Archive of SID

بررسی تاثیر روش‌های کشف اجتماعات در شناسایی گره‌های تاثیرگذار و بیشینه‌سازی اثر در شبکه‌های اجتماعی

یاسمین جلالی[۱]، لیلا اسماعیلی[۲]، سید علیرضا هاشمی گلپایگانی[۳*]، مریم شهسواری[۴]

[۱] دانشجوی کارشناسی ارشد، دانشکده مهندسی کامپیوتر و فناوری اطلاعات، دانشگاه صنعتی امیرکبیر، تهران، ایران، Yasmin.jalali@aut.ac.ir

[۲] دانشجوی دکتری، دانشکده مهندسی کامپیوتر و فناوری اطلاعات، دانشگاه صنعتی امیرکبیر، تهران، ایران، Leila.esmaeili@aut.ac.ir

[۳] استادیار و عضو هیئت علمی، دانشکده مهندسی کامپیوتر و فناوری اطلاعات، دانشگاه صنعتی امیرکبیر، تهران، ایران، Sa.hashemi@aut.ac.ir

[۴] کارشناس ارشد، دانشکده مهندسی کامپیوتر و فناوری اطلاعات، دانشگاه صنعتی امیرکبیر، تهران، ایران، Maryamshahsavari@aut.ac.ir

چکیده

شناسایی گره‌های (افراد) تاثیرگذار در ساختارهای شبکه اجتماعی با هدف بیشیه‌سازی اثر یکی از مسائل پرکاربرد در حوزه تحلیل شبکه-های اجتماعی است. این مسئله به ویژه در حوزه تجارت اجتماعی به دلیل کاربرد آن در بازاریابی و انتشار تبلیغات کالا و خدمات بسیار مورد توجه می‌باشد. پایه‌ای ترین روش در شناسایی گره‌های تاثیرگذار با هدف بیشینه‌سازی اثر در ساختار شبکه، الگوریتم حریصانه است که پاسخی نزدیک به بهینه فراهم می‌کند؛ مسئله‌ی یافتن گره‌های پرنفوذ یک مسئله NP-سخت می‌باشد، از این رو تلاش‌های بسیاری با هدف کاهش محاسبات آن انجام شده است. از جمله این روش‌ها، شناسایی جوامع همپوشان و شروع انتشار از گره‌های همپوشان انتخاب شده در اجتماعات می‌باشد. در این مقاله تاثیر روش های کشف اجتماعات همپوشان در شناسایی گره های تاثیرگذار و بیشینه سازی اثر در شبکه های اجتماعی کوچک مورد بررسی قرار گرفته است. همچنین با توجه به مفروض بودن ساختار کلی شبکه اجتماعی و مدل انتشار، نه تنها بیشینه‌سازی تاثیر در کل شبکه، بلکه پوشش حداکثری اجتماعات در شبکه نیز مورد توجه است.

واژه‌های کلیدی

تحلیل شبکه‌های اجتماعی، کشف اجتماعات، اجتماعات همپوشان، حداکثرسازی اثر، انتشار.

۱- مقدمه

یافتن رئوس تأثیرگذار در یک شبکه‌ی اجتماعی کاربردهای عملی زیادی در بازاریابی، سیاست و حتی کنترل بیماری‌ها دارد. در چنین شبکه‌هایی، رئوس غالبا قادر به اثرگذاری بر روی رئوس همسایه‌ی خود هستند و یک رأسِ تحت تأثیر می‌تواند یک رفتار یا ویژگی را از رئوس همسایه‌ی خود کسب کند. به عنوان مثال، شخصی که با افرادی که از یک وب‌سایت خبری خاص بازدید می‌کنند دوست است، احتمال دارد که شروع به بازدید آن وب‌سایت کند. بعضی از رئوس یک شبکه از بقیه‌ی رئوس تأثیرگذارتر هستند.

یافتن رئوسی که دارای بیش‌ترین تأثیرگذاری هستند مورد توجه مدیران و تحلیل‌گران شبکه‌های اجتماعی است. مدیران بازاریابی ممکن است علاقه‌مند به یافتن افراد تأثیرگذار و ارائه‌ی پیشنهاد تخفیف خرید یا محصولات رایگان خود به آنان باشند، به این امید که این افراد دوستان خود را ترغیب به خرید این محصولات کنند. عاملین سیاسی نیز مایل به یافتن افراد تأثیرگذار برای گسترش انتشار پیام‌های خود می‌باشند [۱]. هم‌چنین، مدل‌های انتشاری که برای تحلیل تأثیر میان افراد به کار می‌روند می‌توانند در شیوع بیماری‌های واگیردار نیز به کار گرفته شوند. رئوس تأثیرگذارِ آلوده‌شده قادر به آلوده‌سازی بخش بزرگ‌تری از جمعیت نسبت به رئوس با تأثیرگذاری کم‌تر هستند. بنابراین، مسئولین بهداشت عمومی می‌توانند با انتشار تلقیح به رئوس تأثیرگذار، نتایج بهتری در درمان بیماری بگیرند.

از طرف دیگر، شبکه‌ها عموماً هم‌جنس نیستند و رئوسی که از یک نوع هستند غالباً در بخش‌هایی محلی از شبکه با یکدیگر یک گروه یا خوشه را تشکیل می‌دهند. از این خوشه‌بندی‌ها می‌توان در جهت افزایش انتشار اثرات بهره برد، به طوری که در ابتدا پرتأثیرترین رأس در هر خوشه برای فعال‌سازی اولیه انتخاب می‌شود و سپس اثر از این رأس انتشار می‌یابد.

* نویسنده مسئول

www.SID.ir

Archive of SID

الگوریتم‌هایی نیز برای یافتن اجتماعات (خوشه‌ها) و شناسایی این گروه‌ها طراحی شده‌اند.

مسئله‌ی بیشینه‌سازی انتشار، انتخاب k رأس دلخواه به عنوان رئوس اولیه برای شروع انتشار در یک شبکه می‌باشد، به طوری که رئوس اولیه انتخاب‌شده به طرز مؤثری بتوانند اطلاعات را میان بقیه‌ی رئوس شبکه منتشر کنند و انتشار اثر در شبکه را بیشینه کنند. در پژوهش‌ها، انتشار اطلاعات تحت یک مدل مفروض شبیه‌سازی می‌شود که روش آبشاری مستقل عمومیت بیشتری دارد. لازم به ذکر است مدل‌های انتشار و روش-های یافتن گره‌های تأثیرگذار مفاهیمی کاملاً مستقل هستند.

یکی از روش‌های یافتن گره‌های تأثیرگذار روش حریصانه‌ی پیشنهادی توسط Kempe و همکارانش [۲] می‌باشد. این روش که اولین بار در سال ۲۰۰۳ برای یافتن گره‌ها پرنفوذ معرفی شد، از زمان اجرای بالایی برخورد است و برای شبکه‌های بزرگ مقیاس‌پذیر نیست. یکی از روش‌های جاگزین روش حریصانه، شناسایی اجتماعات در شبکه‌ها و سپس انتشار اثر است، به طوری که مطمئن باشیم در هر اجتماع یا خوشه حداقل یک رأس اولیه‌ی فعال‌شده وجود دارد. مزیت این روش نسبت به الگوریتم حریصانه مقیاس-پذیر بودن و سرعت بهبودیافته‌ی آن در انتخاب رئوس اولیه می‌باشد.

در این مقاله تاثیر برخی از روش‌های کشف اجتماعات و معیارهای مرکزیت در شناسایی گره‌های تاثیرگذار و بیشینه‌سازی اثر در شبکه‌های اجتماعی کوچک هدف می‌باشد. همچنین با توجه به مفروض بودن ساختار کلی شبکه اجتماعی و مدل انتشار آبشاری مستقل، نه تنها بیشینه‌سازی تاثیر در کل شبکه، بلکه پوشش حداکثری اجتماعات در شبکه نیز مورد توجه است. بر اساس نتایج این تحقیق می‌توان در شرایطی که یافتن k راس با هدف بیشینه‌سازی تاثیر در شبکه‌های کوچک مسئله است، بهترین روش کشف اجتماع را به منظور کاهش محاسبات و میزان تاثیرگذاری قابل قبول انتخاب نمود.

ساختار این مقاله به این صورت تنظیم شده است: در بخش دوم، شرحی از فعالیت‌های مرتبط انجام‌شده در زمینه‌ی یافتن اجتماعات، انتخاب رئوس اولیه و انتشار اثر ارائه شده است. در بخش سوم، بیان دقیق مسئله به همراه جزئیات مورد توجه این مقاله آمده است. مراحل دقیق پیاده‌سازی هر یک از روش‌های خوشه‌بندی و انتشار اثر به کار رفته در این مقاله نیز در بخش چهارم بیان شده‌اند. در بخش پنجم نیز نتایج حاصل از پیاده‌سازی این روش‌ها ارائه شده‌اند. نهایتاً، این مقاله در بخش ششم جمع-بندی و نتیجه‌گیری شده و کارهای آتی را مشخص کرده است.

۲- کارهای پیشین

مقالات بسیاری روش‌های یافتن رئوس تأثیرگذار در شبکه‌های اجتماعی را مطالعه کرده‌اند. پژوهش‌ها معمولاً از روش حریصانه (که یک مسئله‌ی NP-hard است) برای یافتن بهترین k رأس استفاده کرده‌اند که روشی زمان‌بر و پرهزینه است. علاوه بر این، با استفاده از یافتن اجتماعات، تعداد رئوس فعال‌شده‌ی نهایی حداکثر می‌شود. در این مقاله مراجعی که در بازه‌ی زمانی سال ۲۰۰۵ الی ۲۰۱۳ که مرتبط با مفاهیم اصلی مقاله بوده‌اند را بررسی کرده‌ایم که در این قسمت مروری بر این مقالات خواهیم داشت.

در مقاله‌ی [۳]، مسئله‌ی انتخاب همسایگان مؤثر هر گره بررسی شده است. نویسندگان این مقاله به طور تجربی طراحی استراتژی‌های مؤثرتر انتخاب همسایگان به منظور بیشینه‌سازی انتشار اطلاعات را مطالعه کرده‌اند. نتایج این مقاله نشان می‌دهند که یک استراتژی مؤثر در این رابطه، استفاده از اطلاعات درجه‌ی رئوس برای انتشار کوتاه مدت، و استفاده از روش انتخاب تصادفی برای انتشار بلندمدت است. بدین ترتیب، حداقل نیمی از شبکه پوشش داده می‌شود.

در مقاله‌ی [۴]، الگوریتمی به نام الگوریتم حریصانه‌ی مبتنی بر اجتماعات برای یافتن بهترین k رأس تأثیرگذار در شبکه‌های اجتماعی موبایل ارائه شده است. نتایج آزمایش‌های تجربی این مقاله نشان می‌دهند که الگوریتم پیشنهادی سریع‌تر از الگوریتم معمولی حریصانه اجرا می‌کند و خطای آن نیز قابل چشم‌پوشی است. نویسندگان مقاله‌ی [۵] نیز با توجه به مدل انتشار اثر دهان به دهان، حداکثرسازی انتشار یک نوآوری یا رفتار را درون یک شبکه‌ی اجتماعی بررسی کرده‌اند. در این مقاله، یک مدل طبیعی و کلی از انتشار اثر به نام مدل آبشاری کاهشی ارائه شده است که تعمیمی از مدل‌های استفاده‌شده در جوامع اجتماعی و اقتصادی می‌باشد. هم‌چنین، نویسندگان این مقاله مسئله‌ی انتخاب مجموعه‌ی هدف برای فعال‌سازی اولیه را نیز مطالعه کرده‌اند، به طوری که آبشار شروع‌شده از این مجموعه‌ی اولیه‌ی فعال تا حد ممکن بزرگ باشد. در این مقاله با استفاده از پارامتری کردن دوباره‌ی فضای احتمال، اثبات شده است که تعداد موردانتظار رئوس تأثیریافته تابعی یکنواخت و زیرپیمانه‌ای از مجموعه‌ی هدف است.

در مقالات پیشین، فرض شده است که یک رأس در شبکه می‌تواند رئوس دیگر را با احتمال یکنواختی تحت تأثیر قرار دهد. در این مقالات، اثر ترجیحات کاربر در انتشار اثر در نظر گرفته نشده است، با وجود این‌که این عامل، نقش زیادی در دقت نتایج دارد. در مقاله‌ی [۶]، یک الگوریتم جستجوی دو مرحله‌ای برای یافتن تأثیرگذارترین رئوس در یک موضوع خاص پیشنهاد شده است. در گام اول این الگوریتم، از روش فیلترگذاری مشارکتی برای تعیین ترجیح کاربران در یک موضوع استفاده می‌شود. سپس در مرحله‌ی بعدی الگوریتم، از الگوریتم حریصانه برای یافتن بهترین k رأس تأثیرگذار در شبکه استفاده می‌شود. نتایج آزمایش‌های تجربی این مقاله نشان می‌دهند که این الگوریتم می‌تواند با موفقیت بهترین رئوس تأثیرگذار در یک موضوع خاص را بیابد. در جدول ۱، مشخصات این مقالات مشاهده می‌شود.

در [۷] نیز از شناسایی اجتماعات برای شناسایی گره‌های تاثیرگذار با استفاده از مدل انتشار حرارت[1] استفاده شده است و پیدا کردن گره‌های تاثیرگذار به دو بخش شناسایی اجتماعات و شناسایی گره‌های پرنفوذ تقسیم شده است.

[1] Heat Diffusion Model

www.SID.ir

Archive of SID

جدول ۱: خلاصه ادبیات مرتبط

مرجع	سال	رویکرد یافتن رئوس تأثیرگذار	مشخصات مجموعه‌داده‌ی آزمایش‌ها	مدل انتشار
[1]	۲۰۱۳	خوشه‌بندی شبکه با استفاده از دو روش خوشه‌بندی طیفی و روش تراکمی Tang، و سپس انتخاب رأس با درجه‌ی بالاتر در هر اجتماع	چهار مجموعه‌داده‌ی واقعی فوتبال دانشکده‌های آمریکا، موسیقیدانان جاز، شبکه‌ی صفحات وب Webkb، و شبکه‌ی ارجاع انتشارات علمی Cora	آبشاری مستقل
[3]	۲۰۱۲	استفاده از چهار معیار انتخاب تصادفی، درجه، انتخاب حجمی و انتخاب حجمی وزن‌دار	چهار مجموعه‌داده‌ی واقعی PGP، ایمیل، وبلاگ و فیس‌بوک	آبشاری مستقل
[4]	۲۰۱۰	الگوریتم حریصانه‌ی مبتنی بر اجتماعات (CGA)	شبکه‌ی جزئیات تماس‌های تلفنی از بزرگ‌ترین فراهم‌آورنده‌ی خدمت ارتباطی چین	آبشاری مستقل
[5]	۲۰۰۵	الگوریتم حریصانه	نامشخص	آبشاری کاهشی
[6]	۲۰۱۱	الگوریتم دومرحله‌ای جستجوی k عدد از مؤثرترین رئوس (GAUP)	یک شبکه‌ی نویسندگی مشارکی دانشگاهی استخراج‌شده از DBLP	آبشاری مستقل گسترش‌یافته

۳- چارچوب تحقیق

بیشینه‌سازی تأثیرگذاری، یافتن مؤثرترین رئوس در یک شبکه است. فرض می‌شود که رئوس یک شبکه قادر به اتخاذ یک ایده، خرید یک محصول یا مواردی مشابه هستند. این فرآیند فعال‌سازی نام دارد. همچنین فرض می‌شود که رئوس فعال‌شده توانایی اثرگذاری (فعال‌سازی) روی همسایگان بلاواسطه‌ی خود را دارند که این همسایگان نیز به نوبه‌ی خود می‌توانند رئوس دیگر را فعال سازند. مسئله‌ی مورد بحث در این‌جا انتخاب بهترین رئوس برای فعال‌سازی اولیه به منظور بیشینه‌سازی تعداد رئوس فعال‌شده در انتهای فرآیند فعال‌سازی است.

۳-۱- استفاده از روش حریصانه‌ی Kempe

Kempe اظهار داشته است که مسئله‌ی بیشینه‌سازی اثرگذاری از مسائل حل‌نشدنی است. بر این اساس برای حل این نوع مسئله، الگوریتم حریصانه با تقریب قابل اثبات می‌تواند نتایج بهتری نسبت به دیگر روش‌ها حاصل کند. رویکرد حریصانه‌ی استفاده شده توسط Kempe، با یافتن بهترین رأس برای فعال‌سازی با استفاده از روش جستجوی فراگیر شروع می‌شود. یک رأس فعال می‌شود و سپس مدل انتشار به دفعات دلخواه به کار گرفته می‌شود. پس از آزمایش همه‌ی رئوس، رأسی که تعداد بیش‌تری از رئوس را فعال کرده باشد انتخاب می‌شود. سپس هر یک از رئوس باقی‌مانده به این رأس اولیه اضافه می‌شود و دوباره شبیه‌سازی‌ها برای انتخاب بهترین رأس برای اضافه کردن به مجموعه رئوس اولیه اجرا می‌شود. این فرآیند آن‌قدر تکرار می‌شود تا k رأس انتخاب شوند. در شکل ۱، شبه‌کد الگوریتم حریصانه مشاهده می‌شود.

در این روش می‌توان تعداد رئوس اولیه‌ی دلخواه فعال را مشخص کرد. این روش از روش‌های دیگر ساده‌تر است زیرا تنها به گراف شبکه و تعداد دلخواه رئوس تأثیرگذار به عنوان ورودی نیاز دارد. اما در این روش اجتماعات در نظر گرفته نشده‌اند و همچنین، این روش اگرچه کارایی مناسبی دارد اما سرعت بسیار کمی به ویژه در شبکه‌های بزرگ دارد؛ در نتیجه استفاده از این روش معقول نیست.

Algorithm 1 GeneralGreedy(G, k)

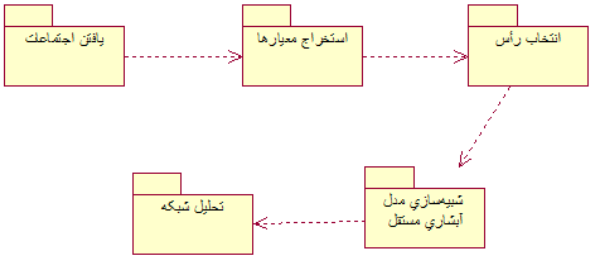
intialize $S = \emptyset$ and $N = 20000$
for $i = 1$ to k **do**
  for each vertex $v \in V \setminus S$ **do**
    $s_v = 0$.
    for $i = 1$ to N **do**
      $s_v \mathrel{+}= |RanCas(S \cup \{v\})|$
    end for
    $s_v = s_v / N$
  end for
  $S = S \cup \{\arg\max_{v \in V \setminus S}\{s_v\}\}$
end for
output S.

شکل ۱: شبه‌کد الگوریتم حریصانه‌ی عمومی [۱۶]

۳-۲- استفاده از روش‌های یافتن اجتماعات و سپس انتخاب رئوس بااهمیت

روش دیگر، مبتنی بر شناسایی اجتماعات در شبکه‌ها و سپس انتشار اثر است، به طوری که مطمئن باشیم در هر اجتماع یا خوشه حداقل یک رأس اولیه‌ی فعال‌شده وجود دارد. در روش انتشار اثر با استفاده از اجتماعات، چهار نوع مرکزیت برای هر رأس استخراج و رأس با مرکزیت بیش‌تر در هر خوشه، به عنوان عضوی از مجموعه‌ی اولیه برای فعال‌سازی برگزیده می‌شود.

شکل ۲، نمودار مؤلفه‌های فرآیند پیاده‌سازی و تحلیل در روش یافتن اجتماعات را نشان می‌دهد. در ادامه، مؤلفه‌های اصلی این رویکرد شرح داده می‌شوند.



شکل ۲: نمودار مؤلفه‌های فرآیند پیاده‌سازی و تحلیل در روش یافتن اجتماعات

www.SID.ir

Archive of SID

۳-۲-۱- یافتن اجتماعات

Wang و همکارانش روش جدیدی به نام "الگوریتم حریصانه‌ی مبتنی بر اجتماعات برای یافتن K رأس از تأثیرگذارترین رئوس" ارائه داده‌اند که شبکه را به تعدادی اجتماع تقسیم می‌کند و سپس تعدادی از این اجتماعات را برای یافتن K رأس از تأثیرگذارترین رئوس انتخاب می‌کند [۴]. ساختار اجتماعات یکی از ویژگی‌های اصلی شبکه‌های اجتماعی است: افرادِ درون هر اجتماع با یکدیگر ارتباط مکرر دارند؛ اما افرادِ اجتماعات مختلف ارتباط کم‌تری با یکدیگر دارند و بنابراین کم‌تر احتمال دارد که بر یکدیگر اثر بگذارند. این ویژگی اظهار می‌دارد که ممکن است یافتن رئوس تأثیرگذار درون اجتماعات (به جای کلّ شبکه) تقریب خوبی برای این مسئله باشد.

فرآیند یافتن اجتماعات در یک شبکه مشابه خوشه‌بندی در داده‌کاوی است. هدف خوشه‌بندی، در یک گروه قرار دادن نمونه‌ها به طریقی است که فاصله‌ی درون گروه‌ها حداقل و فاصله‌ی میان آن‌ها حداکثر شود. در خوشه‌بندی معمولا از یک تابع فاصله میان هر زوج از نمونه‌ها استفاده می‌شود. الگوریتم‌های یافتن اجتماعات از ساختار پیوندها استفاده می‌کنند (دو رأس یا با هم پیوند دارند یا نه). هدف این امر با توجه به نوع الگوریتم تفاوت می‌کند، اما معمولا هدف از آن حداکثرسازی تعداد پیوندها در درون اجتماعات و حداقل‌سازی آن‌ها در میان اجتماعات است. در این مقاله، از سه روش کشف اجتماعات Louvain، CPM و CONGA استفاده شده است که در ادامه شرح داده می‌شوند:

روش Louvain: این روش که از سرعت زیادی برخوردار است، برای یافتن اجتماعات مجزا در گراف‌های با تعداد رئوس زیاد و هم‌چنین در شبکه‌های ad hoc با ساختار شناخته‌شده از اجتماعات استفاده می‌شود و شامل دو مرحله است: (۱) ابتدا اجتماعات کوچک با بهینه‌سازی پیمانگی[۲] به صورت محلی یافت می‌شوند. (۲) رئوس یک اجتماع با یکدیگر ترکیب می‌شوند و شبکه‌ی جدیدی تشکیل می‌دهند که رئوس آن اجتماعات شبکه‌ی اصلی هستند. این مراحل تکرار می‌شوند تا مقدار تقریبی حداکثری از پیمانگی حاصل شود به طوری که در هر مرحله اجتماعات بزرگ و بزرگ‌تری ایجاد می‌شوند و سلسله‌مراتبی از ترکیب شبکه حاصل می‌شود [۸] و [۲۰]. این روش در نرم‌افزار Gephi پیاده‌سازی شده است.

روش CPM[۳]: این روش برای یافتن اجتماعات همپوشان یا گروهک[۴]های Kتایی در گراف‌های بدون وزن/وزن‌دار و بدون جهت/جهت‌دار استفاده می‌شود. تعریف اجتماعات در این روش بر پایه‌ی این اصل است که تعداد معمولی از اعضای یک اجتماع با تعداد زیادی از دیگر اعضا پیوند دارند، اما نه الزاماً تمامی دیگر رئوس آن اجتماع. به بیان دیگر، یک اجتماع می‌تواند به عنوان مجموعه‌ای از زیرگراف‌های کامل و کوچک‌تری (کاملا همبند) که رئوس را به اشتراک می‌گذارند تعبیر شود. چنین زیرگراف‌های کاملی گروهک‌های Kتایی نام دارند که K تعداد رئوس آن زیرگراف است و یک اجتماع از گروهک‌های Kتایی، اجتماعی از تمامی گروهک‌های Kتایی است که از هر یک می‌توان از طریق یک گروهک Kتایی به همسایه دیگری رسید. این روش در نرم‌افزار CFinder پیاده‌سازی شده است و مزایای استفاده از آن به شرح زیر هستند:

- بر اساس تراکم یال‌ها است.
- محلی است.
- منجر به تولید رئوس یا یال‌های برشی[۵] نمی‌شود.
- اجتماعات هم‌پوشان تولید می‌کند: یک رأس می‌تواند به طور هم‌زمان عضو تعداد متفاوتی از اجتماعات باشد و اجتماعات می‌توانند با به اشتراک‌گذاری رئوس، با یک‌دیگر هم‌پوشانی کنند [۹].

روش CONGA[۶]: روشی برای کشف اجتماعات هم‌پوشان در شبکه‌ها است که با گسترش روش بینابینی Girvan و Newman طراحی شده است. مانند الگوریتم اصلی، CONGA نیز به صورت سلسله‌مراتبی یک شبکه را به تعداد دلخواهی اجتماع خوشه‌بندی می‌کند اما برخلاف الگوریتم Girvan و Newman، در این‌جا هم‌پوشانی اجتماعات مجاز است. این الگوریتم به صورت زیر عمل می‌کند:

۱. محاسبه‌ی بینابینی تمامی یال‌ها در شبکه
۲. یافتن یال با بالاترین میزان بینابینی و حذف آن
۳. محاسبه‌ی دوباره‌ی بینابینی تمام یال‌های باقی‌مانده
۴. بازگشت به مرحله‌ی ۲ تا زمانی که هیچ یالی باقی نمانده باشد [۹].

این روش توسط برنامه مبتنی بر جاوا GONGA.jar قابل انجام است.

۳-۲-۲- استخراج معیارها و انتخاب رئوس بااهمیت

پس از خوشه‌بندی شبکه، به معیارهایی برای تعیین بهترین رئوس اولیه برای فعال‌سازی نیاز داریم. در انجام این تحقیق، از چهار نوع مرکزیت رأس برای تعیین این رئوس استفاده شده است. مرکزیت یک رأس، مکان ساختاری آن رأس برای یافتن اهمیت نسبی آن در گراف را تعریف می‌کند. انواع مختلفی از معیارهای مرکزیتِ یک رأس برای یافتن اهمیت آن در ساختار شبکه‌ی اجتماعی وجود دارد که در ادامه به چهار مورد به کار رفته در این مقاله اشاره می‌شود.

مرکزیت درجه[۷]: در گراف $G=(V,E)$ با تعداد $|V|$ رأس و $|E|$ یال، مرکزیت درجه‌ی رأس v به صورت زیر نشان داده می‌شود [۹]:

$$C_D(v)=\deg(v) \qquad (1)$$

مرکزیت نزدیکی[۸]: مرکزیت نزدیکی یک رأس u حاصل جمع فواصل تمامی کوتاه‌ترین مسیرها از رأس u به $n-1$ رأس دیگر گراف است. از

---

[2] modularity

[3] Clique Percolation Method (CPM)

[۴] Clique: زیرمجموعه‌ای از رئوس در یک گراف بدون جهت است که هر دو رأس در این زیرمجموعه به وسیله یک لبه (یال) به هم متصل شده‌اند.

[5] Cut-nodes and Cut-edges

[6] Cluster-Overlap Newman Girvan Algorithm (CONGA)

[7] Degree Centrality

[8] Closeness Centrality

www.SID.ir

Archive of SID

آن‌جایی که این حاصل جمعِ فواصل وابسته به تعداد رئوس این گراف است، این مرکزیت توسط مقدار حداقل حاصل جمع فواصل ممکن نرمال‌سازی شده و به صورت زیر بیان می‌شود:

$$C(u)=\frac{n-1}{\sum_{v=1}^{n-1} d(u,v)} \qquad (2)$$

به طوری که $d(u,v)$ فاصله‌ی کوتاه‌ترین مسیر میان v و u، و n تعداد رئوس گراف است [۱۹].

مرکزیت بینابینی[9]: تعداد دفعاتی است که یک رأس به عنوان پلی در کوتاه‌ترین مسیر میان دو رأس دیگر عمل می‌کند و توسط Linton Freeman معرفی شده است:

$$C_B(i)=\frac{\sum g_{jl}(i)}{g_{jl}} \qquad (3)$$

که در آن g_{jl} برابر تعداد کوتاه‌ترین مسیرهایی میان رئوس j و l است که رأس i را نیز در بر دارند [۱۰].

مرکزیت بردار ویژه[10]: این مرکزیت معیاری برای اندازه‌گیری اثر رأس در شبکه است. این مرکزیت، براساس این مفهوم امتیازهایی نسبی به تمامی رئوس شبکه تخصیص می‌دهد به طوری که برقراری ارتباط با رئوس پرامتیاز منجر به امتیاز بیش‌تری برای رأس مورد نظر می‌شود. مقدار این مرکزیت با استفاده از ماتریس همسایگی محاسبه می‌شود. برای گراف مفروض $G=(V,E)$ با تعداد $|V|$ رأس، $A=(a_{v,t})$ ماتریس همسایگی است. امتیاز مرکزیت رأس v به شکل زیر محاسبه می‌شود:

$$x_v=\frac{1}{\lambda}\sum_{t\in M(v)} x_t=\frac{1}{\lambda}\sum_{t\in G} a_{v,t}x_t \qquad (4)$$

به طوری که $M(v)$ مجموعه‌ی همسایگان رأس v و λ یک مقدار ثابت است [۱۰].

۳-۲-۳- شبیه‌سازی مدل آبشاری مستقل

پس از انتخاب رئوس اولیه‌ی فعال‌سازی، اثر باید از این رئوس در کل شبکه شروع به انتشار کند. Kempe در [۲] مدل‌های مختلفی از انتشار را مورد بحث قرار داده است که رفتار فعال‌سازی یک رأس را توصیف می‌کنند. در آزمایش‌های این مقاله مدل پرکاربرد آبشاری مستقل انتخاب شده است. این مدلِ ارائه‌شده توسط Goldenberg، Libai و Muller با مجموعه‌ای از رئوس فعالِ A_0 شروع می‌شود، سپس فرآیند، آبشاری از فعالیت‌ها را در مراحل گسسته آغاز می‌کند. زمانی که رأس u در مرحله‌ی t فعال می‌شود، تنها یک فرصت برای فعال‌سازی هر یک از همسایگان غیرفعالش (رئوس v) با احتمال موفقیت P(u,v) دارد. ترتیب تلاش‌های فعال‌سازیِ همسایگان غیرفعال اختیاری است.

مستقل از موفقیت یا عدم موفقیت فعال‌سازی توسط راس u، تا پایان فرآیند u دیگر نمی‌تواند برای فعال‌سازی v تلاش کند. فرآیند تا زمانی اجرا می‌شود که دیگر هیچ فعال‌سازی‌ای امکان‌پذیر نباشد. این فرآیند مستقل نام دارد زیرا فعال‌سازی هر رأسِ v مستقل از سابقه‌ی فعال‌سازی‌ها است. همانند Kempe، در این مقاله نیز برای سهولت تمامی احتمالات P(u,v) یکسان در نظر گرفته شده‌است.

۳-۳- ارزیابی

به منظور مقایسه نتایج معیارهای زیر در پایان فرایند انتشار مورد استفاده قرار گرفتند:

- نسبت رئوس (گره‌های) فعال به کل رئوس شبکه
- متوسط نسبت رئوس فعال به کل رئوس خوشه‌ها در هر شبکه
- پوشش رئوس فعال در روش‌های مبتنی بر اجتماعات که نسبت اشتراک رئوس فعال در دو روش مفروض به یکی از روش‌هاست.

۴- پیاده‌سازی و ارزیابی

در این بخش، پس از معرفی مجموعه داده‌ها و ارائه‌ی مشخصات ساختاری آن‌ها، نتایج اجرای دو رویکرد انتشار اثر تحت شبیه‌سازی‌های مختلف ارائه می‌شوند. به منظور ارزیابی نتایج حاصل از پیاده‌سازی‌ها، تعداد و درصد رئوس فعال‌شده در دو روش مقایسه می‌شوند.

۴-۱- مشخصات مجموعه‌داده‌ها

در انجام این پروژه، از چهار مجموعه‌ی داده که در ساختار و اندازه با یکدیگر متفاوت بودند استفاده شده است. در جدول ۲، مشخصات ساختاری هر یک از این مجموعه داده‌ها مشاهده می‌شود.

جدول ۱: مشخصات ساختاری مجموعه داده‌ها (بدون جهت=UD، بدون وزن=UW، جهت‌دار=D و وزن‌دار=W)

مجموعه‌ی داده	|V|	|E|	نوع شبکه	چگالی شبکه	ضریب خوشه‌بندی
DS1: شبکه فوتبال دانشکده‌های آمریکا [۱۱] و [۱۲]	۱۱۵	۶۱۳	UD UW	۰,۰۹۴	۰,۴۰۳
DS2: موسیقیدانان جاز [۱۳] و [۱۴]	۱۹۸	۲۷۴۲	D W	۰,۰۷۰	۰,۳۰۹
DS3: باشگاه کاراته زاکری [۱۵] و [۱۶]	۳۴	۱۵۶	D W	۰,۱۳۹	۰,۵۷۱
DS4: شبکه اجتماعی دلفین‌ها [۱۷] و [۱۸]	۶۲	۱۵۹	UD UW	۰,۰۸۴	۰,۳۰۳

۴-۲- خوشه‌بندی مجموعه‌ی داده‌ها به سه روش خوشه‌بندی

تعداد خوشه‌های یافت‌شده در هر مجموعه داده توسط هر یک از سه روش خوشه‌بندی در جدول ۳ آمده است. همچنین نمودار توزیع رئوس در هر خوشه برای هر چهار مجموعه داده در شکل ۳ خلاصه شده است. همانطور که ملاحظه می‌شود تعداد خوشه‌ها و توزیع رئوس در هر خوشه برای هر ۳ روش متفاوت است.

[9] Betweenness Centrality

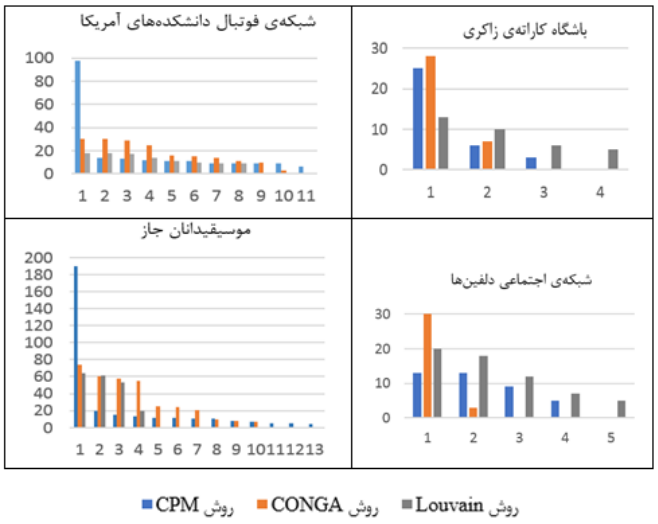
[10] Eigenvector Centrality

www.SID.ir

Archive of SID

جدول ۲: تعداد خوشه‌های یافت شده در هر مجموعه‌ی داده

نام مجموعه‌ی داده	تعداد خوشه‌ها CPM	تعداد خوشه‌ها CONGA	تعداد خوشه‌ها Louvain
DS1	۱۱	۱۰	۹
DS2	۱۳	۱۰	۴
DS3	۳	۲	۴
DS4	۴	۲	۵



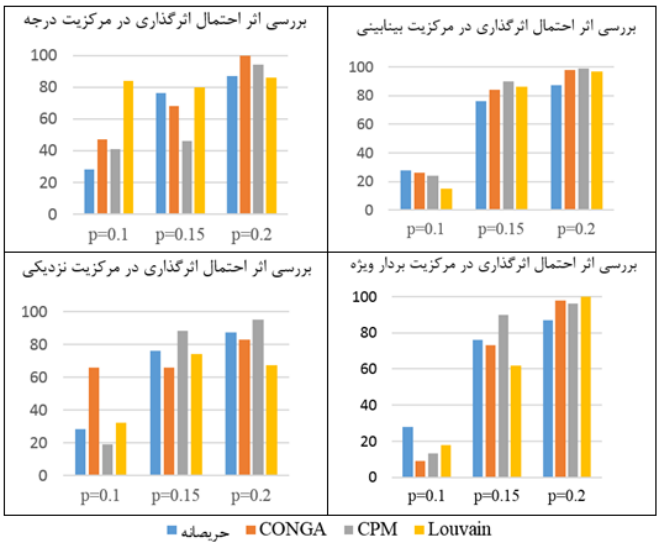
شکل ۱: نمودار توزیع رأس در هر خوشه در چهار مجموعه داده (محور افقی شماره خوشه و محور عمودی تعداد رئوس در هر خوشه است)

۴-۳- استخراج رئوس کاندید بر اساس معیار مرکزیت

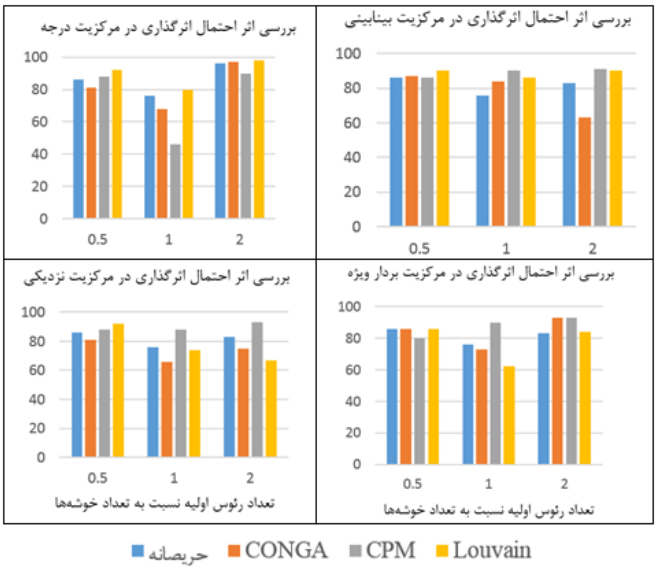
در این مرحله، مقادیر مرکزیت رئوس در هر خوشه محاسبه می‌شود. پس از مرتب‌سازی این مقادیر، در هر خوشه رأس با مقدار مرکزیت بیشتر به عنوان عضوی در مجموعه رئوس اولیه کاندید (seed) انتخاب می‌شود. مرکزیت براساس معیارهای معرفی شده در بخش ۳-۲-۲ به طور مجزا بدست می‌آید تا تاثیر هر یک را در بیشینه‌سازی اثر بتوان بررسی نمود. به دلیل آن‌که بعضی از خوشه‌ها هم‌پوشان هستند، برخی از رئوس کاندید در میان خوشه‌های همپوشان مشترک هستند. بنابراین تعداد رئوس کاندید، حداکثر برابر تعداد خوشه‌های جدا از هم خواهد بود.

۴-۴- انتشار اثر

در این مقاله برای انتشار اثر از مدل آبشاری مستقل استفاده می‌شود. هم‌چنین، از روش حریصانه نیز به عنوان رویکردی پایه با عملکرد مناسب (اما سرعت پایین) به منظور ارزیابی استفاده می‌نماییم. در روش حریصانه، تعداد تکرارها برابر N=1000 در نظر گرفته شده است. به منظور بررسی اثر احتمال اثرگذاری (p) بر روی تعداد رئوس فعال‌شده نهایی، انتشار اثر با مقادیر مختلف p=0.2، p=0.15 و p=0.1 اجرا شده است. همچنین، به منظور بررسی اثر تعداد رئوس اولیه فعال‌شده در تعداد رئوس فعال‌شده نهایی، انتشار اثر با سه مجموعه به اندازه متفاوت از رئوس اولیه یعنی برابر تعداد خوشه‌های یافت‌شده در هر مجموعه داده (C)، نیمی از این تعداد (C/2)، و دو برابر این تعداد (2C) اجرا شده است. در حالت دو برابر، رئوس اولیه اضافی به تصادف از رئوس با مرکزیت بالا انتخاب شده‌اند. در کلیه آزمایش‌های انتشار در زمان محدود که برابر با دو گام می‌باشد انجام شده است. نتایج، بر اساس درصد رئوس فعال‌شده نهایی نسبت به تعداد کل رئوس هر مجموعه داده، و درصد تعداد رئوس فعال‌شده نهایی در روش خوشه‌بندی نسبت به تعداد رئوس فعال‌شده نهایی در روش حریصانه با یکدیگر مقایسه شده‌اند. به دلیل تکثر تعداد نمودارها متناسب با چهار معیار مرکزیت و چهار مجموعه داده، نتایج به صورت خلاصه تنها برای مجموعه داده DS1 ارائه شده است (شکل ۴ و شکل ۵).



شکل ۴: نمودار درصد رئوس فعال‌شده نهایی نسبت به تعداد کل رئوس در مجموعه داده DS1 به روش کشف اجتماعات و حریصانه (محور افقی احتمال تاثیر p و محور عمودی درصد رئوس فعال شده است)



شکل ۵: نمودار درصد رئوس فعال‌شده نهایی نسبت به تعداد کل رئوس در مجموعه داده DS1، به روش کشف اجتماعات و حریصانه، و براساس تعداد رئوس اولیه مختلف (0.5C، C و 2C) و p=0.15

www.SID.ir

Archive of SID

براساس شکل ۴ و با بررسی درصد رئوس فعال‌شده برای هر چهار معیار مرکزیت، مشاهده می‌شود که با افزایش احتمال اثرگذاری، تعداد رئوس فعال‌شده نهایی در تمامی روش‌های مبتنی بر کشف اجتماع و روش حریصانه افزایش می‌یابد. همچنین همان‌طور که در شکل ۵ مشاهده می‌شود، اگر رئوس اولیه از مجموعه رئوس با مرکزیت بالا انتخاب شوند، تعداد آن‌ها در تعداد رئوس فعال‌شده نهایی تأثیر چندانی نخواهد داشت.

میزان پوشش نسبی رئوس فعال در روش مبتنی بر اجتماعات برای مجموعه داده DS3 در شکل ۶ آمده است. برای مثال ردیف چهارم مشخص می‌کند میزان اثرگذاری در روش‌های CONGA و Louvain با احتمال‌های اثرگذاری مختلف نسبت به روش CPM چه مقدار است. زمانی که p=01 است، هر دو روش CONGA و Louvain به نسبت یکسانی رئوس را فعال نموده‌اند. در احتمال p=0.15 و p=0.2 میزان پوشش روش Louvain و CPM بیشتر است.

متوسط اثرگذاری در اجتماعات کشف شده در هر روش و میزان اثرگذاری کلی در شبکه در شکل ۷ آمده است. صرف نظر از احتمال اثرگذاری، روش Louvain و CPM علاوه بر اثرگذاری بیشتر در شبکه، موجب متوسط اثرگذاری بیشتر در اجتماعات نیز می‌شوند.

انتخاب رئوس مبتنی بر مرکزیت درجه									
$\frac{x \cap y}{y}$	CPM			CONGA			Louvain		
p	**0.1**	**0.15**	**0.2**	**0.1**	**0.15**	**0.2**	**0.1**	**0.15**	**0.2**
CPM	100	100	100	43.75	54.55	33.33	43.75	72.73	83.33
CONGA	63.64	85.71	33.33	100	100	100	45.45	57.14	66.67
Louvain	70	53.33	41.67	50	26.67	33.33	100	100	100
انتخاب رئوس مبتنی بر مرکزیت نزدیکی									
$\frac{x \cap y}{y}$	CPM			CONGA			Louvain		
p	**0.1**	**0.15**	**0.2**	**0.1**	**0.15**	**0.2**	**0.1**	**0.15**	**0.2**
CPM	100	100	100	16.67	50	14.29	83.33	83.33	28.57
CONGA	42.86	54.55	50	100	100	100	85.71	72.78	25
Louvain	75	66.67	44.44	30	53.33	11.11	100	100	100
انتخاب رئوس مبتنی بر مرکزیت بینابینی									
$\frac{x \cap y}{y}$	CPM			CONGA			Louvain		
p	**0.1**	**0.15**	**0.2**	**0.1**	**0.15**	**0.2**	**0.1**	**0.15**	**0.2**
CPM	100	100	100	84.21	42.86	20	63.16	57.14	40
CONGA	88.89	50	100	100	100	100	66.67	50	100
Louvain	80	40	12.5	80	30	6.25	100	100	100
انتخاب رئوس مبتنی بر مرکزیت بردار ویژه									
$\frac{x \cap y}{y}$	CPM			CONGA			Louvain		
p	**0.1**	**0.15**	**0.2**	**0.1**	**0.15**	**0.2**	**0.1**	**0.15**	**0.2**
CPM	100	100	100	70	70.69	50	80	69.23	25
CONGA	43.75	16.67	54.55	100	100	100	81.25	33.33	36.36
Louvain	44.44	64.29	50	72.22	14.29	66.67	100	100	100

شکل ۶: پوشش تاثیرگذاری در روش مبتنی بر اجتماعات در DS3

P=0.2		P=0.15		P=0.1		روش کشف اجتماعات	معیار مرکزیت
Inf in Net (%)	Avg Inf in Comm (%)	Inf in Net (%)	Avg Inf in Comm (%)	Inf in Net (%)	Avg Inf in Comm (%)		
47.06	53.3	32.35	37	17.65	37	CPM	درجه
32.35	32.5	20.59	19.5	17.65	17.5	CONGA	
29.4	37.75	44.1	61.7	35.3	47.5	Louvain	
52.94	48.7	35.29	31	41.18	43.7	CPM	نزدیکی
20.59	35.5	32.35	37.5	11.76	20	CONGA	
58.8	61.75	44.1	48.75	26.5	31	Louvain	
55.88	57.3	17.65	37	14.71	32	CPM	بینابینی
52.94	55.5	17.65	17.5	2.94	9	CONGA	
44.1	42	29.4	31	47.1	54	Louvain	
29.41	38.7	41.18	47.33	38.24	41	CPM	بردار ویژه
47.059	51.5	14.71	16	32.35	32.5	CONGA	
52.94	50.5	41.18	50	17.65	23.75	Louvain	

شکل ۷: میزان تاثیرگذاری در کل شبکه (Inf in Net) و اجتماعات (Avg Inf in Comm) در روش مبتنی بر اجتماعات در DS3

P=0.2. گام زمانی برابر با ۲، مدل انتشار برابر با آبشاری مستقل و تعداد رئوس اولیه برابر با تعداد خوشه‌ها													
مجموعه داده	حریصانه	CONGA				CPM				Louvain			
		درجه	نزدیکی	بینابینی	بردار ویژه	درجه	نزدیکی	بینابینی	بردار ویژه	درجه	نزدیکی	بینابینی	بردار ویژه
DS1	87	100	83	98	98	94	97	99	96	86	67	97	100
DS2	92	95	96	96	96	97	95	96	97	92	95	96	95
DS3	29	70	29	70	26	73	97	88	55	55	79	73	61
DS4	32	24	35	51	50	56	62	43	19	30	27	48	24
P=0.15. گام زمانی برابر با ۲، مدل انتشار برابر با آبشاری مستقل و تعداد رئوس اولیه برابر با تعداد خوشه‌ها													
مجموعه داده	حریصانه	CONGA				CPM				Louvain			
		درجه	نزدیکی	بینابینی	بردار ویژه	درجه	نزدیکی	بینابینی	بردار ویژه	درجه	نزدیکی	بینابینی	بردار ویژه
DS1	76	68	66	84	73	46	88	90	90	80	74	86	62
DS2	88	94	93	86	95	97	95	92	94	94	84	88	94
DS3	41	55	41	32	41	52	23	55	26	64	61	47	70
DS4	33	46	27	22	82	53	62	46	33	27	32	56	33
P=0.1. گام زمانی برابر با ۲، مدل انتشار برابر با آبشاری مستقل و تعداد رئوس اولیه برابر با تعداد خوشه‌ها													
مجموعه داده	حریصانه	CONGA				CPM				Louvain			
		درجه	نزدیکی	بینابینی	بردار ویژه	درجه	نزدیکی	بینابینی	بردار ویژه	درجه	نزدیکی	بینابینی	بردار ویژه
DS1	28	47	66	26	9	41	19	24	13	84	32	15	18
DS2	91	90	90	90	90	90	92	90	90	87	93	92	90
DS3	29	14	50	35	20	58	11	11	20	23	29	55	55
DS4	14	12	6	19	4	14	9	16	19	12	16	25	9

شکل ۸: درصد رئوس فعال‌شده نهایی نسبت به تعداد کل رئوس

۴-۵- تحلیل نتایج

با توجه به نتایج آزمایش برای هر چهار مجموعه داده که در این مقاله بدلیل محدودیت تنها نمونه‌هایی ارائه شد، می‌توان چنین استنباط کرد که:

- اگر رئوس اولیه رئوسی با مرکزیت بالا باشند، اندازه مجموعه اولیه (تعداد رئوس اولیه فعال‌شده) عامل مهمی در تعداد رئوس فعال‌شده نهایی نخواهد بود. این نتیجه همراستا با نتیجه [۳] است (شکل ۵).
- چنانچه زمان انتشار اثر محدود باشد (دو گام زمانی در آزمایش‌ها)، روش حریصانه پاسخی بهتر از روش مبتنی بر یافتن اجتماعات ندارد (البته استثنائاتی نیز وجود دارد) و به طور میانگین روش مبتنی بر کشف اجتماعات بهتر است (شکل ۴).
- با توجه به پیچیدگی زمانی، استفاده از هر الگوریتم یافتن اجتماعات تقریباً در تمامی موارد از استفاده از رویکرد حریصانه سریع‌تر عمل خواهد کرد؛ [۱] نیز این مهم را اشاره کرده است.
- رویکرد مبتنی بر اجتماعات به طور متوسط در تمامی موارد نتایجی تقریباً به خوبی نتایج الگوریتم حریصانه تولید کرده‌است (شکل ۴).
- میزان پوشش رئوس فعال شده در روش‌های مختلف کشف اجتماع و با احتمال‌های اثرگذاری مختلف در دو روش Louvian و CPM به یک دیگر نزدیک تر است (شکل ۶). البته اگر معیار انتخاب رئوس مرکزیت بردار ویژه باشد نمی‌توان با قطعیت نظر داد.
- چنانچه علاوه بر حداکثرسازی تاثیر در کل شبکه، حداکثری تاثیر در اجتماعات نیز هدف باشد، روش‌های Louvian و CPM مناسب هستند.
- نتایج نشان می‌دهد که برای بعضی از مجموعه داده‌ها، رویکرد مبتنی بر اجتماعات بهتر از روش حریصانه است (مجموعه داده‌های فوتبال DS1 و جاز DS2). این برتری ارتباطی با چگالی این دو مجموعه ندارد. این دو مجموعه داده به صورت ذاتی دارای اجتماع هستند؛ منطقی به نظر می‌رسد که رویکرد مبتنی بر اجتماعات (به ویژه اجتماعات همپوشان) نتایج بهتری در آن‌ها تولید کند. یعنی، نتایج هم به ساختار معنایی و هم ساختار ارتباطات شبکه بستگی دارد (شکل ۸).

www.SID.ir

Archive of SID

- نتایج به دست‌آمده از تعداد رئوس فعال‌شده نهایی با رویکرد اجتماعات نسبت به تعداد رئوس فعال‌شده با روش حریصانه، این فرض را تأیید می‌کند که حداکثرسازی تأثیرگذاری مبتنی بر اجتماعات در زمان محدود می‌تواند نتایجی را به دست دهد که با روش‌های معمول رقابت کند و پاسخی نزدیک و حتی بهتر از روش حریصانه ارائه دهد (شکل ۸).

۵- نتیجه‌گیری و کارهای آتی

یافتن رئوس تأثیرگذار یکی از مسائل قابل توجه است که برای مدیران در بازاریابی، سیاست و سایر حوزه‌ها کاربرد بسیار دارد. الگوریتم‌هایی برای یافتن مجموعه‌ای اولیه از رئوس به منظور فعال‌سازی حداکثری رئوس شبکه پیشنهاد شده‌اند. این مسئله در ابتدا از مسائل حل‌نشدنی شناخته شده بود، اما بعدها الگوریتم‌های تقریبی پاسخی نزدیک به بهینه ولی بسیار کُند برای آن فراهم کردند. یکی از موارد جالب توجه و توسعه‌یافته این مسئله، گسترش تأثیرگذاری با تحت پوشش قرار دادن اجتماعات متفاوت در شبکه است.

در این مقاله، از الگوریتم‌های یافتن اجتماعات نه تنها برای حداکثرسازی رئوس فعال‌شده، بلکه برای تحت پوشش قرار دادن هر چه بیش‌تر اجتماعات مختلف استفاده شد. در آزمایش‌ها نشان داده شد که رویکرد استفاده از یافتن اجتماعات در بسیاری از مجموعه داده‌ها با الگوریتم معمول حریصانه قابل رقابت است و در حالی که رویکرد حریصانه تقریباً همیشه بهتر از هر روشی عمل می‌کند، استفاده از رویکردهای یافتن اجتماعات نیز نتایج خوبی به دست می‌دهد.

قابل توجه‌ترین نتیجه حاصل از انجام این تحقیق، پوشش‌دهی اجتماعات در نتیجه تاثیرات است. اگر ساختار اجتماعات یک مجموعه مشخص و بدیهی باشد، می‌توان از یک الگوریتم مناسب یافتن اجتماعات برای آن مجموعه به منظور حداکثرسازی پوشش‌دهی اجتماعات به جای روش حریصانه استفاده کرد و نتایج قابل قبولی دریافت نمود [۱].

به عنوان پیشنهادی برای کارهای آتی، می‌توان آزمایش‌ها را بر روی مجموعه داده‌هایی با تعداد رئوس و یال‌های بیش‌تر نیز انجام داد تا نتایج برای شکبه‌های بزرگ نیز قابل تعمیم یا رد شوند. هم‌چنین، می‌توان از معیارهای دیگری برای اندازه‌گیری مرکزیت رئوس استفاده نمود و دیگر روش‌های کشف اجتماعات (اعم از اجتماعات مجزا و هم‌پوشان) را بررسی کرد.

منابع و مراجع

[1] Scripps, J. Discovering Influential Nodes in Social Networks through Community Finding. in WEBIST. 2013.

[2] Kempe, D., J. Kleinberg, and É. Tardos. Maximizing the spread of influence through a social network. in Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining. 2003. ACM.

[3] Kim, H. and E. Yoneki. Influential neighbours selection for information diffusion in online social networks. in 2012 21st International Conference on Computer Communications and Networks (ICCCN). 2012. IEEE.

[4] Wang, Y., et al. Community-based greedy algorithm for mining top-k influential nodes in mobile social networks. in Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining. 2010. ACM.

[5] Kempe, D., J. Kleinberg, and É. Tardos. Influential nodes in a diffusion model for social networks. in International Colloquium on Automata, Languages, and Programming. 2005. Springer.

[6] Zhang, Y., J. Zhou, and J. Cheng. Preference-based top-k influential nodes mining in social networks. in 2011IEEE 10th International Conference on Trust, Security and Privacy in Computing and Communications. 2011. IEEE.

[7] Chen, Y.-C., W.-C. Peng, and S.-Y. Lee, Efficient algorithms for influence maximization in social networks. Knowledge and information systems, 2012. 33(3): p. 577-601.

[8] Blondel, V.D., et al., Fast unfolding of communities in large networks. Journal of statistical mechanics: theory and experiment, 2008. 2008(10): p. P10008.

[9] Gregory, S. An algorithm to find overlapping community structure in networks. in European Conference on Principles of Data Mining and Knowledge Discovery. 2007. Springer.

[10] Rabade, R., N. Mishra, and S. Sharma, Survey of influential user identification techniques in online social networks, in Recent advances in intelligent informatics. 2014, Springer. p. 359-370.

[11] Jain, A; Dubes, R. Algorithms for clustering data, Prentice-Hall, Inc, 1988.

[12] Porter, M; Onnela, J; Mucha, P. “Communities in Networks”, *Notices of the American Mathematical Society*, 2009.

[13] Tang, L; Xufei W; Huan, L; Lei, W. "A Multi-resolution Approach to Learning with Overlapping Communities." *Proceedings of the First Workshop on Social Media Analytics - SOMA '10*, 2010.

[14] Chen, Y; Chang, S; Chou, Ch; Peng, W; Lee, S. “Exploring Community Structures for Influence Maximization in Social Networks”, *The 6th SNA-KDD Workshop’12 (SNA-KDD’12)*, 2012.

[15] Jung, K; Heo, W; Chen, W. "IRIE: Scalable and Robust Influence Maximization in Social Networks." *IEEE 12th International Conference on Data Mining*, 2012.

[16] Chen, W; Wang, Y; Yang, S. "Efficient Influence Maximization in Social Networks." *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '09*, 2009.

[17] Blondel, V; Guillaume, J; Lambiotte, R; Lefebvre, E. "Fast Unfolding of Communities in Large Networks." *J. Stat. Mech. Journal of Statistical Mechanics: Theory and Experiment*, 2008.

[18] Blondel, V; Guillaume, J; Lambiotte, R; Lefebvre, E. “Louvain method: Finding communities in large networks”, *findcommunities*. Availability: https://sites.google.com/site/findcommunities/

[19] Shi, J; Malik, J. “Normalized Cuts and Image Segmentation”, *IEEE Transactions on Pattern Analysis and Machine Itelligence*, 2000.

[20] Rabade, R; Mishra, N; Sharma, S. “Survey of Influential User Identification Techniques in Online Social Networks”, *Advances in Intelligent Systems and Computing Recent Advances in Intelligent Informatics*, 359-370, 2014.

www.SID.ir