دوره ۸، بهار و تابستان ۱۳۹۴
مجله فناوری اطلاعات در طراحی مهندسی
Information Technology in Engineering Design
http://ited.sinaweb.net

ارائه روشی مبتنی بر رتبه‌بندی و الگوریتم انتشار گرما برای استخراج جامعه در شبکه‌های اجتماعی

ملیحه ابراهیمی نژاد*[۱]، سمانه حسن زاده [۲]، مهرداد جلالی[۳]

[۱]گروه مهندسی کامپیوتر، دانشکده ۱۷ شهریور کرج، دانشگاه فنی و حرفه ای استان البرز، ایران *

[۲]گروه مهندسی کامپیوتر و فناوری اطلاعات، موسسه آموزش عالی اقبال لاهوری، مشهد، ایران

[۳]گروه مهندسی کامپیوتر، واحد مشهد، دانشگاه آزاد اسلامی، مشهد، ایران

چکیده:

شبکه های اجتماعی در دهه گذشته توسعه سریعی داشته‌اند. استخراج جامعه مسئله مهمی در تحلیل شبکه‌های اجتماعی است. کشف جامعه در شبکه اجتماعی به‌معنای شناسایی مجموعه‌ای از گره‌هاست به‌گونه‌ای که اعضای آن بیشترین ارتباط را بایکدیگر و ارتباط کمی با گره‌های خارج از مجموعه دارند. الگوریتم کلاسیک خوشه‌بندی K میانه روش کارایی در این حوزه است ولی این الگوریتم حساس به نقاط اولیه ورودی است. برای حل این مشکل دراین پژوهش ابتدا K نقطه ثقل با استفاده از PageRank شناسایی و سپس ساختار جامعه با استفاده از شباهت انتشارگرما و K میانه استخراج می‌شود. با استفاده از شباهت انتشارگرما می‌توان اطلاعات سراسری شبکه را برای هر زوج از رأس‌ها بدست‌آورد. ارزیابی و مقایسه روش پیشنهادی با الگوریتم‌های K-Rank، K میانه، BGLL، OSLOM، LPA و Infomap نشانگر افزایش دقت در بیشتر شبکه‌های اجتماعی مورد بررسی است و برای انواع مختلف شبکه ها کاربردی است.

واژه‌های کلیدی:

استخراج جامعه، انتشار گرما، PageRank، شبکه‌های اجتماعی

۱-مقدمه

در دهه فعلی، شبکه‌های اجتماعی با رشد روز افزونی در حال توسعه‌اند به نحوی که در حال

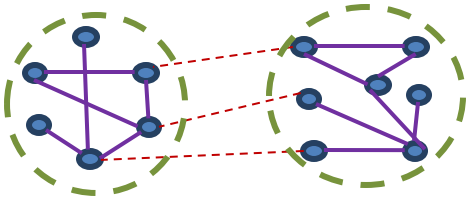
*عهده دار مکاتبات

نشانی: البرز، کرج، حصارک، بلوارآقارضایی، دانشکده فنی و حرفه ای دختران کرج.

تلفن: ۰۹۱۵۳۸۲۰۸۷۹ **پست الکترونیک:** Ebrahiminejad.ma@gmail.com

حاضر بعضی از این شبکه‌ها همانند فیسبوک میلیون‌ها کاربر دارند، بنابراین تجزیه و تحلیل داده‌های این شبکه‌ها به یکی از موضوعات حائز اهمیت در حوزه داده‌کاوی تبدیل شده است [۱]. به طور معمول شبکه‌های اجتماعی را می‌توان در قالب گراف نمایش داد که در این گراف‌ها، گره‌ها معادل افراد یا بازیگران شبکه‌های اجتماعی و یال‌ها نشان‌دهنده ارتباط بین این افراد هستند [۲]. با توجه به ساختار شبکه اجتماعی و یک‌طرفه یا دوطرفه بودن ارتباطات، گراف متناظر می‌تواند گراف جهت‌دار یا بدون‌جهت باشد. همچنین درصورتی که شدت و وضعیت ارتباطات بین افراد در شبکه‌های اجتماعی یکسان نباشد، گراف متناظر با شبکه، یک گراف وزن‌دار خواهد بود که در آن وزن هر یال متناظر با قوت ارتباط است.

در اغلب شبکه‌های اجتماعی مجموعه‌ای از بازیگران وجود دارند که ارتباطات مستحکم‌تری با یکدیگر داشته و موضوعات مورد علاقه آن‌ها نیز مشابه است که به این مجموعه از بازیگران جامعه گفته می‌شود [۳] (شکل (۱)).



شکل (۱) اجتماعات در شبکه‌های اجتماعی

استخراج جامعه در گراف شبکه‌های اجتماعی به معنای شناسایی مجموعه‌ای از گره‌ها است به‌گونه‌ای که اعضای این مجموعه‌ها بیشترین ارتباط را بایکدیگر و ارتباط کمی با گره‌های خارج از مجموعه دارند. به کمک روش‌های استخراج جامعه می‌توان رفتارهای گروهی متداول را دسته‌بندی و مطالعه‌ی هر قسمت از شبکه را به صورت متمرکز انجام داد. هنگامی‌که در شبکه‌های اجتماعی کاربران در جوامع براساس شباهت‌هایی که دارند گروه‌بندی شوند، این امر کمک فراوانی به توصیه و تطبیق می‌کند. به عنوان مثال می‌توان به پژوهش‌هایی که در زمینه توصیه فرد خبره در حوزه‌های مختلف از جمله حوزه‌های علمی انجام می‌گیرد اشاره‌کرد [٤]. در حوزه استخراج جامعه، محدودیت روش‌های استخراج و همچنین تفسیر نتایج خروجی تحقیق بیشتری را در این زمینه می‌طلبد.

در این مقاله برای استخراج جامعه با دقت بالاتر ما در ابتدا گره‌های با اهمیت را در شبکه اجتماعی با استفاده از الگوریتم انتشار گرما وPageRank مشخص می‌کنیم سپس جامعه را با این گره‌های بااهمیت با استفاده از الگوریتم خوشه‌بندی K میانه[۱] تشکیل می‌دهیم. ارزیابی و مقایسه روش پیشنهادی با الگوریتم‌های K-Rank، Kمیانه، BGLL[۲]، OSLOM[۳]، LPA[٤]وInfomap[٥] نشانگر افزایش دقت در بیشتر شبکه‌های اجتماعی مورد بررسی است. این روش قابلیت اعمال بر روی انواع شبکه‌ها را دارد و برای شبکه‌های بزرگ و خلوت مناسب است. روش به کار رفته در مرحله انتخاب نقاط ثقل اولیه و همچنین

---

۱ Kmeans

۲ Blondel,Guillaume, Lefebvre and Lambiotte

۳ Order Statistics Local Optimization Method

٤ Label Propagation Algorithm

٥ Information mapping

استفاده از ساختار سراسری شبکه الگوریتم پیشنهادی را کاراتر ساخته است.

ساختار مقاله در ادامه به این شرح است: در بخش ۲ مروری بر کارهای محققان این حوزه داریم. در بخش ۳ و ٤ به‌ترتیب الگوریتم انتشار گرما و PageRank توضیح داده شده است. در بخش ٥ الگوریتم پیشنهادی را معرفی می‌کنیم. نتایج پیاده سازی روش پیشنهادی بر روی شبکه‌های اجتماعی واقعی در بخش ٦ بیان می‌شود. بخش ۷ شامل نتیجه گیری و کارهای آینده است.

۲- روش‌های مختلف استخراج جامعه

به‌طور کلی، روش‌های شناسایی و استخراج جامعه در شبکه‌های اجتماعی به چهار گروه اجتماعات مبتنی بر گره، اجتماعات مبتنی بر گروه، اجتماعات مبتنی بر شبکه و اجتماعات سلسله مراتبی تقسیم می‌شوند[٥].

معیار تشخیص اجتماعات مبتنی بر گره بر این اساس است که هر گره در گروه حاوی خصوصیات مشخصی از قبیل تقابل و دسترس‌پذیری باشد. روش‌هایی همانند برات فورس[٦] [۳] و CPM[۷] [٦] [۳] از جمله روش‌های این گروه هستند. عیب اصلی الگوریتم‌های ذکرشده این است که برای شبکه‌های کوچک مناسب هستند. در روش مبتنی بر مرکزیت گروه بایستی کلیه اتصالات درون گروه را در نظر گرفت. از جمله روش‌های این گروه می‌توان به روش ابلو[۸] اشاره کرد. این روش طی دو مرحله جستجوی محلی و هرس کردن اکتشافی، جوامع را استخراج می‌کند [۷].

روش‌های مبتنی بر شبکه، ساختار کلی شبکه‌را در نظر می‌گیرند و گره‌ها را درون مجموعه‌هایی مجزا تقسیم‌بندی می‌کنند. در این روش‌ها یک معیار تعریف شده در سراسر شبکه و نه تنها در یک‌گروه از گره‌ها بهینه می‌شود. یکی از پارامترهای کیفی جامعه اکثریت[۹] [۸] است که خوبی یک دسته‌بندی گراف را مشخص می‌کند. در سال‌های اخیر از این پارامتر در اغلب الگوریتم‌های استخراج جامعه استفاده شده است و این روش‌ها به دنبال بدست آوردن حداکثر اکثریت هستند [۹]. همچنین می‌توان به روش‌های مختلف تقسیم‌بندی گراف [۱۰] و الگوریتم‌های خوشه‌بندی همانند K میانه [۱۱] نیز اشاره کرد. مسأله اصلی در الگوریتمKمیانه‌انتخاب مراکز اولیه است که در دقت خروجی تاثیر می‌گذارد. برای حل این مشکل الگوریتم‌های مختلفی پیشنهاد شده‌اند. از میان آن‌ها الگوریتم K++ میانه مشهورتر است که استراتژی خاصی در زمان انتخاب نقاط اولیه دارد [۱۲]. K++ میانه اکثر رأس‌ها در شبکه را قبل از انتخاب نقطه اولیه جدید مقایسه می‌کند که اگر شبکه بزرگ باشد، پیچیدگی این پردازش خیلی زیاد است.

در گروه چهارم روش‌های استخراج جامعه یک ساختار سلسله مراتبی از جامعه‌ها بر اساس ساختار شبکه ایجاد می‌شود. در این گروه می‌توان روش‌های متراکم شونده Single-Link، Average-Link و Complete Link و روش تقسیم کننده نیومن را نام برد [۱۳و۱٤].

---

٦ Brute-force

۷ Clique Percolation Method

۸ Abello

۹ Modularity

از جمله روش‌های نوین و کاربردی استخراج جامعه که از ترکیب روش‌های ذکرشده استفاده می‌کنند و همچنین به منظور مقایسه با روش پیشنهاد شده در این مقاله به کار می‌روند الگوریتم‌های K-Rank [۱۵]، BGLL [۱۶]، Infomap[۱۷]، OSLOM [۱۸] و LPA [۱۹] می‌باشند. الگوریتمK-rank روش کارایی برای استخراج جامعه بر اساس رتبه‌بندی مرکزیت و شباهت سیگنالی است. الگوریتمBGLL یک روش حریصانه مبتنی بر اکثریت برای استخراج جامعه در شبکه‌های بزرگ است. در این روش شناسایی جامعه‌ی کوچکتر از یک اندازه خاص غیر ممکن است و خروجی الگوریتم به ترتیبی که گره‌ها درنظر گرفته می‌شوند بستگی دارد. الگوریتم Infomap یک روش نظری اطلاعات برای استخراج ساختار جامعه در شبکه‌های وزن‌دار و جهت‌دار است که معمولا همگرایی کندی دارد. الگوریتم OSLOM یک تکنیک چندهدفه به منظور مدیریت گراف‌های جهت‌دار، غیرجهت‌دار و نیزگراف‌های وزن‌دار و غیروزن‌دار است که قادر است همپوشانی خوشه‌ها راشناسایی کند و یک سلسله مراتب از خوشه‌ها ایجاد کند. OSLOMیک الگوریتم محلی است و برای رسیدن به نتیجه دقیق‌تر احتیاج به تعداد تکرارهای زیادی دارد. الگوریتم LPA تنها از ساختار شبکه، به‌عنوان راهنمای خود استفاده می‌کند و نیازمند هیچ تابع هدفی و یا اطلاعاتی از پیش مشخص از جامعه‌ها نمی‌باشد ولی این الگوریتم همانند K میانه حساس به ورودی اولیه است.

۳- انتشار گرما

انتشـار گرما یک پدیده فیزیکی است. درحالت کلی، گرمـا از موقعیت‌هـای بـا دمـای بـالا بـه موقعیت‌های با دمای پایین حرکت می‌کند. اخیراً، روش‌هـای مبتنی بر انتشـــار گرما در حوزه‌های مختلف ازجمله کاهش ابعاد [۲۰]، رتبه‌بندی [۲۱]، ســیســتم‌های توصــیه‌گر [۲۲و۲۳] و بازاریابی شبکه‌های اجتماعی [۲٤] استفاده‌شده است. در این بخش به معرفی الگوریتم انتشار گرما در گراف‌های مختلف می‌پردازیم.

۳-۱- انتشـــار گرما در گراف‌های غیر جهت دار

گراف غیرجهت‌دار G=(V,E) را در نظر می‌گیریم که در آن V مجموعه رأس‌های گراف، E مجموعه یال‌های گراف و N برابر تعداد رأس‌ها است. یال (v_i, v_j) را به عنوان یالی در نظر می‌گیریم که رأس‌های v_i و v_j را به هم ارتباط می‌دهد. مقدار $f_i(t)$ نشان‌دهنده مقدار گرما در رأس v_i در لحظه t است با در نظر گرفتن این که مقدار گرما رأس v_i در آغاز برابر $f_i(0)$ بوده است. $f(t)$ نشان‌دهنده برداری است که مؤلفه‌های آن $f_i(t)$ هستند. مدل ریاضی انتشار گرما در گراف های غیرجهت‌دار به صورت زیر است:

$$F(t) \cong (I + \frac{\alpha}{m} * (H))^m F(0) \qquad (۱)$$

در رابطه (۱)، α ضریب انتقال گرما میان رأس‌ها ، I ماتریس یکه N تایی، m تعداد مراحل اجرای الگوریتم و ماتریس H با استفاده از رابطه زیر بدست می‌آید:

$$H_{ij} = \begin{cases} 1 & (v_i, v_j) \in E \\ -d(v_i) & i = j \\ 0 & \text{otherwise} \end{cases} \quad (۲)$$

$d(v_i)$ درجه رأس v_i است. براساس نتایج تجربی که در مقالات [۲۱][۲۳] بدست آمده ، α برابر یک و m برابر ۱۰ در نظر گرفته می‌شود.

۳-۲- انتشار گرما در گراف‌های جهت‌دار

مدل انتشار گرما در گراف‌های جهت دار متفاوت است و انتقال گرما تنها در جهت هر یال، بین رأس‌ها انتشار می‌یابد. گراف جهت دار G=(V,E,W) را در نظر می‌گیریم که در آن V مجموعه رأس‌های گراف، E مجموعه یال‌های گراف و W احتمال وجود یال‌ها (وزن) است. یال (v_i, v_j) را به عنوان یالی در نظر می‌گیریم که رأس v_i را به v_j مرتبط می‌کند و w_{ij} احتمال وجود یال (وزن) بین رأس v_i و v_j است. مقدار $f_i(t)$ نشان دهنده مقدار گرما در رأس v_i در لحظه t است با در نظر گرفتن این که مقدار گرما در آغاز برابر $f_i(0)$ بوده است. f(t) نشان دهنده برداری است که مؤلفه های آن $f_i(t)$ هستند. مدل ریاضی انتشار گرما در گراف‌های جهت‌دار به صورت زیر است:

$$F(t) \cong (I + \frac{\gamma}{m} * (H))^m F(0) \quad (۳)$$

در رابطه (۳) γ ضریب انتقال گرما میان رأس‌ها، I ماتریس یکه N تایی و ماتریس H با استفاده از رابطه زیر بدست می آید:

$$H_{ij} = \begin{cases} -\tau_i & i = j \\ \frac{w_{ji}}{\sum_{k:(j,k)\in E} w_{jk}} & (v_j, v_i) \in E \\ 0 & \text{otherwise} \end{cases} \quad (۴)$$

در رابطه (٤) τ_i اگر v_i به گره‌ای اشاره کند برابر ۱ و در غیر اینصورت ۰ است. براساس نتایج تجربی که در مقالات [۲۱][۲۳] بدست آمده ، γ برابر یک و m برابر ۱۰ در نظر گرفته می‌شود.

اگر E را برابر تعداد یال‌های گراف در نظر بگیریم پیچیدگی رابطه (۱) و (۳) O(mE) می‌شود.

هنگامی که گراف شبکه اجتماعی بزرگ است اجرای الگوریتم انتشار گرما زمان زیادی را صرف می‌کند. برای غلبه بر این مشکل در ابتدا یک زیر گراف که از گره منبع گرما شروع شود استخراج می‌شود. سپس با در نظر گرفتن این منبع گرما، زیرگراف با استفاده از الگوریتم جستجوی پیمایش عمقی در گراف اصلی ایجاد می‌شود. جستجو هنگامی خاتمه می‌یابد که تعداد گره‌ها از یک مقدار از پیش تعریف شده ای بیشتر شود. سپس فرآیند انتشار روی این زیر گراف به طور کارا و موثری اجرا می‌شود. این روش کارایی فرآیند انتشار گرما را کاهش نمی‌دهد زیرا گره‌هایی که خیلی از منبع گرما دور هستند معمولا به آن مرتبط نیستند.

۴- الگوریتم رتبه‌بندی صفحه

الگوریتم رتبه‌بندی صفحه گوگل یا Page Rank، الگوریتم تحلیل لینکی است که که به هر کدام از صفحات اینترنت یک رتبه نسبت می‌دهد. روش کار این الگوریتم در ادامه توضیح داده شده است [۳۲-۲۸].

اگر هریک از صفحات موجود در اینترنت را برابر یک حالت و لینک‌های خروجی بین صفحات را احتمالات انتقال بین آن حالات در نظر بگیریم، از آنجا که احتمال بودن در یک صفحه (حالت) تنها به حالت قبلی لینک دهنده وابسته است، فرآیند جابه جایی بین صفحات اینترنت یک زنجیره مارکوف خواهد بود. در توجیه این مساله می‌توان گفت که رفتار کاربر در انتخاب لینک‌های یک صفحه معمولا به صورت تصادفی است و لذا کاربر که در ابتدا در یک صفحه خاص قرار دارد، در مرحله بعد با یک احتمال مشخص (که برابر ۱ تقسیم بر تعداد لینک های آن صفحه است) بر روی یکی از لینک‌ها کلیک کرده و به صفحه بعد منتقل می‌شود. به همین منوال یکی از لینک‌های صفحه جدید انتخاب خواهد شد.

$$T = [T_{ij}] \quad , \quad T_{ij} = \frac{1}{|O_i|} \qquad (۵)$$

در رابطه T_{ij}(۵) نشان دهنده احتمال انتخاب صفحه jاز صفحه i و$|O_i|$ تعداد لینک‌های صفحه i و T ماتریس انتقال حالت است. اگر فرآیند جابه‌جایی بین صفحات یک فرآیند مارکوف باشد، ماتریس حالت سکون، احتمال قرار داشتن کاربر در هر یک از صفحات یا رتبه صفحه خواهد بود. برای احراز شرایط سکون زنجیره مارکوف، دو اصلاح روی ماتریس انتقال T انجام می شود:

الف) برای داشتن شرط دسترسی پذیری می‌بایست تمام گره‌ها از هر گره قابل دسترس باشند، لذا برای حالت‌هایی که لینک خروجی نداشته و نمی‌توان از آنها به دیگر صفحات منتقل شد احتمال خروج مساوی $1/n$ قرار داده می‌شود. ماتریس حاصل از این مرحله را T' می‌نامیم:

$$T' = [T_i] + \frac{1}{n} \qquad (۶)$$

که اگر i صفحه ای است که هیچ لینک خروجی ندارد و احتمال انتخاب هر صفحه j برابر صفر است $(T_{ij} = 0)$

ب) برای احراز شرط نادوره‌ای بودن، چنین در نظر می‌گیریم که ممکن است کاربر مستقیما و بدون استفاده از لینک‌ها به یک صفحه خاص برود(مثلا آدرس صفحه را تایپ کند) و لذا احتمالی برای این حالت در نظر گرفته می‌شود که صرف‌نظر از صفحه‌ای که در آن قرار داریم به طور مستقیم بتوان به تمام صفحات پرش نمود. این بدان معنی است که در هر حالت یک چرخه به طول ۱ خواهیم‌داشت که شرط نادوره‌ای بودن را برآورده می‌نماید. به این منظور پارامتر α را برابر احتمال استفاده از لینک‌ها در جابه‌جایی بین صفحات، و پارامتر $1-\alpha$ را برابر احتمال پرش مستقیم به یک صفحه خاص در نظر می‌گیریم (مقدار α توسط گوگل معمولا برابر ۰.۸۵ استفاده می‌شود). در این حالت ماتریس اصلاح‌شده انتقال را T'' می‌نامیم‌که برابر خواهد بود با :

$$T'' = \alpha T' + (1-\alpha)E \quad , \quad 0 \le \alpha \le 1 \qquad (۷)$$

که مقدار تمامی درایه های ماتریس E برابر 1/n است .

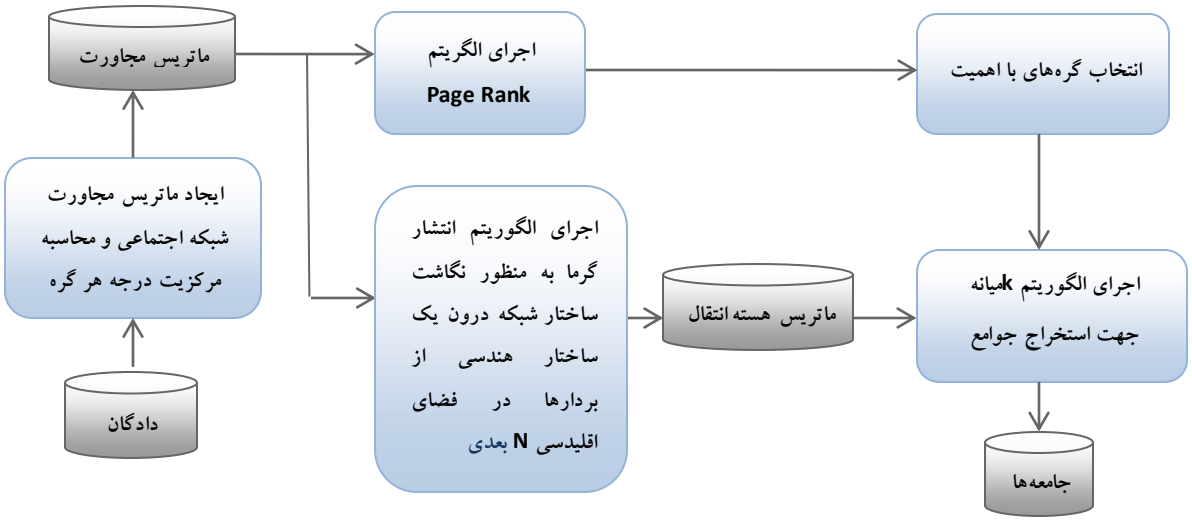
حال با یافتن ماتریس سکون می‌توان احتمال رجوع به هر صفحه را محاسبه کرد. این ماتریس سکون پاسخ معادله برگشتی زیراست:

$$r_i = \sum_{j \in I_i} \frac{r_j}{|O_j|} \qquad (۸)$$

در این رابطه r_i برابر رتبه صفحه iام و r_j رتبه صفحه‌های لینک‌دهنده به صفحه i است. $|O_j|$ نیز تعداد لینک‌های صفحه j ام و I_i مجموعه تمام صفحه‌های لینک‌دهنده به صفحه i است. معنی این رابطه این است که هر صفحه، ارزش رتبه خود را به طور مساوی بین همه صفحه‌هایی که لینک می‌دهد تقسیم می‌کند و صفحات بعدی هم به همین منوال به طوری که در نهایت این رابطه برای تمامی گره‌ها برقرار است و همان ماتریس سکون خواهد بود. پیچیدگی زمانی الگوریتم رتبه بندی صفحه $O(\log(n))$ است که در آن n تعداد صفحات وب می‌باشد.

۵-روش پیشنهادی

در این بخش مسائل مرتبط با روش پیشنهاد شده شامل ایجاد ماتریس مجاورت و محاسبه مرکزیت درجه، اجرای الگوریتم انتشارگرما، چگونگی انتخاب گره‌های بااهمیت و خوشه‌بندی با Kمیانه توضیح داده می‌شود. چارچوب روش پیشنهادی در شکل (۲) نمایش داده شده است.



شکل(۲) چهارچوب روش پیشنهادی

$$A_{v_1 v_2} = \begin{cases} W(v_1, v_2) & v_1 \; R \; v_2 \\ 0 & otherwise \end{cases} \qquad (۹)$$

در این رابطه R ارتباط بین دو کاربر v_1 و v_2 را تعیین می‌کند. این ارتباط می‌تواند هم تالیفی در یک مقاله، فرستادن ایمیل به یکدیگر، علاقه‌مند بودن به یک کالا یا موضوع مشخص و ارتباطات بسیار دیگری که در شبکه‌های اجتماعی مطرح هستند در

۵-۱ ایجاد ماتریس مجاورت شبکه اجتماعی و محاسبه مرکزیت درجه هر گره

ماتریس مجاورت شبکه اجتماعی بر اساس رابطه (۹) ایجاد می‌شود:

نظر گرفت. W در این رابطه بر اساس نوع شبکه اجتماعی متفاوت خواهد بود در شبکه‌های بدون‌وزن برابر ۱ و در شبکه‌های وزن‌دار برابر وزن ارتباط است به عنوان مثال می‌تواند نرخ انتقال پیام بین دو کاربری که به یکدیگر ایمیل می‌فرستند باشد. سپس مرکزیت درجه[۱] هر گره محاسبه می‌شود [۲۵]. مرکزیت درجه در گراف‌های غیرجهت‌دار برابر درجه هر گره است. در گراف‌های جهت‌دار به دو دسته، مرکزیت درجه درونی ومرکزیت درجه خارجی تقسیم می‌شود [۲۶].

در گراف‌های وزن‌دار درجه مرکزیت هر گره برابر مجموع وزن یال‌های متصل به آن در نظر گرفته می‌شود. این درجه را در گراف‌های وزن‌دار قدرت گره نیز می‌نامند. اپسهل[۲] و همکاران تعریف دیگری را برای درجه مرکزیت گراف‌های وزن‌دار پیشنهاد دادند. بر اساس تعریف آن‌ها درجه مرکزیت در گراف‌های وزن‌دار برابر با حاصلضرب درجه هر گره در توان α ام میانگین وزن یال‌های متصل به آن گره است [۲۷].

۵-۲ اجرای الگوریتم انتشار گرما

الگوریتم انتشار گرما بر اساس نوع گراف شبکه اجتماعی اجرا می‌شود. در این روش هر گره شبکه به عنوان یک سیستم تحریک‌پذیر در نظر گرفته می‌شود که می‌تواند گرما را دریافت، ارسال و ذخیره کند. در ابتدا یک گره به عنوان منبع گرما در نظر گرفته شده و با یک گرمای اولیه مقداردهی می‌شود و مابقی گره‌ها بدون گرما در نظر گرفته می‌شوند. سپس گره منبع این مقدار گرما را به همسایه‌هایش انتقال می‌دهد. بعد از این، گره‌هایی که گرما دارند می‌توانند گرمایشان را به همه همسایه‌های خود انتقال دهند و از همسایه‌هایشان گرما دریافت و مقدار مشخصی از گرما را ذخیره کنند. بعد از m مرحله میزان توزیع گرما در سراسر شبکه به عنوان اثر گره منبع روی کل شبکه در نظر گرفته می‌شود. گره‌هایی که در یک جامعه هستند اثر مشابهی بر روی کل شبکه دارند .

برای شبکه ای با N گره توزیع گرما می‌تواند به صورت یک بردار N بعدی در نظر گرفته شود. اجرای الگوریتم انتشار گرما بر روی شبکه، ساختار شبکه را درون یک ساختار هندسی از بردارها در فضای اقلیدسی N بعدی نگاشت می‌کند. ماتریس هسته انتقال مهمترین خروجی الگوریتم انتشار گرما است و سطر i ام ماتریس هسته انتقال به عنوان اثر گره i بر روی کل شبکه در نظر گرفته می‌شود. در مراحل بعدی از این ماتریس برای انتخاب مراکز اولیه و خوشه‌بندی استفاده می‌شود.

۵-۳ انتخاب گره‌های با اهمیت

برای رتبه‌بندی گره‌های شبکه اجتماعی از الگوریتم PageRank به دلیل کارایی و پیچیدگی زمانی مناسبی که دارد استفاده می‌شود. در ابتدا الگوریتم PageRank اجرا می‌شود و به هریک از گره‌ها رتبه‌ای اختصاص می‌یابد. سپس گره‌ها بر اساس این رتبه مرتب می‌شوند. گره‌هایی که

---

[۱] Degree centrality

[۲] Opsahl

بیشترین رتبه دارند، احتمال بیشتری برای انتخاب به عنوان نقاط اولیه ورودی الگوریتم K میانه دارند. این نقاط اولیه انتخابی که تحت عنوان گره‌های با اهمیت در این تحقیق ذکر شده است، به عنوان مراکز اولیه الگوریتم خوشه بندی K میانه انتخاب می‌شوند. اگر فرض کنیم که شبکه اجتماعی نمونه K جامعه دارد، در یک نگاه ساده شاید K گره‌ای که بیشترین رتبه را دارند انتخاب شوند ولی مطابق با الگوریتم PageRank اگر یک صفحه رتبه بالایی داشته باشد، رتبه صفحه‌های دیگری که با این صفحه ارتباط دارند نیز بالا می‌رود، بنابراین اگر K گره با رتبه بالا به عنوان گره‌های با اهمیت انتخاب شوند، واضح است که این نقاط اولیه خیلی به هم نزدیک هستند و کاندیدهای مناسبی برای انتخاب به عنوان نقاط اولیه نمی‌باشند.

بنابراین در ابتدا اولین نقطه را که بالاترین رتبه دارد انتخاب می‌کنیم و گره بعدی به گونه‌ای انتخاب می‌شود که هم رتبه بالاتری دارد وهم شباهت آن با نقطه اولیه کمتر از آستانه ε باشد و همین طور تا انتخاب k گره، انتخاب گره‌ها به همین صورت ادامه می‌یابد.

بدست آوردن آستانه ε کار سختی است. در انتخاب نقاط اولیه این مقدار آستانه فاکتور مهمی است و بر این باوریم که چگونگی تخمین آن هنوز یک مساله باز است. اگر مقدار آن کوچک باشد برای انتخاب مناسب نقاط باید بخش زیادی از نقاط را در نظر بگیریم که شباهت بین نقاط انتخابی از ε کمتر باشد و گره‌های خوبی انتخاب نمی‌شوند زیرا با این روش بیشتر نقاط انتخابی رتبه پایینی دارند. اگر مقدار ε بزرگ باشد، برای انتخاب مناسب نقاط باید بخش کمی از نقاط را در نظر بگیریم و به عبارتی این همان روش انتخاب k نقطه با رتبه بالاتر است و باز هم گره‌های خوبی انتخاب نمی شوند. بنابراین مقدار ε نه باید خیلی کوچک و نه خیلی بزرگ باشد.

از آنجایی که تخمین این مقدار مشکل است بر اساس نتایج تجربی که در حین پیاده سازی الگوریتم بر روی دادگان‌های نمونه بدست آوردیم، اگر تعداد جامعه‌ها کم باشد مقدار ε، کوچکی انتخاب می‌شود و اگر تعداد جامعه‌ها زیاد باشد ε بزرگ بهتر است یعنی:

$$\varepsilon \propto K \qquad (۱۰)$$

۴-۵ خوشه بندی با الگوریتم K میانه

در این مرحله با استفاده از گره‌های با اهمیت که در مرحله قبل انتخاب شد و به عنوان گره‌های ثقل اولیه الگوریتم k میانه در نظر گرفته شدند و همچنین ماتریس هسته انتقال که با اجرای الگوریتم انتشار گرما حاصل شد، الگوریتم خوشه‌بندی K میانه اجرا می‌شود. برای خوشه بندی نیاز به محاسبه شباهت رأس‌ها داریم. این شباهت با استفاده محاسبه فاصله اقلیدسی بردارهایی که در طی فرآیند انتشار گرما به هر رأس اختصاص یافته اند محاسبه می‌شود. ما این نحوه محاسبه شباهت را "شباهت انتشار گرما" می‌نامیم. در نهایت الگوریتم ۱، الگوریتم روش پیشنهادی برای استخراج جامعه در شبکه‌های اجتماعی را نمایش می‌دهد.

الگوریتم ۱: الگوریتم روش پیشنهادی

ورودی: تعداد جامعه‌ها(K)، ماتریس مجاورت شبکه(A)، آستانه شباهت(ε)، حداکثر تکرار(N_{max}).

خروجی: جوامع شبکه

۱. اجرای الگوریتم انتشار گرما به منظور نگاشت ساختار شبکه درون یک ساختار هندسی از بردارها در فضای اقلیدسی N بعدی.

۲. انتخاب K گره با اهمیت مطابق با روش ذکر شده در بخش ۵-۳ (شباهت رأس‌ها به وسیله فاصله اقلیدسی محاسبه می‌شود).

۳. اجرای الگوریتم k میانه به منظور استخراج K جامعه تا زمانی‌که که دیگر همه نقاط ثقل یا جامعه‌های استخراج شده تغییر نکنند یا تعداد تکرار به N_{max} برسد.

۵-۵ تحلیل پیچیدگی زمانی الگوریتم پیشنهادی

پیچیدگی زمانی الگوریتم پیشنهادی به فرآیند انتشار گرما، انتخاب گره‌های اولیه و خوشه‌بندی وابسته است. همانطور که در بخش ۳ ذکر شد پیچیدگی زمانی فرآیند انتشار گرما برابر O(mE) است که در آن E تعداد یال‌ها و m تعداد تکرار فرآیند انتشار گرما بوده و طبق نتایج تجربی برابر ۱۰ در نظر گرفته می‌شود [۲۱][۲۳]. بهترین الگوریتم مرتب سازی N عنصر، در حالت متوسط دارای پیچیدگی زمانی O(N. logN) است. پیچیدگی زمانی الگوریتم K میانه، O(NtK) بوده که در آن K تعداد خوشه‌ها، N تعداد گره‌ها و t تعداد تکرار است [۱۰].

بنابراین پیچیدگی زمانی الگوریتم پیشنهادی برابر O(mE+NtK+N.logN) است.

۶- ارزیابی روش پیشنهادی

به منظور ارزیابی روش پیشنهادی از شبکه‌های اجتماعی واقعی جدول (۱) استفاده شده است. در این جدول N تعداد گره‌ها، E تعداد یال‌ها و K تعداد جوامع است. در جدول (۱) بر اساس مقدار سطح خلوتی [۳۱] که برابر یک منهای چگالی است، شبکه اجتماعیBlogs نسبت به مابقی شبکه‌ها خلوت‌تر است. علت اصلی انتخاب این دادگان‌ها این است که در اکثر پژوهش‌ها و مقاله‌های معتبر در حوزه تحقیقاتی استخراج جوامع آن‌ها استفاده شده‌اند و این امر کمک فروانی به ارزیابی روش پیشنهادی با تعداد زیادی از روش‌های استخراج جامعه می‌کند. در این تحقیق از معیارهای ارزیابیNMI وصحت که از جمله معیارهای مرسوم ارزیابی دقت، در اغلب الگوریتم‌های استخراج جامعه می‌باشند استفاده شده است [۲]. معیارNMI شباهت بین مجموعه درست و نتایج روش پیشنهادی را به‌کمک تعیین یک نگاشت یک به یک بین جامعه‌های شناسایی شده در آن‌ها اندازه می‌گیرد و معیار صحت تمام زوج‌های ممکن از گره‌ها را در نظر می‌گیرد و بررسی می‌کند که آیا آن‌ها در یک جامعه یکسان قرار دارند یا خیر.

جدول(۱) شبکه های اجتماعی واقعی

ردیف	شبکه اجتماعی	N	E	K	سطح خلوتی	ε	منبع
۱	Zachary's club	۳۴	۷۸	۲	۰.۸۶۰۹	۰.۵	[۳۲]
۲	Risk	۴۲	۸۳	۶	۰.۹۰۳۶	۰.۷	[۱۵]
۳	Dolphins	۶۲	۱۵۹	۲	۰.۹۱۵۹	۰.۵	[۳۳]
۴	Lesmis	۷۷	۲۵۴	۱۱	۰.۹۱۳۲	۰.۸	[۳۴]
۵	Political books	۱۰۵	۴۴۱	۳	۰.۹۱۹۲	۰.۶	[۳۵]
۶	Football	۱۱۵	۶۱۳	۱۲	۰.۹۰۶۵	۰.۹	[۳۶]
۷	Blogs	۱۴۹۰	۱۹۵۱۵	۲	۰.۹۹۱۲	۰.۵	[۳۷]
۸	PPI	۱۶۲۸	۱۱۸۹۳	۴۰۸	۰.۹۹۱۰	۱	[۳۸]

جدول (۲) مقایسه دقت روش پیشنهادی و الگوریتم‌های مختلف با معیارNMI

ردیف	شبکه	روش پیشنهادی	K-Rank	K-means	LPA	BGLL	Infomap	OSLOM
۱	Zachary's club	**۱**	**۱**	**۱**	۰.۶۴	۰.۵۸	۰.۶۹	۰.۸۴
۲	Risk	۰.۹۲	**۰.۹۶**	**۰.۹۶**	۰.۹۰	۰.۹۴	۰.۹۴	۰.۴۲
۳	Dolphins	**۱**	**۱**	**۱**	۰.۶۰	۰.۴۶	۰.۵۳	۰.۵۵
۴	Lesmis	**۰.۸۲**	۰.۸۰	۰.۶۶	۰.۷۴	۰.۸۱	۰.۷۸	۰.۶۴
۵	Political books	۰.۵۵	**۰.۵۷**	۰.۵۵	۰.۵۳	۰.۵۱	۰.۵۳	۰.۵۵
۶	Football	۰.۹۲	۰.۹۲	۰.۹۰	۰.۸۶	**۰.۹۳**	۰.۹۲	۰.۹۱
۷	Blogs	**۰.۶۸**	۰.۵۱	۰.۳۸	۰.۲۹	۰.۳۳	۰.۲۹	۰.۵۰
۸	PPI	**۰.۹۸**	۰.۹۴	–	۰.۹۴	۰.۹۵	۰.۹۰	۰.۹۳

جدول (۳) مقایسه دقت روش پیشنهادی و الگوریتم‌های مختلف با معیار صحت

ردیف	شبکه	روش پیشنهادی	K-Rank	K-means	LPA	BGLL	Infomap	OSLOM
۱	Zachary's club	**۱**	**۱**	**۱**	۰.۹۱	۰.۶۴	۰.۸۲	۰.۹۷
۲	Risk	۰.۹۵	**۰.۹۸**	**۰.۹۸**	۰.۸۵	۰.۸۵	۰.۸۵	۰.۴۲
۳	Dolphins	**۱**	**۱**	**۱**	۰.۷۷	۰.۵۰	۰.۵۸	۰.۸۷
۴	Lesmis	**۰.۸۷**	۰.۷۱	۰.۶۲	۰.۷۲	۰.۷۷	۰.۷۶	۰.۵۹
۵	Political books	۰.۷۶	**۰.۸۵**	**۰.۸۵**	۰.۸۲	۰.۷۲	۰.۷۸	۰.۸۰
۶	Football	۰.۹۱	**۰.۹۳**	۰.۹۰	۰.۸۰	۰.۹۰	۰.۹۱	۰.۹۱
۷	Blogs	**۰.۹۴**	۰.۸۹	۰.۸۲	۰.۶۳	۰.۷۹	۰.۵۸	۰.۸۸
۸	PPI	**۰.۹۵**	۰.۹۱	–	۰.۸۷	۰.۸۸	۰.۷۵	۰.۸۵

جدول‌های (۲) و (۳) نتیجه مقایسه دقت روش پیشنهادی با الگوریتم‌هایOSLOM[۱۸]، Infomap[۱۷]، BGLL[۱۶]، K-Rank[15] و K میانه[۱۱] هستند. بر طبق نتایج حاصل از پیاده‌سازی، الگوریتم پیشنهادی درمقایسه با الگوریتم‌های ذکر شده دقت بالایی در مورد بیشتر شبکه‌های اجتماعی مورد بررسی دارد و در برخی موارد دقت الگوریتم‌های BGLL، Kمیانه و K-Rank بهتر است. به دلیل اینکهK میانه و LPA هر دو حساس به ورودی اولیه می‌باشند، هر یک ازاین الگوریتم ها ده بار اجرا شده‌اند و نتیجه‌ای که بیشترین مقدار دقت را با هریک از پارامترهای ارزیابی دارد در جدول‌های (۲) و (۳) درج شده است.

علامت "–" در جداول ذکر شده به معنای این است که الگوریتم K میانه به دلیل انتخاب بد نقاط اولیه در شبکه بزرگ PPI که تعداد جوامع آن نیز زیاد است نمی‌تواند جوامع را استخراج کند. بنابراین در شبکه‌های بزرگ که تعداد جوامع آن‌ها نیز زیاد است K میانه جوامع خوبی را استخراج نمی‌کندزیرا انتخاب تصادفی نقاط ثقل اولیه منجر به ایجاد یک خوشه خالی یایک نتیجه خوشه‌بندی با دقت خوشه‌بندی کم می‌شود. الگوریتم پیشنهادی و K-Rank به دلیل انتخاب مناسب نقاط ثقل اولیه از این نتایج بد جلوگیری می‌کنند.

از آنجایی که در مراحل ابتدایی الگوریتم نقاط اولیه در الگوریتم پیشنهادی به درستی انتخاب می‌شوند الگوریتم پیشنهادی بر خلاف K میانه و LPA حساس به ورودی اولیه نیست و در هر بار اجرای الگوریتم به یک دقت واحد استخراج جامعه می‌رسد. همچنین در دادگان‌های بزرگ الگوریتم پیشنهادی دقت بالایی در استخراج جوامع نسبت به تمام الگوریتم‌های ذکر شده دارد. همانطور که از نتایج مشخص است خلوت بودن ماتریس مجاورت شبکه اجتماعی در دقت تاثیرگذار است به عنوان مثال در شبکه اجتماعیBlogs که نسبت به شبکه‌های اجتماعی دیگر خلوت‌تر است، به ترتیب الگوریتم پیشنهادی، K-Rank و OSLOM دقت خوبی در استخراج جامعه در این شبکه دارند.

همانطور که در بخش۵-٤ ذکر شد الگوریتم پیشنهادی برای خوشه‌بندی نیاز به محاسبه‌شباهت رأس‌ها دارد و ما "شباهت انتشار گرما" را بر اساس ساختار شبکه به کار بردیم. بنابراین با این شیوه می‌توان اطلاعات سراسری شبکه را برای‌هر زوج از رأس‌ها بدست آورد. این اطلاعات‌سراسری است که الگوریتم پیشنهادی ما را کاراتر ساخته است. بسیاری از روش‌ها تنها از اطلاعات محلی شبکه استفاده می‌کنند.

نتیجه مقایسه پیچیدگی زمانی الگوریتم پیشنهادی و الگوریتم‌های مختلف در جدول (٤) نمایش داده شده‌است. علامت "–" در جدول (٤) به معنای این است که پیچیدگی زمانی الگوریتم OSLOM به ویژگی‌های خاص ساختار جامعه مورد مطالعه بستگی دارد و نمی‌توان تخمینی زد.

جدول(۴) مقایسه پیچیدگی زمانی الگوریتم پیشنهادی و الگوریتم های مختلف

الگوریتم	پیچیدگی زمانی	منبع
الگوریتم پیشنهادی	O(mE+NtK+N.logN)	
K-Rank	$O(c(k+1)N^2+NtK)$	[۱۵]
K میانه	O(NKt)	[۱۰]
LPA	O(N+E)	[۱۹]
BGLL	O(E)	[۱۶]
Infomap	O(E)	[۱۷]
OSLOM	-	[۱۸]

اگرچه برخی از این الگوریتم‌ها نسبت به الگوریتم پیشنهادی از نظر پیچیدگی زمانی سریع‌تر هستند ولی براساس نتایج جداول (۲) و (۳) ، دقت خوبی در استخراج جامعه در شبکه‌های مختلف ندارند. K میانه همگرایی سریعی دارد ولی حساس به ورودی اولیه است. در الگوریتم پیشنهادی از این همگرایی سریع K میانه استفاده شده و در عین حال مشکل ذکر شده نیز برطرف شده است. همچنین الگوریتم LPA علاوه بر این که حساس به ورودی اولیه است در استخراج شبکه‌های خلوت همانند Infomap دقت کمی دارد. هر چندالگوریتم BGLL برای استخراج جوامع شبکه‌های بزرگ مناسب است و الگوریتم سریعی است ولی برای شبکه‌های خلوت مناسب نیست. همچنین در مقایسه با روشK-Rankالگوریتم پیشنهادی پیچیدگی زمانی بهتری دارد.

۷- جمع بندی و کارهای آینده

از جمله پیامدهای جدیدی که به واسطه نفوذ اینترنت در دنیای امروز رشد و توسعه یافت، ایجاد شبکه‌های اجتماعی است که به کاربران فرصت می‌دهند تا اطلاعات و علاقه‌مندی‌های خود رانخیره نموده و با دیگران به اشتراک بگذارند. از این رو حجم وسیعی ازاطلاعات بسیار ارزشمند در فضای اینترنت موجب پیشنهاد الگوریتم‌هایی شده است که با استفاده هرچه بهینه‌تر از این اطلاعات به کمک کاربران اینترنت بشتابند. ازآن جمله الگوریتم‌های استخراج جوامع هستند.

در این مقاله در ابتدا با استفاده از اطلاعات معتبرموجود، ماتریس مجاورت شبکه ایجاد و درجه مرکزیت هر گره محاسبه شد. سپس با استفاده از الگوریتم انتشار گرما و PageRank گره‌های بااهمیت شناسایی شدند. در نهایت با استفاده از این گره‌های بااهمیت که به عنوان نقاط ثقل اولیه الگوریتم خوشه بندیKمیانه در نظرگرفته می‌شوند، جوامع استخراج شدند.

ارزیابی و مقایسه روش پیشنهادی با الگوریتم‌های K-Rank، Kمیانه، BGLL، OSLOM، LPA وInfomap با معیارهای ارزیابیNMI و صحت نشانگر افزایش معیارهای ذکرشده درمورد بیشتر شبکه‌های اجتماعی مورد بررسی است. این

روش قابلیت اعمال بر روی انواع شبکه‌ها را دارد و برای شبکه‌های بزرگ و خلوت مناسب است. همچنین همانند K میانه همگرایی سریعی دارد. روش به کار رفته در مرحله انتخاب نقاط ثقل اولیه و همچنین استفاده از ساختار سراسری شبکه الگوریتم پیشنهادی را کاراتر ساخته است.

از جمله چشم اندازهای احتمالی برای توسعه روش پیشنهادی این است که علاوه بر در نظر گرفتن ساختار شبکه، محتوای روابط کاربران نیز در نظر گرفته شود. به خصوص در مواردی که استخراج جوامع برای پیشنهاد فرد خبره در حوزه‌های مختلف استفاده می‌شود. همچنین دربرخی از شبکه‌های اجتماعی همانند فیسبوک، فلیکر و ... کاربران بر اساس سلایق و حوزه های کاری مختلف، تمایل به عضویت در جوامع مختلف را دارند بنابراین در این موارد جوامع همپوشانی دارند. بنابراین بعد از استخراج جوامع در شبکه‌های اجتماعی با استفاده از الگوریتم پیشنهادی،مشخص کردن گره‌هایی که متعلق به جامعه های‌متفاوت هستند را می‌توان در نظر گرفت.

۸- مراجع

[1] C. C. Aggarwal, *An introduction to social network data analytics*. Springer US, 2011.

[2] S. Fortunato, "Community detection in graphs," *Phys. Rep.*, vol. 486, no. 3–5, pp. 75–174, Feb. 2010.

[3] T. Lei, H. Liu, and L. Tang, *Community detection and mining in social media*, vol. 2, no. 1. Morgan & Claypool Publishers, pp. 1–137, 2010.

[4] Y. Xu, X. Guo, J. Hao, J. Ma, R. Y. K. Lau, and W. Xu, "Combining social network and semantic concept analysis for personalized academic researcher recommendation," *Decis. Support Syst.*, vol. 54, no. 1, pp. 564–573, 2012.

[5] L. Tang and H. Liu, "Graph mining applications to social network analysis," in *Managing and Mining Graph Data*, "Springer, pp. 487–513, 2010.

[6] G. Palla, I. Derényi, I. Farkas, and T. Vicsek, "Uncovering the overlapping community structure of complex networks in nature and society," *Nature*, vol. 435, no. 7043, pp. 814–818, 2005.

[7] J. Abello, M. G. C. Resende, and S. Sudarsky, "Massive quasi-clique detection," in *Theoretical Informatics*, Springer, pp. 598–612, 2002.

[8] M. E. J. Newman, "Modularity and community structure in networks," *Proc. Natl. Acad. Sci.*, vol. 103, no. 23, pp. 8577–8582, 2006.

[9] R. Shang, J. Bai, L. Jiao, and C. Jin, "Community detection based on modularity and an improved genetic algorithm," *Phys. A Stat. Mech. its Appl.*, vol. 392, no. 5, pp. 1215–1231, Mar. 2013.

[10] H. Jiawei, K. Micheline, and P. Jian, *Data Mining concepts and techniques*, pp. 452–454, 2012.

[11] J. MacQueen and others, "Some methods for classification and analysis of multivariate observations," in *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, " vol. 1, no. 281–297, p. 14, 1967.

[12] D. Arthur and S. Vassilvitskii, "k-means++: The advantages of careful seeding," in *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*, " pp. 1027–1035, 2007.

[13] M. Newman and M. Girvan, "Finding and evaluating community structure in networks," *Phys. Rev. E*, vol. 69, no. 2, 2004.

[14] M. Newman, "Fast algorithm for detecting community structure in networks," *Phys. Rev. E*, vol. 69, no. 6, p. 66133, 2004.

[15] Y. Jiang, C. Jia, and J. Yu, "An efficient community detection method based on rank centrality," *Phys. A Stat. Mech. its Appl.*, vol. 392, no. 9, pp. 2182–2194, May 2013.

[16] V. D. Blondel, J.-L. Guillaume, E. Lefebvre, and R. Lambiotte, "Fast unfolding of communities in large networks," *J. Stat. Mech. Theory Exp.*, 2008.

[17] M. Rosvall and C. T. Bergstrom, "Maps of random walks on complex networks reveal community structure," *Proc. Natl. Acad. Sci.*, vol. 105, no. 4, pp. 1118–1123, 2008.

[18] A. Lancichinetti, F. Radicchi, J. J. Ramasco, and S. Fortunato, "Finding statistically significant communities in networks," *PLoS One*, vol. 6, no. 4, p. e18961, 2011.

[19] U. N. Raghavan, S. Kumara, and R. Albert, "Near linear time algorithm to detect community structures in large-scale networks," *Phys. Rev. E*, vol. 76, no. 3, 2007.

[20] M. Belkin and P. Niyogi, "Laplacian Eigenmaps for Dimensionality Reduction and Data Representation," vol. 15, no. June, pp. 1373–1396, 2003.

[21] H. Yang, I. King, and M. R. Lyu, "DiffusionRank:A Possible Penicillin for Web Spamming," pp. 431–438, 2007.

[22] S. Aarthi and S. Sampath, "A Heat Diffusion Method for Mining Web Graphs for Recommendations Using Recommendation Algorithm," *Int. J. Eng. Res. Technol.*, vol. 2, no. 8, pp. 961–966, 2013.

[23] H. Ma, I. King, S. Member, and M. R. Lyu, "Mining Web Graphs for Recommendations," vol. 24, no. 6, pp. 1051–1064, 2012.

[24] H. Ma, H. Yang, M. R. Lyu, and I. King, "Mining Social Networks Using Heat Diffusion Processes for Marketing Candidates Selection," pp. 233–242, 2008.

[25] M. E. J. Newman, "Analysis of weighted networks," *Phys. Rev. E*, vol. 70, no. 5, p. 56131, 2004.

[26] A. Barrat, M. Barthelemy, R. Pastor-Satorras, and A. Vespignani, "The architecture of complex weighted networks," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 101, no. 11, pp. 3747–3752, 2004.

[27] T. Opsahl, F. Agneessens, and J. Skvoretz, "Node centrality in weighted networks: Generalizing degree and shortest paths," *Soc. Networks*, vol. 32, no. 3, pp. 245–251, 2010.

[28] T. Pagerank, C. Ranking, and B. Order, “The pagerank citation ranking: bringing order to the web,” *Tech. Report, Stanford Univ.*, pp. 1–17, 1999.

[29] N. Duhan, A. K. Sharma, and K. K. Bhatia, “Page Ranking Algorithms:A Survey,” in *IEEE International Advance Computing Conference*, no. March, pp. 6–7, 2009.

[30] A. Jain, R. Sharma, G. Dixit, and V. Tomar, “Page Ranking Algorithms in Web Mining, Limitations of Existing Methods and a New Method for Indexing Web Pages,” in *2013 International Conference on Communication Systems and Network Technologies*, pp. 640–645, 2013.

[31] Q. Shambour and J. Lu, “A trust-semantic fusion-based recommendation approach for e-business applications,” *Decis. Support Syst.*, pp. 38–41, 2012.

[32] W. W. Zachary, “An information flow model for conflict and fission in small groups,” *J. Anthropol. Res.*, vol. 33, pp. 452–473, 1977.

[33] D. Lusseau, K. Schneider, O. J. Boisseau, P. Haase, E. Slooten, and S. M. Dawson, “The bottlenose dolphin community of Doubtful Sound features a large proportion of long-lasting associations,” *Behav Ecol Sociobiol*, vol. 54, no. 4, pp. 396–405, Sep. 2003.

[34] D. E. Knuth, *The Stanford GraphBase: a platform for combinatorial computing*, vol. 4. Addison-Wesley Reading, pp. 1–3, 1993.

[35] V. Krebs, “http:\\www.orgnet.com,” ,2014.

[36] M. Girvan and M. E. J. Newman, “Community structure in social and biological networks.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 99, no. 12, pp. 7821–6, Jun. 2002.

[37] L. A. Adamic and N. Glance, “The political blogosphere and the 2004 US election: divided they blog,” in *Proceedings of the 3rd international workshop on Link discovery*, pp. 36–43, 2005.

[38] J. Vlasblom and S. J.Wodak, “Markov clustering versus affinity propagation for the partitioning of protein interaction graphs,” *BMC Bioinformatics*, Oct. 2009.