابر با تعداد گره‌های تحت تأثیر قرارگرفته توسط مجموعه سبد S پس از اتمام فرآیند شبیه‌سازی انتشار است. الگوریتمی که دارای مقدار گسترش تأثیر بیشتر باشد به‌عنوان الگوریتم اثربخشتر انتخاب خواهد شد. محققین در سال‌های اخیر تلاش‌های بسیاری به منظور بهبود همزمان این دو معیار کلیدی داشته‌اند. در این مقاله، تحقیقات برجسته‌ی ارائه‌شده در این حوزه از نظر این دو معیار بررسی خواهند شد.

۲.۴. مقایسه کلی الگوریتم‌های موجود

الگوریتم‌های حریصانه‌ی بررسی‌شده در بخش ۱.۳ ازجمله الگوریتم حریصانه‌ی کمپ[۴] و CELF [۳۵] دارای بالاترین میزان اثربخشی نسبت به روش‌های مکاشفه‌ای بودند. اما معضل اساسی آنها سربار محاسباتی زیاد در گراف‌های بزرگ بود. الگوریتم‌های NewGreedy [۳۹]، MixGreedy [۳۹]، StaticGreedy(DU) [۴۰]، PrunedMC [۴۱]، CELF++ [۳۶] و UBLF [۳۸] تلاش های نسبتاً ناموفقی برای افزایش سرعت الگوریتم‌های حریصانه‌ی کمپ و CELF [۳۵] بودند. در ادامه الگوریتم‌های TIM (TIM+) [۴۲] و IMM [۴۳] توانستند که کارایی زمانی الگوریتم حریصانه‌ی کمپ را بهبود دهند، اما باوجود ارائه ضمانت تئوری قادر به کنترل شبکه‌های‌اجتماعی بزرگ در زمان منطقی نبودند.

پیچیدگی زمانی بالای الگوریتم‌های حریصانه و ناکارآمدی آن‌ها در مواجهه با شبکه‌های اجتماعی بسیار بزرگ منجر به ارائه طیف گسترده‌ای از الگوریتم‌های مکاشفه‌ای شد. این دسته از الگوریتم‌ها (مبتنی‌بر مرکزیت درجه، مبتنی‌بر مسیر و مبتنی‌بر رتبه‌بندی) که در بخش ۲.۳ مطالعه شدند در کل زمان اجرای بهتری نسبت به الگوریتم‌های حریصانه داشتند. بااین‌وجود، این الگوریتم‌ها ازلحاظ تئوری ضعیف بودند و اثربخشی کمتری داشتند. البته بعضی از آنها مانند IRIE [۵۰] و PMIA [۴۶] به میزان اثربخشی قابل قبولی دست یافتند. اما باز هم در عصر کلان داده‌ها ناکارمد هستند.

در سال های اخیر که اندازه شبکه های اجتماعی بصورت افسارگسیخته در حال بزرگ شدن است، محققان به رویکرد های مبتنی بر شناسایی اجتماعات برای افزایش کارایی زمان و اثربخشی روی آورده‌اند. تحلیل آنها بر این نکته استوار است که گسترش تأثیر یک گره در اجتماعش می‌تواند تخمین خوبی از تأثیر آن در کل شبکه‌ی‌اجتماعی باشد. از طرف دیگر تقسیم یک شبکه‌ی‌اجتماعی بزرگ به اجتماعات کوچکتر باعث بهبود قابل‌ملاحظه‌ای در کارایی زمانی خواهدشد. مطالعات انجام‌شده درگیر چالش‌های زیر بودند.

۱. الگوریتم شناسایی اجتماعات: هر چه تعداد اجتماعات بدست‌آمده بیشتر و تعداد گره‌های اجتماعات متوازن‌تر باشد، کارایی زمانی بهتری بدست خواهد آمد. از طرف دیگر اگر مکانیسم کشف اجتماعات با مکانیسم مدل‌های انتشار اطلاعات هم‌راستا باشد، میزان گسترش تأثیر افزایش پیدا خواهد کرد. به‌عنوان مثال الگوریتم‌هایی مانند CGA [۵۸]، INCIM [۶۲]، IM-LPA [۶۴]، CI2 [۷۲] و HybridIM [۷۵] که از روش های LPA [۲۷] و SLPA [۲۸] به‌منظور شناسایی اجتماعات بهره برده‌اند، اثربخشی نسبتاً خوبی داشته‌اند. البته استفاده از الگوریتم Louvain [۲۲] (مبتنی بر بهینه‌سازی مدولاریته) در الگوریتم‌هایی مانند CoFIM [۶۶]، IMPC [۷۷]، PHG [۷۸]، TI-SC [۸۵] و CAOM [۶۸] نیز منجر به نتایج قابل‌قبولی در زمان اجرا و اثربخشی شده‌است.

۲. الگوریتم محاسبه میزان گسترش تأثیر: نحوه محاسبه میزان گسترش تأثیر گره‌ها در اجتماعات چالش بزرگی به حساب می‌آید. مدل‌هایی مانند ComPath [۶۱] و TOPSIS [۷۴] که از روش‌های مبتنی بر مرکزیت استفاده کرده‌اند، علیرغم زمان اجرای قابل‌قبول اثربخشی کمتری داشتند. در طرف مقابل مدل‌هایی مانند INCIM [۶۲]،BCRIM [۷۳]، CI2 [۷۲] و HybridIM [۷۵] که از روش های حریصانه بهره برده‌اند با وجود تولید گره های تأثیرگذار باکیفیت، کارایی زمانی را از دست داده اند.

۳. انتخاب اجتماعات کلیدی: بعضی از الگوریتم‌ها گره‌های تأثیرگذار را از دل همه‌ی اجتماعات بصورت کورکورانه انتخاب می‌کنند. ازآن‌جاکه همه‌ی اجتماعات چک خواهند شد، اثربخشی قابل قبولی بدست آمده ولی زمان اجرای آنها زیاد می شود. در مقابل مدل‌هایی مانند CAOM [۶۸] و CIMA [۶۷] ابتدا اجتماعات بااهمیت‌تر را انتخاب کرده و سپس از داخل آنها به انتخاب گره‌های تأثیرگذار می‌پردازند. این سیاست منجر به افزایش سرعت خواهد شد. البته ازآن جا که احتمال دارد که گره های تأثیر گذار در اجتماعات نادیده‌گرفته‌شده قرارگرفته‌باشند، اثربخشی کم می‌شود. ثانیاً پیداکردن معیار و سنجه‌ای مناسب برای تعیین درجه اهمیت اجتماعات دشوار است و باعث افت کارایی می‌شود. در مجموع الگوریتم های CoFIM [۶۶]، IM-LPA [۶۴] و CAOM [۶۸] از شاخص‌ترین (ازنظر دو معیار اثربخشی و زمان اجرا) الگوریتم‌های این دسته به شمار می‌آیند.

۵. تحقیقات آینده

محققان زیادی در حوزه تجزیه و تحلیل شبکه های اجتماعی، مسئله‌ی بیشینه‌سازی تأثیر را مورد مطالعه قرار دادند. در این حین چالش‌های متعددی مطرح شده است. اگرچه برخی از این چالش‌ها توسط محققان حل شده است، بسیاری دیگر به اندازه کافی پوشش داده نشده‌اند یا باید بیشتر مورد مطالعه قرار گیرند. این بخش برخی از چالش ها و مسیرهای بعدی برای تحقیق را نشان می‌دهد.

• شبکه‌های ناهمگن[120]: تنوع شبکه‌های اجتماعی یکی از چالش های معروف است. اکثر مدل‌های پیشنهادشده، بر مبنای یک شبکه همگن طراحی شده‌اند. با این حال در دنیای واقعی، ما با شبکه‌های ناهمگن روبرو هستیم که گره‌ها و یال‌ها از انواع مختلفی هستند. در واقع، هر شبکه طیف وسیعی از فعالیت‌های کاربر و معانی خود را پوشش می‌دهد. فرض همگنی شبکه ممکن است منجر به از دست رفتن اطلاعات معنایی کلیدی و عدم درک کامل اطلاعات شود. بنابراین، شناسایی گره‌های تأثیرگذار تحت شبکه‌های ناهمگن [۸۸–۹۰] می‌تواند در تحقیقات آینده مورد مطالعه‌ی بیشتر قرار گیرد.

• شبکه های پویا[121]: همانطور که مشخص است، پویایی یک موضوع اجتناب‌ناپذیر در شبکه‌های اجتماعی است. در حقیقت، اکثر مدل‌های موجود مدل‌هایی هستند که روی شبکه‌های ایستا کار می‌کنند. حتی مدل‌هایی که به شبکه‌های پویا علاقه دارند، در واقع هنوز روی شبکه‌های ایستا کار می‌کنند. زیرا هر بار از وضعیت فعلی شبکه اجتماعی عکس‌برداری[122] کرده و پویایی شبکه را در زمان واقعی تحلیل نمی‌کنند. به عبارت دیگر، یک شبکه اجتماعی واقعی می‌تواند خیلی سریع تکامل یابد. ایجاد یک مدل موثر و پویا برای مدیریت این تکامل سریع دشوار است. از آنجا که تعداد الگوریتم‌های پویای پیشنهادی محدود است [۹۱،۹۲]، تحقیقات بیشتری به منظور توسعه الگوریتم‌های پویا نیاز است.

• شبکه های چندگانه[123]: امروزه، افراد تمایل دارند که به طور همزمان به چندین شبکه اجتماعی بپیوندند و اطلاعات به طور همزمان در بین آنها منتشر می‌شود، در نتیجه ابعادی جدید برای مسئله بیشینه سازی تأثیر در شبکه‌های اجتماعی چندگانه ایجاد می‌شود [۹۳]. سیستم‌های شبکه‌ای مدرن دارای ساختارهای پیچیده‌تر و لایه‌های چندگانه هستند، که برخورد با این شبکه‌ها را برای اجرای تکنیک های تعیین افراد تأثیر گذار دشوار می‌کند. بنابراین، یافتن گره‌های تأثیرگذار در این شبکه‌های چندگانه واقعی مبحثی باز برای تحقیق است.

• شبکه‌های علامت‌دار[124]: اکثر مطالعات انجام‌شده در مسئله بیشینه‌سازی تأثیر، بر روی شبکه‌های اجتماعی بی‌علامت تمرکز کرده و از قطبی بودن روابط بین کاربران چشم پوشی می‌کنند. در حقیقت، روابط می‌توانند به روابط مثبت (به عنوان مثال، تاییدکردن[125]) و روابط منفی (به عنوان مثال، عدم تایید[126]) تقسیم شوند. شبکه‌های اجتماعی که به طور همزمان شامل روابط مثبت و منفی کاربران هستند شبکه های اجتماعی علامت‌دار نامیده می‌شوند. در مسئله بیشینه‌سازی تأثیر، نادیده گرفتن قطب‌های رابطه‌ی بین کاربران و رفتار با شبکه‌های اجتماعی علامت‌دار به عنوان شبکه‌های بی‌علامت، به طور تقریبی منجر به برآورد بیش از حد تأثیر مثبت می‌شود که در برنامه‌های کاربردی انعکاس بدی خواهد داشت. از آنجا که تعداد کمی از تحقیقات در زمینه شبکه‌های اجتماعی علامت‌دار وجود دارد [۹۴،۹۵]، این موضوع می‌تواند موضوعی جالب برای مطالعات بیشتر باشد.

• روش‌های موازی[127]: همانطور که در بخش‌های قبلی گفته شد، چالش‌های اصلی رویکردهای موجود در مسئله بیشینه سازی تاثیر مربوط به کارآیی زمان و مقیاس پذیری است. این معظل را می توان با طراحی مدل‌های موازی و توزیع‌شده کنترل کرد [۹۶،۹۷]. از این رو، شناسایی گره‌های تأثیرگذار را می‌توان تحت مدل‌های توزیع شده و موازی مورد مطالعه‌ی بیشترقرار داد.

۶. جمع‌بندی

یافتن تاثیرگذارترین افراد به منظور بیشینه‌سازی تأثیر در شبکه‌های اجتماعی موضوع داغی است که هنوز باید مورد بررسی قرار گیرد. کارآیی زمانی و میزان اثربخشی دو چالش مهم در تولید و تجزیه و تحلیل الگوریتم‌های بیشینه سازی تأثیر است. در عصر کلان داده‌ها و با رشد بدون توقف اندازه شبکه‌های‌اجتماعی، الگوریتم‌های مبتنی بر شناسایی اجتماعات می‌توانند راه‌حلی کارا جهت کنترل معضل زمان در مسئله‌ی بیشینه سازی تأثیر باشند. این مقاله چارچوبی را برای محققان فراهم آورده است تا همزمان با تمرکز بر رویکردهای مبتنی بر اجتماع، به طور مختصر با رویکردهای برجسته در این زمینه آشنا شوند. در این مقاله الگوریتم‌ها و روش‌های پیشنهادی بررسی و معیارهای اثربخشی و کارایی زمانی آنها را تجزیه و تحلیل شده است. همچنین نقاط قوت و ضعف الگوریتم‌ها و مدل‌های پیشنهادی را توضیح داده شده است. در نهایت، چندین زمینه برای تحقیقات آینده در این زمینه پیشنهاد شد. انتظار می‌رود که این مقاله به عنوان مرجع مفیدی برای محققان علاقه مند باشد تا مسئله بیشینه‌سازی تأثیر را بیشتر و عمیق‌تر مورد بررسی قرار دهند.

مراجع

[1] S. Banerjee, M. Jenamani, D.K. Pratihar, A survey on influence maximization in a social network, Knowl. Inf. Syst. 62 (2020) 3417–3455. https://doi.org/10.1007/s10115-020-01461-4.

[2] Y. Li, J. Fan, Y. Wang, K.-L.L. Tan, Influence Maximization on Social Graphs: A Survey, IEEE Trans. Knowl. Data Eng. 30 (2018) 1852–1872. https://doi.org/10.1109/TKDE.2018.2807843.

[120] Heterogenous networks
[121] Dynamics networks
[122] snapshot
[123] Multiplex networks
[124] Signed networks
[125] like
[126] dislike
[127] Parallel methods

[3] P. Domingos, M. Richardson, Mining the network value of customers, in: Proc. Seventh ACM SIGKDD Int. Conf. Knowl. Discov. Data Min., Association for Computing Machinery (ACM), 2001: pp. 57–66. https://doi.org/10.1145/502512.502525.

[4] D. Kempe, J. Kleinberg, É. Tardos, Maximizing the spread of influence through a social network, in: Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min., ACM Press, New York, New York, USA, 2003: pp. 137–146. https://doi.org/10.1145/956750.956769.

[5] S. Peng, Y. Zhou, L. Cao, S. Yu, J. Niu, W. Jia, Influence analysis in social networks: A survey, J. Netw. Comput. Appl. 106 (2018) 17–32. https://doi.org/10.1016/j.jnca.2018.01.005.

[6] M. Jaouadi, L. Ben Romdhane, Influence maximization problem in social networks: An overview, in: Proc. IEEE/ACS Int. Conf. Comput. Syst. Appl. AICCSA, IEEE Computer Society, 2019. https://doi.org/10.1109/AICCSA47632.2019.9035366.

[7] S. Peng, S. Yu, P. Mueller, Social networking big data: Opportunities, solutions, and challenges, Futur. Gener. Comput. Syst. 86 (2018) 1456–1458. https://doi.org/10.1016/j.future.2018.05.040.

[8] K. Li, L. Zhang, H. Huang, Social Influence Analysis: Models, Methods, and Evaluation, Engineering. 4 (2018) 40–46. https://doi.org/10.1016/j.eng.2018.02.004.

[9] N. Sumith, B. Annappa, S. Bhattacharya, Influence maximization in large social networks: Heuristics, models and parameters, Futur. Gener. Comput. Syst. 89 (2018) 777–790. https://doi.org/10.1016/j.future.2018.07.015.

[10] M. Han, Y. Li, Influence analysis: A survey of the state-of-the-art, Math. Found. Comput. 1 (2018) 201–253. https://doi.org/10.3934/mfc.2018010.

[11] N. Hafiene, W. Karoui, L. Ben Romdhane, Influential nodes detection in dynamic social networks: A survey, Expert Syst. Appl. 159 (2020) 113642. https://doi.org/10.1016/j.eswa.2020.113642.

[12] B.S. Khan, M.A. Niazi, Network Community Detection: A Review and Visual Survey, ArXiv. (2017). http://arxiv.org/abs/1708.00977 (accessed December 26, 2020).

[13] S. Gupta, D.P. Singh, Recent trends on community detection algorithms: A survey, Mod. Phys. Lett. B. (2020). https://doi.org/10.1142/S0217984920504084.

[14] E.R. Barnes, An Algorithm for Partitioning the Nodes of a Graph, Http://Dx.Doi.Org/10.1137/0603056. 3 (2006) 541–550. https://doi.org/10.1137/0603056.

[15] B.W. Kernighan, S. Lin, An Efficient Heuristic Procedure for Partitioning Graphs, Bell Syst. Tech. J. 49 (1970) 291–307. https://doi.org/10.1002/J.1538-7305.1970.TB01770.X.

[16] J. MacQueen, Some methods for classification and analysis of multivariate observations, Https://Doi.Org/. 5.1 (1967) 281–298. (accessed October 20, 2021).

[17] M. Fiedler, Algebraic connectivity of graphs, Czechoslov. Math. J. 23 (1973) 298–305. http://eudml.org/doc/12723.

[18] M. Girvan, M.E.J. Newman, Community structure in social and biological networks, Proc. Natl. Acad. Sci. U. S. A. 99 (2001) 7821–7826. https://doi.org/10.1073/pnas.122653799.

[19] M.E.J. Newman, Modularity and community structure in networks, Proc. Natl. Acad. Sci. U. S. A. 103 (2006) 8577–8582. https://doi.org/10.1073/pnas.0601602103.

[20] M.E.J. Newman, Fast algorithm for detecting community structure in networks, Phys. Rev. E - Stat. Physics, Plasmas, Fluids, Relat. Interdiscip. Top. 69 (2003) 5. https://doi.org/10.1103/physreve.69.066133.

[21] A. Clauset, M.E.J. Newman, C. Moore, Finding community structure in very large networks, Phys. Rev. E - Stat. Physics, Plasmas, Fluids, Relat. Interdiscip. Top. 70 (2004) 6. https://doi.org/10.1103/PhysRevE.70.066111.

[22] V.D. Blondel, J.L. Guillaume, R. Lambiotte, E. Lefebvre, Fast unfolding of communities in large networks, J. Stat. Mech. Theory Exp. 2008 (2008). https://doi.org/10.1088/1742-5468/2008/10/P10008.

[23] R. Guimerà, L.A. Nunes Amaral, Functional cartography of complex metabolic networks, Nat. 2005 4337028. 433 (2005) 895–900. https://doi.org/10.1038/nature03288.

[24] M. Rezapoor Mirsaleh, M. Reza Meybodi, A Michigan memetic algorithm for solving the community detection problem in complex network, Neurocomputing. 214 (2016) 535–545. https://doi.org/10.1016/J.NEUCOM.2016.06.030.

[25] S. Sobolevsky, R. Campari, A. Belyi, C. Ratti, A General Optimization Technique for High Quality Community Detection in Complex Networks, Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys. 90 (2013). https://doi.org/10.1103/physreve.90.012811.

influence maximization in social networks, in: Proc. 20th Int. Conf. Companion World Wide Web, WWW 2011, 2011: pp. 47–48. https://doi.org/10.1145/1963192.1963217.
[37] A. Arora, S. Galhotra, S. Ranu, Debunking the myths of influence maximization: An in-depth benchmarking study, Proc. ACM SIGMOD Int. Conf. Manag. Data. Part F127746 (2017) 651–666. https://doi.org/10.1145/3035918.3035924.
[38] C. Zhou, P. Zhang, W. Zang, L. Guo, On the Upper Bounds of Spread for Greedy Algorithms in Social Network Influence Maximization, IEEE Trans. Knowl. Data Eng. 27 (2015) 2770–2783. https://doi.org/10.1109/TKDE.2015.2419659.
[39] W. Chen, Y. Wang, S. Yang, Efficient influence maximization in social networks, in: Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min., ACM Press, New York, New York, USA, 2009: pp. 199–207. https://doi.org/10.1145/1557019.1557047.
[40] S. Cheng, H. Shen, J. Huang, G. Zhang, X. Cheng, StaticGreedy: Solving the scalability-accuracy dilemma in influence maximization, in: Int. Conf. Inf. Knowl. Manag. Proc., ACM Press, New York, New York, USA, 2013: pp. 509–518. https://doi.org/10.1145/2505515.2505541.
[41] N. Ohsaka, T. Akiba, Y. Yoshida, K. Kawarabayashi, Fast and accurate influence maximization on large networks with pruned Monte-Carlo simulations, in: Proc. Twenty-Eighth AAAI Conf. Artif. Intell., 2014: pp. 138–144. https://dl.acm.org/doi/10.5555/2893873.2893897 (accessed December 27, 2020).
[42] Y. Tang, X. Xiao, Y. Shi, Influence maximization: Near-optimal time complexity meets practical efficiency, in: Proc. ACM SIGMOD Int. Conf. Manag. Data, Association for Computing Machinery, New York, New York, USA, 2014: pp. 75–86. https://doi.org/10.1145/2588555.2593670.
[43] Y. Tang, Y. Shi, X. Xiao, Influence maximization in near-linear time: A martingale approach, in: Proc. ACM SIGMOD Int. Conf. Manag. Data, Association for Computing Machinery, New York, New York, USA, 2015: pp. 1539–1554.
https://doi.org/10.1145/2723372.2723734.
[44] W. Chen, Y. Yuan, L. Zhang, Scalable influence maximization in social networks under the linear threshold model, in: Proc. - IEEE Int. Conf. Data Mining, ICDM, 2010: pp. 88–97. https://doi.org/10.1109/ICDM.2010.118.
[45] M. Kimura, K. Saito, R. Nakano, H. Motoda, Extracting influential nodes on a social
[26] G. Palla, I. Derényi, I. Farkas, T. Vicsek, Uncovering the overlapping community structure of complex networks in nature and society, Nature. 435 (2005) 814–818. https://doi.org/10.1038/nature03607.
[27] U.N. Raghavan, R. Albert, S. Kumara, Near linear time algorithm to detect community structures in large-scale networks, Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys. 76 (2007). https://doi.org/10.1103/PhysRevE.76.036106.
[28] J. Xie, B.K. Szymanski, X. Liu, SLPA: Uncovering overlapping communities in social networks via a speaker-listener interaction dynamic process, in: Proc. - IEEE Int. Conf. Data Mining, ICDM, 2011: pp. 344–349. https://doi.org/10.1109/ICDMW.2011.154.
[29] J.F. Douglas, Random walks and random environments, vol. 2, random environments, J. Stat. Phys. 87 (1997) 961–962. https://doi.org/10.1007/bf02181260.
[30] L. Page, L. Page, S. Brin, R. Motwani, T. Winograd, The PageRank Citation Ranking: Bringing Order to the Web, -. (1998). http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.38.5427 (accessed September 25, 2020).
[31] M. Rosvall, C.T. Bergstrom, Maps of random walks on complex networks reveal community structure, Proc. Natl. Acad. Sci. U. S. A. 105 (2007) 1118–1123. https://doi.org/10.1073/pnas.0706851105.
[32] C. Tantipathananandh, T. Berger-Wolf, D. Kempe, A framework for community identification in dynamic social networks, Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min. (2007) 717–726. https://doi.org/10.1145/1281192.1281269.
[33] L. Hajdu, M. Krész, A. Bóta, Evaluating the role of community detection in improving influence maximization heuristics, Soc. Netw. Anal. Min. 2021 111. 11 (2021) 1–11. https://doi.org/10.1007/S13278-021-00804-5.
[34] Z. Wang, C. Sun, J. Xi, X. Li, Influence maximization in social graphs based on community structure and node coverage gain, Futur. Gener. Comput. Syst. 118 (2021) 327–338. https://doi.org/10.1016/j.future.2021.01.025.
[35] J. Leskovec, A. Krause, C. Guestrin, C. Faloutsos, J. Vanbriesen, N. Glance, Cost-effective outbreak detection in networks, in: Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min., ACM Press, New York, New York, USA, 2007: pp. 420–429. https://doi.org/10.1145/1281192.1281239.
[36] A. Goyal, W. Lu, L.V.S. Lakshmanan, CELF++: Optimizing the greedy algorithm for

Deterministic Linear Threshold Model, Knowledge-Based Syst. 161 (2018) 111–123. https://doi.org/10.1016/j.knosys.2018.07.040.
[55] Q. He, X. Wang, Z. Lei, M. Huang, Y. Cai, L. Ma, TIFIM: A Two-stage Iterative Framework for Influence Maximization in Social Networks, Appl. Math. Comput. 354 (2019) 338–352. https://doi.org/10.1016/J.AMC.2019.02.056.
[56] W. Liu, Y. Li, X. Chen, J. He, Maximum likelihood-based influence maximization in social networks, Appl. Intell. 2020 5010. 50 (2020) 3487–3502. https://doi.org/10.1007/S10489-020-01747-8.
[57] C. Wang, Q. Shi, W. Xian, Y. Feng, C. Chen, Efficient diversified influence maximization with adaptive policies, Knowledge-Based Syst. 213 (2021) 106692. https://doi.org/10.1016/J.KNOSYS.2020.106692.
[58] Y. Wang, G. Cong, G. Song, K. Xie, Community-based Greedy Algorithm for Mining top-K Influential Nodes in Mobile Social Networks, in: Proc. 16th ACM SIGKDD Int. Conf. Knowl. Discov. Data Min., ACM, New York, NY, USA, 2010: pp. 1039–1048. https://doi.org/10.1145/1835804.1835935.
[59] Y.C. Chen, W.Y. Zhu, W.C. Peng, W.C. Lee, S.Y. Lee, CIM: Community-based influence maximization in social networks, ACM Trans. Intell. Syst. Technol. 5 (2014) 1–31. https://doi.org/10.1145/2532549.
[60] H. Li, S.S. Bhowmick, A. Sun, J. Cui, Conformity-aware influence maximization in online social networks, VLDB J. 24 (2015) 117–141. https://doi.org/10.1007/s00778-014-0366-x.
[61] K. Rahimkhani, A. Aleahmad, M. Rahgozar, A. Moeini, A fast algorithm for finding most influential people based on the linear threshold model, Expert Syst. Appl. 42 (2015) 1353–1361.
https://doi.org/10.1016/j.eswa.2014.09.037.
[62] A. Bozorgi, H. Haghighi, M. Sadegh Zahedi, M. Rezvani, INCIM: A community-based algorithm for influence maximization problem under the linear threshold model, Inf. Process. Manag. 52 (2016) 1188–1199. https://doi.org/10.1016/j.ipm.2016.05.006.
[63] M. Halappanavar, A. V. Sathanur, A.K. Nandi, Accelerating the mining of influential nodes in complex networks through community detection, in: 2016 ACM Int. Conf. Comput. Front. - Proc., Association for Computing Machinery, Inc, New York, New York, USA, 2016: pp. 64–71. https://doi.org/10.1145/2903150.2903181.
[64] Y. Zhao, S. Li, F. Jin, Identification of influential nodes in social networks with network for information diffusion, Data Min. Knowl. Discov. 20 (2010) 70–97. https://doi.org/10.1007/s10618-009-0150-5.
[46] W. Chen, C. Wang, Y. Wang, Scalable influence maximization for prevalent viral marketing in large-scale social networks, in: Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min., ACM Press, New York, New York, USA, 2010: pp. 1029–1038. https://doi.org/10.1145/1835804.1835934.
[47] A. Goyal, W. Lu, L.V.S. Lakshmanan, SIMPATH: An Efficient Algorithm for Influence Maximization under the Linear Threshold Model, in: 2011 IEEE 11th Int. Conf. Data Min., IEEE, 2011: pp. 211–220. https://doi.org/10.1109/ICDM.2011.132.
[48] W. Lu, X. Xiao, A. Goyal, K. Huang, L.V.S. Lakshmanan, Refutations on "Debunking the Myths of Influence Maximization: An In-Depth Benchmarking Study," (2017). https://arxiv.org/abs/1705.05144v3 (accessed August 21, 2021).
[49] R. Narayanam, Y. Narahari, A shapley value-based approach to discover influential nodes in social networks, IEEE Trans. Autom. Sci. Eng. 8 (2011) 130–147. https://doi.org/10.1109/TASE.2010.2052042.
[50] K. Jung, W. Heo, W. Chen, IRIE: Scalable and robust influence maximization in social networks, in: Proc. - IEEE Int. Conf. Data Mining, ICDM, 2012: pp. 918–923. https://doi.org/10.1109/ICDM.2012.79.
[51] Q. Liu, B. Xiang, E. Chen, H. Xiong, F. Tang, J.X. Yu, Influence maximization over large-scale social networks: A bounded linear approach, in: CIKM 2014 - Proc. 2014 ACM Int. Conf. Inf. Knowl. Manag., Association for Computing Machinery, Inc, New York, New York, USA, 2014: pp. 171–180. https://doi.org/10.1145/2661829.2662009.
[52] S. Cheng, H.-W. Shen, J. Huang, W. Chen, X.-Q. Cheng, IMRank: Influence Maximization via Finding Self-Consistent Ranking, SIGIR 2014 - Proc. 37th Int. ACM SIGIR Conf. Res. Dev. Inf. Retr. (2014) 475–484. http://arxiv.org/abs/1402.3939 (accessed September 24, 2020).
[53] S. Galhotra, A. Arora, S. Roy, Holistic Influence Maximization: Combining Scalability and Efficiency with Opinion-Aware Models, Proc. ACM SIGMOD Int. Conf. Manag. Data. 26-June-20 (2016) 743–758. https://doi.org/10.1145/2882903.2882929.
[54] F. Gursoy, D. Gunnec, Influence maximization in social networks under

decision making for influence maximization, Comput. Ind. Eng. 120 (2018) 234–250. https://doi.org/10.1016/j.cie.2018.04.049.
[75] Y.Y. Ko, K.J. Cho, S.W. Kim, Efficient and effective influence maximization in social networks: A hybrid-approach, Inf. Sci. (Ny). 465 (2018) 144–161. https://doi.org/10.1016/j.ins.2018.07.003.
[76] X. Li, X. Cheng, S. Su, C. Sun, Community-based seeds selection algorithm for location aware influence maximization, Neurocomputing. 275 (2018) 1601–1613. https://doi.org/10.1016/j.neucom.2017.10.007.
[77] J. Shang, H. Wu, S. Zhou, J. Zhong, Y. Feng, B. Qiang, IMPC: Influence maximization based on multi-neighbor potential in community networks, Phys. A Stat. Mech. Its Appl. 512 (2018) 1085–1103. https://doi.org/10.1016/j.physa.2018.08.045.
[78] Q. Liqing, J. Wei, Y. Jinfeng, F. Xin, Z. Shuang, PHG: A Three-phase Algorithm for Influence Maximization based on Community Structure, IEEE Access. (2019) 1–1. https://doi.org/10.1109/ACCESS.2019.2912628.
[79] S.S. Singh, A. Kumar, K. Singh, B. Biswas, C2IM: Community based context-aware influence maximization in social networks, Phys. A Stat. Mech. Its Appl. 514 (2019) 796–818. https://doi.org/10.1016/J.PHYSA.2018.09.142.
[80] H. Huang, H. Shen, Z. Meng, H. Chang, H. He, Community-based influence maximization for viral marketing, Appl. Intell. 49 (2019) 2137–2150. https://doi.org/10.1007/s10489-018-1387-8.
[81] S. Banerjee, M. Jenamani, D.K. Pratihar, ComBIM: A community-based solution approach for the Budgeted Influence Maximization Problem, Expert Syst. Appl. 125 (2019) 1–13. https://doi.org/10.1016/j.eswa.2019.01.070.
[82] S.S. Singh, K. Singh, A. Kumar, B. Biswas, CoIM: Community-Based Influence Maximization in Social Networks, in: Commun. Comput. Inf. Sci., Springer Verlag, 2019: pp. 440–453. https://doi.org/10.1007/978-981-13-3143-5_36.
[83] Y. Atif, K. Al-Falahi, T. Wangchuk, B. Lindström, A fuzzy logic approach to influence maximization in social networks, J. Ambient Intell. Humaniz. Comput. 11 (2020) 2435–2451. https://doi.org/10.1007/s12652-019-01286-2.
[84] Z. Zhang, X. Li, C. Gan, Identifying influential nodes in social networks via community structure and influence distribution difference, Digit. Commun. Networks. (2020). https://doi.org/10.1016/j.dcan.2020.04.011.
[85] H.A. Beni, A. Bouyer, TI-SC: top-k community structure based on label propagation, Neurocomputing. 210 (2016) 34–44. https://doi.org/10.1016/j.neucom.2015.11.125.
[65] M. Jaouadi, L. Ben Romdhane, DIN: An efficient algorithm for detecting influential nodes in social graphs using network structure and attributes, in: Proc. IEEE/ACS Int. Conf. Comput. Syst. Appl. AICCSA, IEEE Computer Society, 2016. https://doi.org/10.1109/AICCSA.2016.7945698.
[66] J. Shang, S. Zhou, X. Li, L. Liu, H. Wu, CoFIM: A community-based framework for influence maximization on large-scale networks, Knowledge-Based Syst. 117 (2017) 88–100. https://doi.org/10.1016/j.knosys.2016.09.029.
[67] H. Huang, H. Shen, Z. Meng, Community-based influence maximization in attributed networks, Appl. Intell. 50 (2020) 354–364. https://doi.org/10.1007/s10489-019-01529-x.
[68] Q. He, X. Wang, F. Mao, J. Lv, Y. Cai, M. Huang, Q. Xu, CAOM: A community-based approach to tackle opinion maximization for social networks, Inf. Sci. (Ny). 513 (2020) 252–269. https://doi.org/10.1016/j.ins.2019.10.064.
[69] M. Hosseini-Pozveh, K. Zamanifar, A.R. Naghsh-Nilchi, A community-based approach to identify the most influential nodes in social networks, J. Inf. Sci. 43 (2017) 204–220. https://doi.org/10.1177/0165551515621005.
[70] C. Wang, L. Deng, G. Zhou, M. Jiang, A global optimization algorithm for target set selection problems, Inf. Sci. (Ny). 267 (2014) 101–118. https://doi.org/10.1016/j.ins.2013.09.033.
[71] J. Xie, B.K. Szymanski, LabelRank: A stabilized label propagation algorithm for community detection in networks, in: Proc. 2013 IEEE 2nd Int. Netw. Sci. Work. NSW 2013, 2013: pp. 138–143. https://doi.org/10.1109/NSW.2013.6609210.
[72] A. Bozorgi, S. Samet, J. Kwisthout, T. Wareham, Community-based influence maximization in social networks under a competitive linear threshold model, Knowledge-Based Syst. 134 (2017) 149–158. https://doi.org/10.1016/j.knosys.2017.07.029.
[73] F. Ye, J. Liu, C. Chen, G. Ling, Z. Zheng, Y. Zhou, Identifying Influential Individuals on Large-Scale Social Networks: A Community Based Approach, IEEE Access. 6 (2018) 47240–47257. https://doi.org/10.1109/ACCESS.2018.2866981.
[74] M. Jalayer, M. Azheian, M. Agha Mohammad Ali Kermani, A hybrid algorithm based on community detection and multi attribute

Trans. Knowl. Data Eng. 29 (2017) 359–372. https://doi.org/10.1109/TKDE.2016.2620141.
[92] W. Li, K. Zhong, J. Wang, D. Chen, A dynamic algorithm based on cohesive entropy for influence maximization in social networks, Expert Syst. Appl. 169 (2021) 114207. https://doi.org/10.1016/j.eswa.2020.114207.
[93] A. Kuhnle, M.A. Alim, X. Li, H. Zhang, M.T. Thai, Multiplex Influence Maximization in Online Social Networks With Heterogeneous Diffusion Models, IEEE Trans. Comput. Soc. Syst. 5 (2018) 418–429. https://doi.org/10.1109/TCSS.2018.2813262.
[94] D. Li, C. Wang, S. Zhang, G. Zhou, D. Chu, C. Wu, Positive influence maximization in signed social networks based on simulated annealing, Neurocomputing. 260 (2017) 69–78. https://doi.org/10.1016/j.neucom.2017.03.003.
[95] M. Hosseini-Pozveh, K. Zamanifar, A.R. Naghsh-Nilchi, Assessing information diffusion models for influence maximization in signed social networks, Expert Syst. Appl. (2018). https://doi.org/10.1016/j.eswa.2018.07.064.
[96] J. Kim, S.K. Kim, H. Yu, Scalable and parallelizable processing of influence maximization for large-scale social networks?, in: Proc. - Int. Conf. Data Eng., 2013: pp. 266–277. https://doi.org/10.1109/ICDE.2013.6544831.
[97] S. Kim, D. Kim, J. Oh, J.H. Hwang, W.S. Han, W. Chen, H. Yu, Scalable and parallelizable influence maximization with Random Walk Ranking and Rank Merge Pruning, Inf. Sci. (Ny). 415–416 (2017) 171–189. https://doi.org/10.1016/j.ins.2017.06.018.
influential nodes selection based on community detection and scoring criteria in social networks, J. Ambient Intell. Humaniz. Comput. 1 (2020) 3. https://doi.org/10.1007/s12652-020-01760-2.
[86] M.M.D. Khomami, A. Rezvanian, M.R. Meybodi, A. Bagheri, CFIN: A community-based algorithm for finding influential nodes in complex social networks, J. Supercomput. (2020) 1–30. https://doi.org/10.1007/s11227-020-03355-2.
[87] A. Ahmad, T. Ahmad, A. Bhatt, HWSMCB: A community-based hybrid approach for identifying influential nodes in the social network, Phys. A Stat. Mech. Its Appl. 545 (2020) 123590.
https://doi.org/10.1016/j.physa.2019.123590.
[88] M. Haque, D.S. Banerjee, Accelerating influence maximization using heterogeneous algorithms, J. Supercomput. 76 (2019) 4747–4769. https://doi.org/10.1007/S11227-019-03061-8.
[89] S. Molaei, R. Farahbakhsh, M. Salehi, N. Crespi, Identifying influential nodes in heterogeneous networks, Expert Syst. Appl. 160 (2020) 113580. https://doi.org/10.1016/j.eswa.2020.113580.
[90] M.M. Keikha, M. Rahgozar, M. Asadpour, M.F. Abdollahi, Influence maximization across heterogeneous interconnected networks based on deep learning, Expert Syst. Appl. 140 (2020) 112905.
https://doi.org/10.1016/j.eswa.2019.112905.
[91] G. Song, Y. Li, X. Chen, X. He, J. Tang, Influential Node Tracking on Dynamic Social Network: An Interchange Greedy Approach, IEEE