تاثیر خواص شبکه بر کارایی الگوریتم‌های کشف اجتماع

مرضیه کریمیان خوزانی [*1]، صادق سلیمانی [2]، محبوبه شمسی [3]

۱- گروه مهندسی کامپیوتر، مؤسسه غیرانتفاعی شهاب دانش، قم، ایران

۲- گروه مهندسی کامپیوتر، دانشکده فنی مهندسی، ، دانشگاه کردستان، ایران

۳- گروه مهندسی کامپیوتر، دانشکده فنی مهندسی، دانشگاه صنعتی، قم، ایران

خلاصه

کشف اجتماع (Community Detection) به طور گسترده درچند دهه‌ی گذشته عمدتاً به دلیل این‌که اجتماع در شبکه‌های مختلف از جمله شبکه‌های فناوری، اجتماعی، اطلاعاتی و زیستی نقش مهمی دارد، اهمیت پیدا کرده است. در نظریه‌ی شبکه، یک شبکه‌ی پیچیده، یک گراف با خصوصیات توپولوژیکی خاص است که در شبکه‌های ساده مانند شبکه‌های تصادفی و منظم رخ نمی‌دهد ولی اغلب در شبکه‌های دنیای واقعی رخ می‌دهد. ساختار شبکه می‌تواند بر کارایی روش‌های کشف اجتماع تأثیرگذار باشد، بنابراین روش‌های کشف اجتماع، بسته به نوع شبکه ممکن است نتایج کارآمد یا ناکارآمدی داشته باشند؛ مثلاً برخی روش‌ها برای شبکه‌های جهان کوچک مثل شبکه‌های اجتماعی کارایی خوبی دارند و در عین حال برای شبکه‌های مقیاس آزاد مانند شبکه‌های زیستی از نوع تعامل پروتئین-پروتئین نمی‌توانند به درستی اجتماعات را کشف کنند. در این مقاله ابتدا به بررسی جامع روش‌های کشف اجتماع پرداخته شده است. سپس پایگاه‌داده‌های شبکه‌های مورد آزمایش هر کدام از روش‌ها، از طریق اینترنت یا مکاتبه با نویسندگان مقالات تهیه شده است. آنگاه با تعریف معیارهای کمی، بر اساس ویژگی‌های مقیاس آزاد یا جهان کوچک بودن، پایگاه داده‌های جمع‌آوری شده دسته‌بندی شده‌اند و در نهایت، روش‌های کشف اجتماع نیز بر اساس کارایی در شبکه‌های جهان کوچک یا مقیاس آزاد دسته‌بندی شده‌اند. این پژوهش می‌تواند مقدمه‌ای بر پیشنهاد روشی جامع برای کشف اجتماع در تمام شبکه‌ها باشد.

کلمات کلیدی: کشف اجتماع، جهان کوچک، مقیاس آزاد، پیمانه، خوشه‌بندی.

۱. مقدمه

در سال‌های اخیر، انقلاب کامپیوتر، دانشمندان را با مقدار زیادی از داده‌ها و منابع محاسباتی مواجه کرده است تا آن‌ها را آنالیز و پردازش کنند. اندازه‌ی شبکه‌های واقعی به‌طور قابل توجهی افزایش یافته است؛ طوری‌که به میلیون‌ها یا حتی میلیاردها

* Corresponding author: دانشجوی کارشناسی ارشد، دانشکده مهندسی، مؤسسه‌ی غیرانتفاعی شهاب دانش قم
Email: M.karimiyan@shahabdanesh.ac.ir

رأس رسیده‌اند. شناخت از این شبکه‌ها و تحلیل صحیح آن‌ها در حیطه‌های مختلف علمی و اجتماعی بسیار سودمند است. یکی از روش‌های شناخت بهتر این شبکه‌ها، کشف اجتماع است [۱].

بیشتر شبکه‌های اجتماعی، زیستی و فن آوری، خصوصیات توپولوژیکی‌ایی دارند که نه کاملاً منظم و نه صرفاً تصادفی هستند. ویژگی‌هایی که شامل یک الگوی خاص در توزیع درجه، ضریب خوشه‌بندی[۱] بالا، ساختار جامعه و ساختار سلسله مراتبی می‌باشد. در مقابل، بسیاری از مدل‌های ریاضی شبکه‌ها که در گذشته مورد مطالعه قرار گرفته‌اند مانند گراف‌های منظم و تصادفی، این ویژگی را نشان نمی‌دهند [۲]. دو کلاس شناخته شده و بسیار مورد مطالعه‌ی شبکه‌های پیچیده، شبکه‌های مقیاس آزاد [۳] و شبکه‌های جهان کوچک [۴, ۵] هستند که هر کدام ویژگی‌های ساختاری مشخص خاص دارند: توزیع قاعده‌ی توان برای مقیاس آزاد و میانگین طول مسیر کوتاه و خوشه بندی بالا برای جهان کوچک.

۱.۱. تشخیص اجتماع

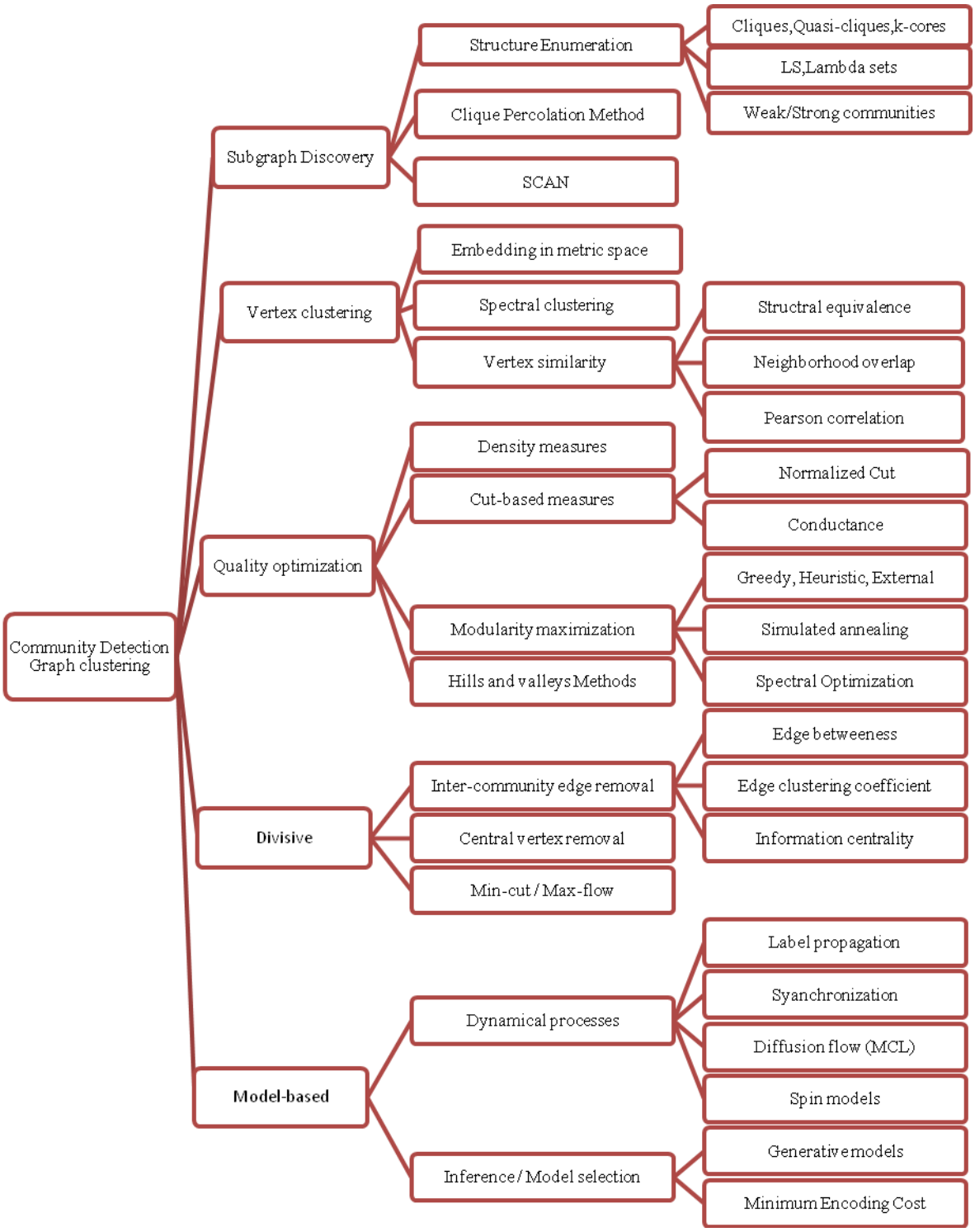
در یک شبکه، اصطلاح ساختار اجتماع[۲] که گاهی خوشه[۳] نیز نامیده می‌شود، به گروه‌هایی از رئوس اشاره می‌کند، که تعداد زیادی از لبه‌ها، آن‌ها را به هم متصل می‌کنند و تعداد کمی از لبه‌های مرتبط با آن‌ها، به رئوسی که درسایر خوشه‌های مختلف قرار گرفته‌اند متصل هستند. یک شبکه‌ی داده شده توسط یک گراف $G = \langle V, E \rangle$ نمایش داده می‌شود که در آن V مجموعه‌ی رئوس و E مجموعه‌ی لبه‌ها است. مسئله‌ی کشف اجتماع شامل پیدا کردن پارتیشن‌هایی ازرئوس در V به فرم $C = \{C_1,...,C_q\}$ است به طوری که هر C_i، $1 \le i \le q$ ساختار جامعه که در بالا توضیح داده شد را نمایش می‌دهد [۶].

مؤلفین مختلف بر اساس تعریف جامعه و انواع روش‌های موجود، روش‌های کشف اجتماع را دسته‌بندی کرده‌اند. در یکی از دسته‌بندی‌های جامع، پاپادپولس[۴] و همکارانش در سال ۲۰۱۲ این روش‌ها را به شکل ۱ طبقه‌بندی کردند [۷]. در این طبقه-بندی، پنج دسته‌ی اصلی به صورت زیر وجود دارد:

۱- کشف زیر گراف چسبنده[۵]: روش‌های این دسته مشخصات ویژگی‌های ساختاری را فرض می‌کنند که یک زیرگراف از شبکه، به منظور این که یک جامعه مطرح شود، باید برآورده سازد. اولین ساختار این زیرگراف، شامل روش‌های شمارش چنین ساختارهایی در شبکه‌ی تحت مطالعه است. تعاریف جامعه‌ی محلی مثل دسته[۶]، n دسته، k هسته، مجموعه‌های LS و مجموعه‌های لامبدا[۷] ، نمونه‌هایی از چنین ساختارهای منسجم هستند. علاوه بر این، روش‌هایی از قبیل روش نفوذ دسته[۸] و الگوریتم اسکن[۹] که منجر به کشف ساختارهای زیرگراف با خواص مشخص خوبی[۱۰] می-شوند، در همین کلاس قرار دارند.

---

[1] Clustering coefficient
[2] Community Structure
[3] Clusters
[4] Papadopoulos
[5] Cohesive subgraph discovery
[6] cliques
[7] lambda
[8] Clique Percolation Method
[9] SCAN
[10] well-specified properties



شکل ۱- طبقه بندی روش‌های کشف اجتماع وخوشه‌بندی گراف با توجه به تعریف جامعه و اصول روش اساسی [۷].

۲- خوشه‌بندی رئوس: چنین تکنیک‌هایی از خوشه‌بندی داده‌های سنتی ناشی شده است و بر اساس فاصله‌ی بین رئوس، خوشه‌بندی طیفی[۱] و دیگر معیارهای شباهت رئوس مثل هم‌ارزی ساختاری[۲] و هم‌پوشانی همسایگی[۳] می‌باشد.

۳- بهینه‌سازی کیفیت جامعه[۴]: روش‌های بسیاری بر اساس بهینه‌سازی اندازه‌گیری کیفیت جامعه مبتنی برگراف وجود دارند. بیشینه کردن پیمانه‌ای[۵]، چگالی[۶] زیرگراف و معیارهای مبتنی بر برش[۷]، مانند برش نرمال و هدایت[۸] در این دسته قرار دارند.

۴- تقسیم‌کننده[۹]: این روش‌ها درشناسایی عناصر شبکه (لبه ها و رئوس) که بین جوامع قرار دارند، تکیه می‌کنند. به عنوان مثال، الگوریتم اصلی توسط گیرون[۱۰] و نیومن[۱۱]، لبه های یک شبکه را بر اساس اندازه‌ی betweeness لبه حذف می‌کند تا زمانی‌که جوامع به صورت مؤلفه‌های غیرمتصل گراف پدیدار شوند. چند نمونه از اندازه‌گیری-های betweeness عبارت است از betweeness لبه، پیاده‌روی تصادفی[۱۲] و جریان متداول[۱۳].

۵- مبتنی بر مدل[۱۴]: این مدل یک رده‌ی گسترده از روش‌ها است که یا فرآیندهای پویایی که جوامع را معلوم می-کنند نشان می‌دهد، و یا آن‌ها یک مدل اساسی از طبیعت آماری را که می‌توانند بخشی از شبکه‌ی درون جوامع را به وجود آورند، در نظر می‌گیرند.

همچنین مقالات دیگری نیز در مورد مرور روش‌های کشف اجتماع در شبکه وجود دارد که عبارتند از طبقه‌بندی پورکاظمی و کیوان‌پور که در ۶ دسته‌ی [۸] کشف اجتماع برای شبکه‌ی اجتماعی علامت‌دار، کشف اجتماع در شبکه‌ی اجتماعی مثبت، کشف اجتماع برای شبکه‌ی اجتماعی ناهمگن[۱۵]، کشف اجتماع در شبکه‌ی اجتماعی ایستا، کشف اجتماع برای شبکه‌ی اجتماعی پویا و کشف اجتماع برای شبکه‌ی اجتماعی جهت‌دار می‌باشد. اما بر اساس بررسی‌های ما، تاکنون هیچ اثری به بررسی روش‌های کشف اجتماع بر اساس خواص شبکه از نظر جهان کوچک بودن یا مقیاس آزاد بودن نپرداخته است و اثر چنین ویژگی‌های مهمی را در عملکرد الگوریتم‌ها و روش‌های کشف اجتماع به حساب نیاورده است.

اکنون خلاصه‌ای از روش‌های کشف اجتماع که در سال‌های اخیر ارائه شده‌اند را مرور می‌کنیم و بر اساس روش پاپادپولس آن‌ها را به صورت طبقه‌بندی شده ارائه می‌کنیم.

---

[1] spectrum
[2] Structural equivalence
[3] Neighborhood overlap
[4] Community quality optimization
[5] modularity maximization
[6] Density
[7] Cut-based
[8] Conductance
[9] Divisive
[10] Girvan
[11] Newman
[12] random-walk
[13] current-flow
[14] Model-based
[15] Heterogeneo

۱.۱.۱. روش‌های بهینه‌سازی کیفیت جامعه

کافیری[1] و همکارانش، الگوریتمی به عنوان پس پردازش برای روش‌های اکتشافی پیشنهاد دادند که برای دو بخشی کردن، از یک پارتیشن شروع کرده و آزمایش را انجام دادند؛ اگر آزمایش ارزشمند باشد به برخی از جوامع تقسیم می‌شود و یا دو تا از آن‌ها را ادغام می‌کند [۶]. لی[2] یک الگوریتم محاسباتی سریع ارائه کرد که رئوس را با حداکثر اندازه‌گیری پیمانه‌ای استاندارد از پارتیشن‌های نتیجه، به گروه‌های مختلف تقسیم می‌کند و آن را روش تکراری طیفی بر اساس گرد کردن[3] (IR) نامیدند. [۹]. دمو[4] و همکارانش یک روش عمومی برای به حداکثر رساندن پیمانه‌ای پیشنهاد دادند. روش ارائه شده که به نام کشف خوشه‌ی شبکه‌ی پیچیده[5] (CONCLUDE) است در دو مرحله کار می‌کند: در مرحله‌ی اول، از مدل انتشار اطلاعات بر اساس پیاده-روی تصادفی و غیر برگشت به عقب[6] به طول محدود استفاده می‌کند. در مرحله‌ی دوم، CONCLUDE با استفاده از فاصله‌ی محاسبه شده در مرحله‌ی اول شبکه را به خوشه‌هایی افراز می‌کند [۱۰]. مهرل[7] و همکارانش، یک تعمیم طبیعی از پیمانه، بر اساس تفاوت بین تعداد واقعی و مورد انتظار پیاده روی در خوشه‌ها ارائه دادند که آن را پیمانه‌ی پیاده‌روی[8] نامیدند [۱۱]. جاروکاسمراتانا[9] و موراتا[10] یک الگوریتم دو مرحله‌ای برای کشف جامعه ارائه دادند. آن‌ها در طول آزمایش خود، با یکی دیگر از انوع زیر شبکه‌ی مقیاس آزاد که آن را "شبکه‌های مقیاس آزاد آمیخته" نامیدند، مواجه شدند. سپس الگوریتم کشف جامعه‌ای را پیشنهاد دادند که می‌تواند بر روی شبکه‌هایی کار کند که در یک زمان شامل هر دو نوع جامعه باشد [۱۲].

۲.۱.۱. روش تقسیم‌کننده

جیا[11] و همکارانش روش EACH را در سال ۲۰۱۴ ارائه کردند. دراین روش ابتدا امتیاز مرکزیت[12] غیر مثلث[13] یال‌ها را که برای اعمال تبعیض لینک‌های بیرونی از لینک‌های درونی برای تشخیص جامعه استفاده می‌شود، برای همه‌ی یال‌های یک شبکه‌ی داده شده محاسبه کرده و در هر تکرار، یال با بالاترین امتیاز را حذف می‌کند تا زمانی‌که امتیازات یال‌های باقی مانده همه صفر شوند [۱].

---

[1] Cafieri
[2] Li
[3] Iterative rounding based spectral method
[4] Meo
[5] COmplex Network CLUster DEtection
[6] Random and Non-backtracking Walks
[7] Mehrle
[8] Walk-modularity
[9] Jarukasemratana
[10] Murata
[11] S. Jia
[12] Centrality
[13] Antitriangle

۳.۱.۱. روش‌های خوشه‌بندی رئوس

لیم[۱] و همکارانش برای کشف هم‌پوشانی جوامع، یک چارچوب جدید از تبدیل فضای لینک[۲] که یک گراف اصلی داده شده را به یک گراف فضای لینک تبدیل می‌کند، ارائه کردند. بر اساس این چارچوب، آن‌ها الگوریتم LinkSCAN را توسعه دادند که خوشه‌بندی ساختاری را در گراف فضای لینک انجام می‌دهد. علاوه بر این، آن‌ها الگوریتم LinkSCAN* که کارایی LinkSCAN را با استفاده از نمونه‌گیری[۳] افزایش می‌دهد، ارائه کردند [۱۳]. سینگ[۴] و هومفریس[۵] عملگرهای تمایلی پس‌گرد را که با صراحت درختان معلق را می‌شمارند، معرفی کردند در نتیجه آن‌ها یک احتمال کوچک برای بازگشت به گره بلافاصله‌ی قبلی، بر خلاف عملگرهای غیر پس گرد که از بازگشت فوری اجتناب کردند، را پذیرفتند [۱۴]. برگس[۶] و همکارانش یک الگوریتم خوشه‌بندی اجتماع برای بهبود کشف جامعه در شبکه‌های ناقص پیشنهاد کردند. این چارچوب از الگوریتم‌های کشف جامعه‌ی موجود استفاده می‌کند تا فرآیندهای شبکه را به الگوریتم پیش‌بینی پیوند، نسبت دهد. سپس خروجی‌های متعدد خود را به یک خروجی نهایی ادغام می‌کند [۱۵]. لی[۷] و لوینا[۸] یک روش ساده و بسیار سریع برای برآورد تعداد جوامع مبتنی بر برخی از خواص طیفی عملگرهای گراف مانند ماتریس غیر پس گرد[۹] و ماتریس هشین بت[۱۰] پیشنهاد کردند [۱۶]. زینگ[۱۱] و همکارانش یک روش جدید توسط گسترش جامعه‌ی محلی به نام OCDLCE پیشنهاد دادند که ابتدا شبکه را با استفاده از اطلاعات ساختاری محلی به جوامع کوچک محلی تقسیم می‌کند، و سپس این جوامع را به ساختارهای جوامع هم-پوشان نهایی ادغام می‌کند [۱۷].

۴.۱.۱. روش‌های مبتنی بر مدل

نیومن در سال ۲۰۱۳ نشان داد که دو روش استنباطی که به طور گسترده استفاده می‌شود (مدل بلاک تصادفی[۱۲] یا نوع اصلاح شده‌ی درجه‌ی[۱۳] آن) را می‌توان به طور مستقیم، بر روی نسخه‌های مسئله‌ی پارتیشن بندی گراف استاندارد حداقل برش نگاشت کرد، که اجازه می‌دهد تا هر یک از الگوریتم‌های پارتیشن‌بندی را که به خوبی درک می‌شوند، برای حل مسئله‌ی کشف جامعه درخواست کنیم [۱۸]. فرئردگارد[۱۴] و هرلا[۱۵] مدل بلوک تصادفی چند شبکه‌ای[۱۶] (MNSBM) را فرموله کردند، که برای جدا کردن شبکه‌ی مشاهده شده به زیرشبکه‌هایی از انواع مختلف، جستجو را انجام می‌دهد که در آن مسئله‌ی ساختار استنتاج در هر زیرشبکه آسان‌تر می‌شود [۱۹].

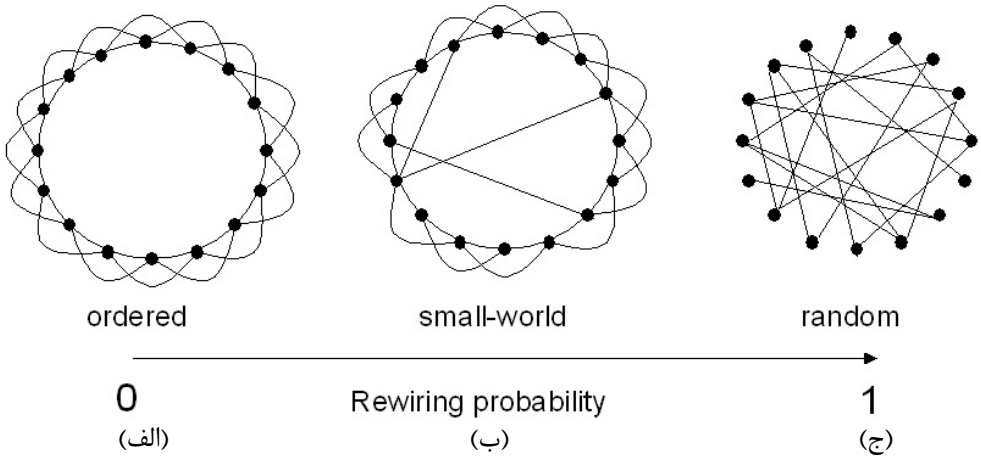
---

[1] S. Lim
[2] Link-space
[3] Sampling
[4] Singh
[5] Humphries
[6] Burgess
[7] Le
[8] Levina
[9] Non-backtracking matrix
[10] Bethe Hessian
[11] Xing
[12] stochastic block model
[13] degree-corrected variant
[14] B. Ø. Fruergaard
[15] T. Herlau
[16] Multiple-networks stochastic blockmodel

۲.۱. ویژگی‌های شبکه

الف) شبکه‌های جهان کوچک[1]

واتس[2] و استرگاتز[3] یک مدل مصنوعی برای شبکه‌های جهان کوچک که در آن ارتباط بین گره‌ها دریک گراف منظم، با یک احتمال معین دوباره سیم کشی[4] شده، ارائه کرده‌اند که در آن گراف‌هایی که در ساختار خود بین منظم و تصادفی بوده را به عنوان شبکه‌های جهان کوچک (SW) اشاره کرده‌اند. شبکه‌هایSW از لحاظ ساختاری به شبکه‌های اجتماعی بسیاری که خوشه‌بندی بالایی دارند، شبیه هستند و تقریباً میانگین مسیر یکسانی نسبت به شبکه‌های تصادفی با همان تعداد گره‌ها و لبه‌ها دارند. شبکه‌هایSW معمولاً خاصیت پیمانه‌ای[5] بالا دارند (گروهی از گره‌ها که به طور متراکم از مابقی شبکه به هم متصل‌اند). در شکل ۲، مثالی از شبکه‌های منظم، جهان کوچک و تصادفی وجود دارد [۲۰].



شکل ۲- اگر شبکه‌های فوق را شبکه‌های دوستی در نظر بگیریم، (الف) در این شبکه‌ی دوستی با قاعده‌ی کامل، هر نفر تنها با ۴ تا از نزدیک‌ترین همسایه‌ی خود دوست است. در این شبکه هر ۲ نفر به طور متوسط درجه‌های مجزای زیادی دارند. (ب) در این شبکه‌ی جهان کوچک، هر نفر هنوز به طور متوسط با ۴ تای دیگر ارتباط دارد، اما کمی دوستان دورتری دارد. در این شبکه متوسط درجه‌ی تفکیک[6] کم است. (ج) در شبکه‌ی تصادفی، هر نفر به طور متوسط با ۴ نفر دیگر ارتباط دارد، اما دوستانش پراکنده هستند؛ تعداد کمی از مردم، دوستان زیادی دارند و برخی دیگر به طور متوسط تنها چند درجه‌ی از هم جدا[7] دارند [۲۰].

[1] Small World
[2] Watts
[3] Strogatz
[4] Rewiring
[5] Modularity
[6] average degree of separation
[7] degrees apart

به بیان دقیق‌تر، شبکه‌های SW دارای ضریب خوشه‌بندی[1] بالا و میانگین طول مسیر[2] کم هستند. ضریب خوشه‌بندی c_i از گره i با درجه‌ی k_i، نسبت تعداد واقعی لینک‌های بین همسایگان i (e_i) و حداکثر تعداد ممکن از لینک‌های بین همسایگان آن می‌باشد (همسایه‌های i، گره‌هایی هستند که به طور مستقیم به گره i متصل هستند).

$$c_i = \frac{2e_i}{k_i(k_i - 1)} \quad (۱)$$

ضریب خوشه‌بندی C شبکه، میانگین تمام ضریب خوشه‌بندی‌های تکی[3] است:

$$C = \frac{1}{N}\sum_{i=1}^{N} c_i \quad (۲)$$

ضرایب خوشه‌بندی بین ۰ و ۱ متفاوت است. ضریب خوشه‌بندی بالا بدان معنی است که گره‌های همسایه به خوبی به هم متصل هستند [۲۱].

مشخصه‌ی[4] طول مسیر، یک معیار شبکه است که نشان می‌دهد چگونه یک شبکه یکپارچه می‌شود و این‌که چگونه اطلاعات می‌تواند در داخل شبکه جریان داشته باشد. طول مسیر یا فاصله‌ی (یا مسیر geodesic) d_{ij} بین دو نود i و j، تعداد کوچکترین لینکی است که i را به j متصل می‌کند. مشخصه‌ی طول مسیر L یک شبکه، متوسط فاصله بین تمام جفت از گره‌ها است [۲۱]:

$$L = \frac{1}{N(N-1)} \sum_{i,j \in N, i \neq j} d_{i,j} \quad (۳)$$

ب) شبکه‌های مقیاس آزاد[5]

یک گروه بزرگ از شبکه‌ها، شبکه‌های مقیاس آزاد (SF) هستند. این نوع از شبکه‌ها دارای توزیع درجه‌ی بسیار ناهمگنی[6] هستند که از قاعده‌ی توان[7] به شکل $P(k) \sim k^{-\gamma}$ پیروی می‌کنند. توزیع درجه، متشکل از درجه‌ی تمام گره‌ها در شبکه می‌باشد (درجه‌ی بزرگ‌تر نشان دهنده‌ی تعداد اتصالات بیشتر می‌باشد). علت نام‌گذاری به مقیاس آزاد آن است که بزرگ یا کوچک کردن هر قسمت از توزیع درجه، شکل آن را تغییر نمی‌دهد؛ در هر سطح از بزرگ‌نمایی، تعداد کم، اما قابل توجهی ازگره با تعداد زیادی از اتصالات وجود دارد و در دنباله‌ی انتهایی، گره‌هایی با تعداد بسیار کمی از اتصالات وجود دارد. شکل ۳ یک شبکه‌ی مقیاس آزاد را نشان می‌دهد [۲۰].

[1] Clustering Coefficient
[2] Average Path Length
[3] Individual
[4] Characteristic
[5] Scale free
[6] Heterogeneous
[7] Power-Law

شکل ۳- یک شبکه‌ی مقیاس آزاد با ۱۳۰ نود. ۵ تا از بزرگترین نودها با قرمز مشخص شده‌اند و آن‌ها با ۶۰ درصد از نودهای دیگر (سبز) در تماس هستند [۲۰].

ج) ویژگی‌های شبکه‌های جهان کوچک و مقیاس آزاد

بر اساس آن‌چه از خصوصیات شبکه‌های مقیاس آزاد و جهان کوچک بیان شد، می‌توان این شبکه‌ها را از لحاظ نوع توزیع درجه، میانگین طول مسیر، ضریب خوشه‌بندی و پیمانه‌ای مورد مقایسه قرار داد (جدول ۱).

جدول ۱- مقایسه‌ی ویژگی‌های شبکه‌های جهان کوچک و مقیاس آزاد. منظور از نوع توزیع درجه، توزیع پواسن[1] که این توزیع به صورت نمایی از قله‌اش به سمت پایین می‌آید، یا توزیع قاعده‌ی توان که اجازه می‌دهد برای تعداد کمی از گره با درجه‌ی بسیار بزرگ به صورت نمایی به تعداد زیادی از گره‌های با درجه‌ی کم نزول یابد، می‌باشد.

نوع شبکه	توزیع درجه	نوع توزیع درجه	میانگین طول مسیر	ضریب خوشه‌بندی	پیمانه‌ای
جهان کوچک	همگن	توزیع پواسن	کم	زیاد	بالا
مقیاس آزاد	بسیار ناهمگن	توزیع قاعده‌ی توان	بستگی به تعداد نودهای با درجه‌ی زیاد دارد.	بستگی به تعداد نودهای با درجه‌ی زیاد دارد.	پایین یا بالا

۳.۱. هدف و ساختار مقاله

در این مقاله ابتدا به بررسی روش‌های کشف اجتماع پرداخته شده است. سپس مجموعه‌داده‌های[2] هر یک از روش‌ها مورد آزمایش قرار گرفته است. آن‌گاه بر اساس ویژگی‌های مقیاس آزاد یا جهان کوچک، مجموعه داده‌های جمع‌آوری شده، دسته‌بندی شده‌اند و در نهایت، روش‌های کشف اجتماع نیز بر اساس کارایی در شبکه‌های جهان کوچک یا مقیاس آزاد دسته‌بندی شده‌اند. این پژوهش می‌تواند زمینه‌ای را برای پیشنهاد روشی جامع برای کشف اجتماع در تمام شبکه‌ها فراهم آورد.

[1] Poisson
[2] Dataset

در بخش ۲ مقاله به بررسی روش پیشنهادی می‌پردازیم. ابتدا مشخصات مجموعه داده‌ها را بیان می‌کنیم. سپس معیار تشخیص شبکه‌های جهان کوچک و مقیاس آزاد را شرح می‌دهیم. در بخش ۳، نتیجه‌ی اعمال معیارهای مورد نظر بر مجموعه داده‌ها و روش‌های تشخیص اجتماع را توضیح می‌دهیم. در بخش ۴ به بحث و تفسیر نتایج می‌پردازیم. در نهایت در بخش ۵ نتیجه‌گیری و کارهای آتی را بیان می‌کنیم.

۲. روش پیشنهادی

با توجه به این‌که در این مقاله می‌خواهیم روش‌ها و مجموعه داده‌های مورد استفاده در کشف اجتماع را بر اساس خواص شبکه دسته‌بندی کنیم، لازم است ابتدا مشخصات مجموعه داده‌ها را بیان کنیم، سپس نحوه‌ی تعیین خواص جهان کوچک و مقیاس آزاد، برای مجموعه داده‌ها را مشخص می‌کنیم تا بتوان از روی آن، روش‌ها و الگوریتم‌های پیشنهادی در این زمینه را طبقه‌بندی کرد.

۳.۱. مجموعه داده‌ها

به منظور دست‌یابی به هدف تحقیق، علاوه بر مرور روش‌های مختلف کشف اجتماع، مجموعه داده‌های آن‌ها نیز لازم به تهیه و بررسی است. از آن‌جا که تعداد مجموعه داده‌های مقالات بررسی شده زیاد است تنها مشخصات مجموعه داده‌ها را بیان می‌کنیم (جدول ۲).

جدول ۲- مشخصات مجموعه داده‌های بررسی شده بر اساس نام، تعداد نود، تعداد یال و کاربرد آن‌ها.

نام مجموعه داده	تعداد نودها	تعداد یال‌ها	مرجع
شبکه‌های اجتماعی[1]			
dolphins	٦٢	١٥٩	شبکه‌ای از دابتفولد ساند[2] و نیوزلند[3] با یال‌هایی که روابط اجتماعی بین افراد را، نشان می‌دهد [٢٢].
Email	١١٣٣	٥٤٥١	شبکه‌ی ایمیل URV [٢٣]
political books	١٠٥	٤٤١	شبکه‌ای درباره‌ی سیاست‌های ایالات متحده که یال‌ها نشان دهنده‌ی این است که کدام کتاب‌ها اغلب با هم در آمازون خریداری شده‌اند[4].
Zachary karate club network	٣٤	٧٨	گره‌های این شبکه افراد و یال‌ها، دوستی-هایی هستند که توسط زاکاری تعیین می-شود [٢٤].

[1] Social
[2] Doubtful Sound
[3] New Zealand
[4] http://www.orgnet.com/

Political Web Blogs Network	۱۴۹۰	۱۹۰۸۷	شبکه‌ای شامل وبلاگ‌های سیاسی در رابطه با سیاست‌های آمریکا که یال‌ها نشان دهنده‌ی لینک‌های وب و جوامع به صورت "محافظه‌کار" یا "لیبرال" می-باشند [۲۵].
Polblogs	۱۲۲۴	۱۶۷۱۵	[۲۵]
Enron-Email	۳۶۶۹۲	۱۸۳۸۳۱	شبکه‌ی ایمیل اینرون[1]
Brightkite	۵۸۲۲۸	۲۱۴۰۷۸	ارائه دهنده‌ی سرویس مبتنی بر مکان[2]
Facebook	۶۳۷۳۱	۸۱۷۰۹۰	شبکه‌ی فیس بوک نیواورلئان[3] [۲۶]
facebook100/Caltech36	۷۶۹	۳۳۳۱۲	شبکه‌ی فیس بوک دانشگاه caltech [۲۷]
facebook100/Haverford76	۱۴۴۶	۱۱۹۱۷۸	شبکه‌ی فیس بوک دانشگاه Haverford [۲۷]
facebook100/Reed98	۹۶۲	۳۷۶۲۴	شبکه‌ی فیس بوک دانشگاه Reed [۲۷]
facebook100/Simmons81	۱۵۱۸	۶۵۹۷۶	شبکه‌ی فیس بوک دانشگاه Simmons [۲۷]
facebook100/Swarthmore42	۱۶۵۹	۱۲۲۱۰۰	شبکه‌ی فیس بوک دانشگاه Swarthmore [۲۷]
PGP	۱۰۶۸۰	۲۴۳۱۶	شبکه کاربران الگوریتم Pretty-Good-Privacy[4]
Ego-facebook	۴۰۳۹	۸۸۲۳۴	شبکه‌ی فیس بوک اگو [۲۸]
شبکه‌ی Coappearance			
les miserables-degree	۷۷	۲۵۴	تعامل شخصیت در کتاب [۲۸]
شبکه‌ی ورزش			
Football	۱۱۵	۶۱۳	بازی‌های فوتبال دانشکده‌ی آمریکا در سال ۲۰۰۰ [۲۹]
شبکه‌های فن‌آوری[5]			

[1] http://snap.stanford.edu/data/
[2] http://snap.stanford.edu/data/
[3] New Orleans
[4] http://www.cs.bris.ac.uk/~steve/networks/copra/
[5] Technological

USAir97	۳۳۲	۲۱۲۶	ارتباطات بین فرودگاه‌های آمریکا[1]
S838-electronic circuits	۵۱۲	۸۱۹	شبکه‌ی مدارات الکترونیکی[2]
PowerGrid	٤۹٤۱	٦٥۹٤	شبکه‌ی توپولوژی برق [٤]
شبکه‌های همکاری[3]			
erdos02	٦۹۲۷	۱۱۸٥۰	شبکه‌ی همکاری نویسندگان[4]
DBLP	۳۱۷۰۸۰	۱۰٤۹۸٦٦	شبکه‌ی همکاری نویسندگان[5]
Amazon	۳۳٤۸٦۳	۹۲٥۸۷۲	شبکه‌ی خرید محصول[6]
شبکه‌های زیستی[7]			
p53 protein	۱۰٤	۲۲٦	شبکه‌ی تعاملات پروتئین [۲۳]
gene regulatory	۱۹۸۹	۹۱۷٥	شبکه‌ی تنظیم ژن [۳۰]
شبکه‌ی هدایت متن[8]			
Adjnoun	۱۱۲	٤۲٥	شبکه‌ی مجاورت صفات و اسم‌ها [۳۱]
شبکه‌های همکاری نویسندگان[9]			
Netscience-main	۳۷۹	۹۱٤	[٦]
Netscience	۱٥۸۹	۲۷٤۲	[۳۲]
Netsci	۱٤٦۱	۲۷٤۲	[۳۱]
Hepth	۷٦۱۰	۱٥۷٥۱	[۳۳]
Astroph	۱٦۰٤٦	۱۲۱۲٥۱	[۳۳]
Condmat	۱٦۲٦٤	٤۷٥۹٤	[۳۳]
ca-GrQc	٥۲٤۱	۱٤٤۸٤	[۳٤]
ca-HepTh	۹۸۷٥	۲٥۹۷۳	[۳٤]
ca-HepPh	۱۲۰۰٦	۱۱۸٤۸۹	[۳٤]
ca-AstroPh	۱۸۷۷۱	۱۹۸۰٥۰	[۳٤]
ca-CondMat	۲۳۱۳۳	۹۳٤۳۹	[۳٤]
شبکه‌های اینترنت			

[1] http://vlado.fmf.uni-lj.si/pub/networks/data/ND/NDnets.htm
[2] http://www.weizmann.ac.il/mcb/UriAlon/
[3] collaboration
[4] http://vlado.fmf.uni-lj.si/pub/networks/data/ND/NDnets.htm
[5] http://dblp.uni-trier.de/xml/
[6] http://snap.stanford.edu/data/
[7] Biological
[8] Text Driven
[9] Scientific Collaboration

Internet	۲۲۹۶۳	٤٨٤٣٦	تصاویر لحظه‌ای از اینترنت[1]
WWW	۳۲٥۷۲۹	١٤٦٩٦٨٠	شبکه‌ی اینترنت[2]
شبکه‌های استنادی[3]			
SciMet	۳۰٨٤	١٠٤١٥	شبکه‌ی استنادی[4]
SmaGri	١٠٥٩	٤٩٢١	شبکه‌ی استنادی[5]
شبکه‌ی مغز			
Hagmann	۹۹٨	٨٦٧٦٣	[۳٦ ,۳٥]

۲.۲. معیار تشخیص شبکه‌های جهان کوچک

برای بررسی این‌که شبکه دارای خواص جهان کوچک است، متوسط طول مسیر (L_{random}) و ضریب خوشه‌بندی (C^1_{random}) را باید در نظر گرفت [۳۷]؛ که به ترتیب بر اساس فرمول (۱) و (۲) محاسبه می‌شود:

$$L_{random} = \ln n / \ln \bar{k} \quad (۱)$$

$$C^1_{random} = \frac{\bar{k}}{n} \quad (۲)$$

که در آن n تعداد کل رأس و $\bar{k}$ متوسط درجه‌ی شبکه می‌باشد.

در این مقاله برای محاسبه‌ی C^1_{actual} و L_{actual} از نرم افزار Pajek[6] استفاده می‌کنیم. بر اساس تعریف واتس و استورگاتز، اگر شبکه‌ای دارای شرایط (۱) $C^1_{actual} >> C^1_{random}$ و (۲) $L_{actual} \geq L_{random}$ باشد، در واقع $\sigma > 1$ ($\sigma = \frac{C/C_{random}}{L/L_{random}} = \frac{\gamma}{\lambda}$) باشد؛ نشان دهنده‌ی این است که این شبکه یک شبکه‌ی جهان کوچک است [۲۱].

همان‌طور که در بالا اشاره شد میانگین طول مسیر یک شبکه باید با یک شبکه تصادفی مقایسه شود. اما، مقایسه‌ی ضریب خوشه‌بندی با شبکه‌ی تصادفی معادل، به درستی رفتار جهان کوچک را نمی‌دهد؛ زیرا ضریب خوشه‌بندی در یک شبکه‌ی جهان کوچک بیشتر شبیه به یک شبکه‌ی منظم است. علاوه بر این، به طور کلی پذیرفته شده است که ضریب خوشه‌بندی در شبکه‌ی اصلی بسیار بیشتر از یک شبکه تصادفی است. تلسفورد[7] و همکارانش در سال ۲۰۱۱ یک متریک جدید برای جهان کوچک به نام ω معرفی کردند. بنابراین آن‌ها از متریک $\omega = \frac{L_{random}}{L} - \frac{C}{C_{lattic}}$ ($C_{lattic} = \frac{3(\bar{k}-2)}{4(\bar{k}-1)}$) برای تشخیص بهتر شبکه‌ی جهان کوچک استفاده کردند. مقدار ω بین ۱ و ۱- می‌باشد. مقادیر نزدیک به صفر اشاره می‌کند که شبکه دارای خصوصیات جهان کوچک می‌باشد. مقادیر مثبت اشاره می‌کند که گراف بیشتر خاصیت شبکه‌ی تصادفی را دارد ($L \approx L_{random}$ و

[1] http://www.routeviews.org/
[2] http://vlado.fmf.uni-lj.si/pub/networks/data/ND/NDnets.htm
[3] Citation
[4] http://vlado.fmf.uni-lj.si/pub/networks/data/ND/NDnets.htm
[5] http://vlado.fmf.uni-lj.si/pub/networks/data/ND/NDnets.htm
[6] http://pajek.imfm.si/doku.php?id=download
[7] Telesford

$C<<C_{lattic}$). مقادیر منفی اشاره می‌کند که گراف بیشتر خاصیت شبکه‌ی منظم را دارد ($C \approx C_{lattic}$ و $L>>L_{random}$) [۲۱].

امتیاز دهی به شبکه‌های جهان کوچک

برای امتیاز دادن[1] به شبکه‌های SW با توجه به شرایط ۱ و ۲ و متریک‌های σ و ω در بالا می‌توانیم ۵ امتیاز را در نظر بگیریم:

۱- امتیاز ۰ (ندارد): اگر یکی از شرط‌ها و یا هر دو شرط را نداشته باشد.
۲- امتیاز ۱ (ضعیف): اگر ضریب خوشه‌بندی پایین (شرط ۱) و متوسط طول مسیر بالا (شرط ۲) داشته باشد (به عبارتی خصوصیات کمی از SW را داشته باشد).
۳- امتیاز ۲ (متوسط): اگر یک شرط را کامل داشته باشد اما دومی را به صورت ضعیف داشته باشد (خوشه‌بندی پایین یا متوسط طول مسیر بالا).
۴- امتیاز ۳ (خوب): اگر هر دو شرط را ارضا کند ($\sigma > 1$).
۵- امتیاز ۴ (عالی): $\sigma > 1$ و مقدار ω نزدیک به صفر باشد.

۲.۳. معیار تشخیص شبکه‌های مقیاس آزاد

توزیع درجه‌ی شبکه‌های مقیاس آزاد، یک توزیع قاعده‌ی توان است. برای پیدا کردن این که شبکه مقیاس آزاد است، از P-value استفاده می‌کنیم. در آمار، P-value تابعی از نتایج نمونه‌ی مشاهده شده (یک آمار) است که برای آزمایش یک فرضیه‌ی آماری استفاده می‌شود. قبل از انجام آزمون، یک مقدار آستانه به نام سطح معنی دار آزمون، که ۵٪ یا ۱٪ است و به صورت α نمایش داده می‌شود، انتخاب می‌شود. اگر P-value برابر یا کوچکتر از سطح معنی دار (α) باشد، نشان می‌دهد که داده‌های مشاهده شده با این فرض، که فرضیه‌ی صفر درست است، مغایر می‌باشد و در نتیجه این فرضیه را باید رد کرد [۶].

پس از انتخاب توزیع قاعده‌ی توان با پارامترهای x_{min} و α ، هر یک از این مجموعه داده‌ها را با داده‌های مشاهده شده مقایسه می‌کنیم تا ببینیم آیا این دو مجموعه داده با انجام آزمون KS (در آمار, آزمون KS یک آزمون برابری مستمر بدون پارامتر[2] و توزیع احتمال یک بعدی است که می‌تواند برای مقایسه یک نمونه[3] با توزیع احتمال مرجع (آزمون KS یک نمونه) و یا مقایسه‌ی دو نمونه (آزمون KS دو نمونه) مورد استفاده قرار بگیرد) از یک توزیع مشابه برای داده‌های مشاهده شده و اطلاعات تولید شده آمده است یا نه. برای مشخص کردن این‌که هر آزمون KS را رد می‌کنیم و یا می‌پذیریم از سطح معنی‌دار ۰/۱ استفاده می‌کنیم؛ بدان معنی‌که اگر مقدار P بزرگتر از ۰/۱ باشد فرض صفر پذیرفته می‌شود و نتیجه می‌گیریم که هر دو مجموعه داده از توزیع یکسان آمده است، در غیر این صورت اگر مقدار P-value کمتر یا برابر با ۰/۱ باشد فرض صفر رد می‌شود و نتیجه می‌گیریم که مجموعه داده از توزیع قاعده‌ی توان پیروی نمی‌کند [۳۸, ۳۹].

امتیاز دهی به شبکه‌های مقیاس آزاد

به طور شهودی شبکه‌ای را مقیاس آزاد گوییم که هر گاه بخشی از آن را جدا کنیم، الگوی توزیع درجات در آن بخش همانند الگوی توزیع درجات در کل شبکه است. اما به صورت رسمی، شبکه‌ای مقیاس آزاد است که توزیع درجه‌ی آن از یک

[1] Ranking
[2] nonparametric test of the equality of continuous
[3] sample

تابع خاص ریاضی به نام قانون قاعده‌ی توان پیروی کند. این عملاً بدان معنی است که علاوه بر این‌که درجه‌ی بسیاری از یال‌ها پایین است، تعداد کمی از یال‌ها درجه‌ی بسیار بالایی دارند (که به بسیاری از یال‌های دیگر متصل هستند).

با توجه به مطالب بیان شده در قسمت قبلی، برای مقیاس آزاد ۳ امتیاز در نظر می‌گیریم:

۱- اگر $p=0$ باشد امتیاز صفر (ندارد)

۲- اگر $0.1 > p >= 0.05$ امتیاز ۱ (متوسط)

۳- اگر $1 >= p > 0.1$ امتیاز ۲ (قوی)

در نهایت اگر امتیاز مقیاس آزاد ۲ باشد آن را SF در نظر می‌گیریم.

امتیاز دهی به مقیاس آزاد بر اساس مقدار p که با استفاده از فایل‌های کد متلب[1] که شامل plfit.m، plplot و plpva [۳۸] می‌باشد، محاسبه می‌شود. فایل‌های کد متلب، توابعی برای بررسی این موضوع است که آیا داده‌ها، یک توزیع قاعده‌ی توان است یا نه.

۳. نتایج محاسباتی

در این مقاله ۱۳ روش و ۴۴ مجموعه داده مرتبط با کشف اجتماع مورد آزمایش قرار گرفته است. بر اساس مباحثی که در قسمت ۲ بیان شد، مجموعه داده‌ها از لحاظ مقیاس آزاد یا جهان کوچک بررسی شده‌اند که در جدول ۳ نشان داده شده است (در جدول ۳، R مخفف کلمه‌ی Random و L مخفف کلمه‌ی Lattice می‌باشد). جدول ۴ نیز روش‌های کشف اجتماع را بر اساس جهان کوچک یا مقیاس آزاد بودن نشان می‌دهد.

با توجه به جدول ۳، مشاهده می‌شود که ۲۴ تا از دیتاست‌ها دارای خاصیت جهان کوچک، ۵ تا خاصیت مقیاس آزاد و ۹ تا هر دو خاصیت را دارند در حالی‌که ۵ تا از آن‌ها هیچ یک از خصوصیات شبکه‌های جهان کوچک یا مقیاس آزاد را ندارند. در جدول ۳ نیز مشاهده می‌شود که ۴ تا از روش‌ها دارای خاصیت جهان کوچک و ۹ تا هر دو خاصیت را دارند.

۴. بحث

در بررسی مجموعه داده‌ها از لحاظ مقیاس آزاد یا جهان کوچک (جدول ۳) مشاهده می‌شود که الزامی ندارد شبکه‌ها حتماً یکی از دو خاصیت جهان کوچک یا مقیاس آزاد را داشته باشند، بلکه ممکن است هر دو یا هیچ یک از خاصیت‌ها را نداشته باشند. مثلاً مقدار ω در مجموعه داده‌های gene regulatory و Brightkite مثبت می‌باشد که به این موضوع اشاره می‌کند که گراف بیشتر خاصیت شبکه‌ی تصادفی را دارد ($L \approx L_{random}$ و $C<<C_{lattic}$) یا در مجموعه داده Facebook، ω منفی است یعنی گراف بیشتر خاصیت شبکه‌ی منظم را دارد ($C \approx C_{lattic}$ و $L>>L_{random}$). از آن‌جا که بعضی از مجموعه داده‌ها، ممکن است تعدادی از شرایط شبکه‌های جهان کوچک یا مقیاس آزاد را نداشته باشند یا آن را به صورت کم‌تری داشته باشند شیوه‌ای از امتیازدهی برای این شبکه‌ها پیشنهاد شده که در قسمت ۲.۲ و ۲.۳ نحوه‌ی امتیازدهی شرح داده شده است.

برای دسته‌بندی روش‌ها (جدول ۳)، مجموعه داده‌هایی که در هر مقاله استفاده شده را بررسی کرده و تعداد SW یا SF در هر روش را به دست می‌آوریم. تعداد هر یک از SW یا SF یا هر دو بیشتر بود آن روش را در آن نوع شبکه قرار می‌دهیم. از

[1] MATLAB

آن‌جا که روشی در SF نداریم ما روش عملگرهای تمایلی پس‌گرد را SF در نظر می‌گیریم زیرا تعداد SF در آن بیشتر از SW می‌باشد.

برخی مقالات، اگاهانه یا ناآگاهانه، مجموعه داده‌هایی را تست گرفته‌اند که با الگوریتم خودشان همخوانی دارند و به عبارت دیگر اگر روش‌شان برای جهان کوچک متناسب بوده، مجموعه داده‌ی متناظر را تست گرفته‌اند و احتمالاً اگر با مجموعه داده نامناسب تست می‌گرفتند جواب خوبی نمی‌گرفتند. بنابراین در کل درست نیست که هر روش کشف اجتماع را با هم مقایسه کنیم مگر این‌که خواص زمینه‌ای شبکه‌ای مشابهی داشته باشند، وگرنه یکی ممکن است برای یک مجموعه داده بهتر باشد در حالی‌که دیگری برای مجموعه داده‌ی دیگری برتر باشد. به طور مثال در روش MNSBM [۱۹] یا روش CONCLUDE [۱۰] که از روش‌های جهان کوچک هستند بیشتر مجموعه داده‌های استفاده شده در آن، جهان کوچک هستند، به عبارتی روش‌های آن‌ها برای جهان کوچک متناسب بوده و مجموعه داده‌های متناظر با آن را مورد آزمایش قرار داده‌اند.

۵. نتیجه‌گیری و کارهای آتی

در این مقاله، ما یک طبقه‌بندی جدید از روش‌های کشف اجتماع و مجموعه داده‌ها بر اساس خصوصیات شبکه ارائه دادیم تا با کمک آن بتوانیم از روش‌های کشف اجتماع به صورت کاربردی‌تر استفاده کنیم و همچنین زمینه را برای ارائه‌ی روش‌های کشف اجتماع با کاربرد جامع‌تر فراهم سازیم. از آن‌جا که ساختار شبکه می‌تواند بر عملکرد روش‌های کشف اجتماع تأثیرگذار باشد، پس این روش‌ها، می‌تواند به نوع شبکه بستگی داشته باشد و هر روش الزاماً در هر نوع شبکه‌ای بهترین عملکرد را نخواهد داشت. ممکن است تعدادی از روش‌ها تنها برای شبکه‌های جهان کوچک کارایی خوبی داشته باشند در حالی‌که تعدادی دیگر فقط برای شبکه‌های مقیاس آزاد مناسب باشند یا حتی برای هر دو نوع شبکه خوب جواب دهند. با این دسته‌بندی جدید روش‌های کشف اجتماع، می‌توان در آینده روشی را ارائه کرد که در هر دو نوع شبکه‌ی جهان کوچک و مقیاس آزاد عملکرد خوبی داشته باشد.

۶. سپاس‌گزاری

از نویسندگان مقاله‌ی [۶]، لیبرتی[1] از دانشکده‌ی صنعتی اکول[2] به خاطر فراهم کردن مجموعه داده‌های p53 protein و netscience main، و نیز از لی از مؤسسه‌ی دانشکده‌ی صنعتی ماکایو[3] برای مجموعه داده‌های مورد استفاده در مقاله‌ی [۹] ، همچنین از جیا[4] از دانشگاه زیدین[5] نویسنده‌ی مقاله‌ی [۱] برای مجموعه داده gene regulatory و از هرلا[6] نویسنده‌ی مقاله‌ی [۱۹] از دانشکده‌ی فنی دانمارک[7] که در تهیه مجموعه داده‌های Hagmann و facebook/100 به ما کمک کردند؛ تشکر و قدرانی می‌کنیم.

[1] Liberti
[2] Ecole Polytechnique
[3] Macau Polytechnic Institute
[4] Jia
[5] Xidian
[6] Herlau
[7] Technical University of Denmark

جدول ۳: بررسی مجموعه داده‌ها از لحاظ مقیاس آزاد یا جهان کوچک.

ردیف	نام مجموعه داده	C^1_{random}	C^1_{actual}	L_{random}	L_{actual}	$\sigma = \frac{\gamma}{\lambda}$	C_{lattic}	ω	مقدار p	امتیاز مقیاس آزاد	امتیاز جهان کوچک	$\bar{k}$	SF	SW
۱	dolphins	۰/۰۸۲۷	۰/۳۰۲۹	۲/۵۲۴	۳/۳۵۶۹	۲/۷۵	۰/۵۶۸۴	۰/۲۲	۰/۱۳	۲	۴	۵/۱۳	+	+
۲	les miserables-degree	۰/۰۸۵۷	۰/۷۳۵۵	۲/۳۰۱۹	۲/۶۴۱۱	۷/۴۸	۰/۶۱۶۱	−۰/۳۲	۰/۷۴۲	۲	۴	۶/۶	+	+
۳	political books	۰/۰۸	۰/۴۸۷۵	۲/۱۸۶۸	۳/۰۷۸۷	۴/۳۲	۰/۶۴۸۶	−۰/۰۴	۰/۲۶	۲	۴	۸/۴	+	+
۴	Football	۰/۰۹۲۷	۰/۴۰۳۲	۲/۰۰۵	۲/۵۰۸۲	۳/۴۷	۰/۶۷۲۴	۰/۲	۰	۰	۴	۱۰/۶۶	−	+
۵	USAir97	۰/۰۳۸۶	۰/۷۴۹۳	۲/۲۷۷	۲/۷۳۸۱	۱۶/۱۴	۰/۶۸۶۴	−۰/۲۶	۰	۰	۴	۱۲/۸۰	−	+
۶	S838-electronic circuits	۰/۰۰۶۲	۰/۰۵۸۴	۵/۳۶۳۳	۶/۸۵۸۵	۷/۳۶	۰/۴۰۹۱	۰/۶۴	۰/۰۰۲	۱	۲	۳/۲	−	−
۷	e-mail network of URV	۰/۰۰۸۴	۰/۲۵۴	۳/۱۰۶۵	۳/۶۰۶۰	۲۶/۰۵	۰/۶۶۳	۰/۴۸	۰	۰	۴	۹/۶۲	−	+
۸	PowerGrid	۰/۰۰۰۵	۰/۱۰۶۵	۸/۶۶۰۵	۱۸/۹۸۹۲	۹۷/۱۴	۰/۳۰۰۹	۰/۱	۰	۰	۲	۲/۶۷	−	−
۹	erdos02	۰/۰۰۰۴	۰/۳۹۸۳	۷/۱۹۱۷	۳/۷۷۶۴	۱۸۹۶	۰/۴۴۰۱	۰/۹۹۹۳	۰	۰	۰	۳/۴۲	−	−
۱۰	p53 protein	۰/۰۴۱۸	۰/۳۹۲۵	۳/۱۵۹۱	۳/۱۰۰۴	۹/۵۷	۰/۵۲۶۱	۰/۲۷	۰/۲۱	۲	۳	۴/۳۵	+	+
۱۱	Netscience-main	۰/۰۱۲۷	۰/۷۹۸۱	۳/۷۷۵۲	۶/۰۴۱۹	۳۹/۲۷	۰/۵۵۳۷	−۰/۸۲	۰/۷۹۶	۲	۴	۴/۸۲	+	−
۱۲	Zachary karate club network	۰/۱۳۴۹	۰/۵۸۷۹	۲/۳۱۴۷	۲/۴۰۸۲	۴/۱۹	۰/۵۴۱۱	−۰/۱۲	۰/۱۳۲	۲	۴	۴/۵۹	+	+
۱۳	Adjnoun	۰/۰۶۷۸	۰/۱۸۹۸	۲/۳۲۸	۲/۵۳۵۵	۲/۵۷	۰/۶۳۶۲	۰/۶۲	۰/۱۸	۲	۲	۷/۵۹	+	−
۱۴	Polblogs(1224)	۰/۰۲۲۳	۰/۳۶	۲/۱۴۹۸	۲/۷۳۷۵	۱۲/۶۸	۰/۷۲۶	۰/۲۹	۰/۴۵۸	۲	۴	۲۷/۳۱	+	+
۱۵	Netsci	۰/۰۲۶	۰/۸۷۸۲	۵/۵۱۳	۵/۸۲۳۲	۳۱۹/۷۸	۰/۴۷۷۳	−۰/۸۹	۰	۰	۳	۳/۷۵	−	+
۱۶	Hepth	۰/۰۰۰۵	۰/۶۳۶۴	۶/۲۹۰۷	۷/۰۲۵۴	۱۱۳۹/۶۹	۰/۵۱۱۱	−۰/۳۵	۰	۰	۴	۴/۱۴	−	+
۱۷	Astroph	۰/۰۰۰۹	۰/۷۲۶۲	۳/۵۶۶۱	۴/۷۹۸	۵۹۹/۷۲	۰/۶۹۶۸	−۰/۳	۰/۰۵۶	۱	۴	۱۵/۱۱	−	+
۱۸	Condmat	۰/۰۰۰۳	۰/۷۳۷۱	۵/۴۸۹۴	۶/۶۲۷۶	۲۰۳۵/۰۴	۰/۵۹	−۰/۴۲	۰/۱۴۴	۲	۴	۵/۸۵	+	+
۱۹	Internet	۰/۰۰۰۲	۰/۳۴۹۹	۶/۹۷۴۱	۳/۸۴۲۴	۳۱۷۵/۴۱	۰/۵۱۷۱	۱/۱۴	۰/۶۴۲	۲	۰	۴/۲۲	+	−
۲۰	Political Web Blogs Network(1490)	۰/۰۱۷۱	۰/۲۳۶۰	۲/۲۵۵۲	۳/۳۹۰۲	۹/۱۸	۰/۷۱۹۴	۰/۳۴	۰/۷۲۴	۲	۴	۲۵/۵۳	+	+
۲۱	gene regulatory	۰/۰۰۴۶	۰/۱۰۱۵	۳/۴۱۹۲	۴/۹۸۰۲	۱۵/۱۵	۰/۶۵۸۷	۰/۵۳	۰/۰۹۴	۱	۲	۹/۲۲	−	− (R)

۲۲	Enron-Email	۰/۰۰۰۲	۰/۷۱۵۶	۴/۵۶۰۶	۴/۰۲۵۱	۴۰۵۴	۰/۶۶۶۸	۰/۰۶	۰	۰	۴	−	+
۲۳	Brightkite	۰/۰۰۰۱	۰/۲۷۰۷	۵/۵۰۰۶	۴/۹۱۷۳	۳۰۲۲	۰/۶۳۱۹	۰/۶۹	۰	۰	۳	−	(R) −
۲۴	Amazon	۰/۰۰۰۰۱	۰/۴۲۹۷	۷/۴۳۸۶	۱۱/۹۵۳۹	۲۶۷۳۹	۰/۵۸۴۴	−۰/۱۱	۰/۰۱۶	۱	۴	−	+
۲۵	Facebook	۰/۰۰۰۴	۰/۲۵۳۲	۳/۴۱	۱۱/۷۱۹۸	۱۸۴	۰/۷۱۹۶	−۰/۰۶	۰	۰	۲	−	(L) −
۲۶	WWW	۰/۰۰۰۰۱	۰/۳۱۱۷	۸/۴۲۷۲	۱۱/۲۶۵۵	۴۷۳۵	۰/۵۳۶۳	۰/۱۷	۰	۰	۴	−	+
۲۷	ca-GrQc	۰/۰۰۱	۰/۶۸۶۵	۵/۰۰۷۸	۶/۰۴۸۵	۵۶۸/۳۸	۰/۵۸۴۴	−۰/۳۵	۰	۰	۴	−	+
۲۸	ca-HepTh	۰/۰۰۰۵	۰/۵۹۹۶	۵/۵۴۰۴	۵/۹۴۵۲	۱۱۱۷/۵۵	۰/۵۷۳۹	−۰/۱۱	۰	۰	۴	−	+
۲۹	ca-HepPh	۰/۰۰۱۶	۰/۶۹۸۴	۳/۱۴۹۳	۴/۶۷۲۶	۲۹۴/۲	۰/۷۱	−۰/۳۱	۰	۰	۴	−	+
۳۰	ca-AstroPh	۰/۰۰۱۱	۰/۶۷۶۸	۳/۲۲۷	۴/۱۹۴	۴۷۳/۴۱	۰/۷۱۲۷	−۰/۱۸	۰	۰	۴	−	+
۳۱	ca-CondMat	۰/۰۰۰۳	۰/۷۰۵۸	۴/۸۰۹۵	۵/۳۵۲۱	۲۱۱۴/۱۵	۰/۶۴۴۱	−۰/۲	۰/۱۶۸	۲	۴	+	+
۳۲	Facebook-links	۰/۰۰۰۴	۰/۲۵۳۲	۳/۴۱	۴/۳۲۱۹	۴۹۹/۴۴	۰/۷۱۹۶	۰/۴۴	۰	۰	۴	−	+
۳۳	SciMet	۰/۰۰۲۲	۰/۱۰۱۱	۴/۲۰۷۳	۴/۲۲۸۹	۴۵/۷۲	۰/۶۱۹۶	۰/۸۳	۰/۸۴	۲	۲	+	−
۳۴	SmaGri	۰/۰۰۸۸	۰/۱۷۴۷	۳/۱۲۴۸	۳/۲۴۲	۱۹/۱۳	۰/۶۵۹۵	۰/۷	۰/۳۷۸	۲	۲	+	−
۳۵	Hagmann	۰/۱۷۴۲	۰/۴۱۹۳	۱/۳۳۸۸	۲/۲۵۶	۱/۴۳	۰/۷۴۵۷	۰/۰۳	۰	۰	۴	−	+
۳۶	facebook100/Caltech36	۰/۱۱۲۷	۰/۴۲۸۸	۱/۴۸۹۳	۲/۳۳۷۸	۲/۴۲	۰/۷۴۱۲	۰/۰۶	۰/۰۰۶	۰	۴	−	+
۳۷	facebook100/Haverford76	۰/۱۱۴	۰/۳۲۶۸	۱/۴۲۵۴	۲/۲۲۷۹	۱/۸۳	۰/۷۴۵۴	۰/۲	۰	۰	۴	−	+
۳۸	facebook100/Reed98	۰/۰۸۱۳	۰/۳۳۰۴	۱/۵۷۵۶	۲/۴۶۱۵	۲/۶	۰/۷۴۰۳	۰/۱۹	۰/۰۰۸	۰	۴	−	+
۳۹	facebook100/Simmons81	۰/۰۵۷۳	۰/۳۲۵۴	۱/۶۴۰۶	۲/۵۷۰۳	۳/۶۲	۰/۷۴۱۳	۰/۲	۰/۰۱۲	۰	۴	−	+
۴۰	facebook100/Swarthmore42	۰/۰۸۸۷	۰/۲۹۹۲	۱/۴۸۵۲	۲/۳۱۸۱	۲/۱۶	۰/۷۴۴۹	۰/۲۴	۰	۰	۴	−	+
۴۱	DBLP	۰/۰۰۰۰۲	۰/۷۳۲۱	۶/۷۰۰۶	۶/۷۹۱۸	۳۶۱۱۳	۰/۶۱۶۵	−۰/۲	۰	۰	۴	−	+
۴۲	Ego-facebook	۰/۰۱۰۸	۰/۶۱۷	۲/۱۹۸۴	۳/۶۹۲۵	۳۴/۰۱	۰/۷۳۲۴	−۰/۲۵	۰	۰	۴	−	+
۴۳	PGP	۰/۰۰۰۴	۰/۴۴۰۳	۶/۱۲۲۳	۷/۴۸۵۵	۹۰۰	۰/۵۳۸۷	۰/۰۰۰۵	۰	۰	۴	−	+
۴۴	Netscience	۰/۰۰۲۱	۰/۸۷۸۲	۵/۹۵۲	۵/۸۲۳۲	۴۲۷	۰/۴۴۳۹	−۰/۹۶	۰	۰	۳	−	+

جدول ۴: بررسی روش‌ها بر اساس جهان کوچک یا مقیاس آزاد.

نام روش	مجموعه داده‌های استفاده شده	مجموعه داده‌های بررسی شده	تعداد SF	تعداد SW	تعداد SW,SF	هیچکدام	نتیجه
روشی به عنوان پس پردازش برای روش-های مکاشفه‌ای	۱۱	۱۱	۰	۳	۵	۳	SW-SF
روش تکراری طیفی بر اساس گردکردن (IR)	۱۳	۱۳	۲	۴	۶	۱	SW-SF
EACH	۵	۵	۰	۱	۳	۱	SW-SF
LinkSCAN	۶	۵	۰	۳	۰	۲	SW
CONCLUDE	۷	۶	۰	۵	۱	۰	SW
مدل بلاک تصادفی	۲	۲	۰	۰	۲	۰	SW-SF
MNSBM	۹	۹	۲	۷	۰	۰	SW
پیاده‌روی پیمانه	۲	۲	۰	۰	۲	۰	SW-SF
عملگرهای تمایلی پس‌گرد	۳	۳	۱	۰	۲	۰	SW-SF
EDGEBOOST	۵	۵	۰	۲	۳	۰	SW-SF
روشی بر اساس خواص طیفی عملگرهای گراف مانند ماتریس غیر پس گرد و ماتریس هشین بت	۵	۵	۰	۱	۴	۰	SW-SF
OCDLCE	۸	۸	۰	۴	۴	۰	SW
روش بهینه‌سازی پیمانه‌ای و وزن یال	۴	۵	۲	۰	۳	۰	SW-SF

مراجع

1. S. Jia, L. Gao, Y. Gao, and H. Wang, (2014),"Anti-triangle centrality-based community detection in complex networks," *IET systems biology*, vol. 8, pp. 116-125.
2. J. Kim and T. Wilhelm, (2008), "What is a complex graph?," *Physica A: Statistical Mechanics and its Applications*, vol. 387, pp. 2637-2652.
3. A.-L. Barabási and E. Bonabeau, (2003), "Scale-free networks," *Scientific American*, vol. 288, pp. 50-59.
4. D. J. Watts and S. H. Strogatz, (1998),"Collective dynamics of 'small-world'networks," *nature*, vol. 393, pp. 440-442.
5. L. A. N. Amaral, A. Scala, M. Barthelemy, and H. E. Stanley, (2000),"Classes of small-world networks," *Proceedings of the national academy of sciences*, vol. 97, pp. 11149-11152.
6. S. Cafieri, P. Hansen, and L. Liberti, (2014),"Improving heuristics for network modularity maximization using an exact algorithm," *Discrete Applied Mathematics*, vol. 163, pp. 65-72.

7. S. Papadopoulos, Y. Kompatsiaris, A. Vakali, and P. Spyridonos, (2012), "Community detection in social media," *Data Mining and Knowledge Discovery,* vol. 24, pp. 515-554.
8. M. Pourkazemi and M. Keyvanpour, (2013),"A survey on community detection methods based on the nature of social networks," in *Computer and Knowledge Engineering (ICCKE), 2013 3th International eConference on*, pp. 114-120.
9. W. Li, (2014), "A Constrained Power Method for Community Detection in Complex Networks," *Mathematical Problems in Engineering*, vol. 2014.
10. P. De Meo, E. Ferrara, G. Fiumara, and A. Provetti, (2014),"Mixing local and global information for community detection in large networks," *Journal of Computer and System Sciences,* vol. 80, pp. 72-87.
11. D. Mehrle, A. Strosser, and A. Harkin, (2014),"Walk modularity and community structure in networks," *arXiv preprint arXiv:1401.6733*.
12. S. Jarukasemratana and T. Murata, (2015),"Community Detection in Scale-Free Networks using Edge Weight and Modularity Optimization Method," *人工知能学会論文誌*, vol. 30, pp. 84-95.
13. S. Lim, S. Ryu, S. Kwon, K. Jung, and J.-G. Lee, (2014),"LinkSCAN*: Overlapping community detection using the link-space transformation," in *Data Engineering (ICDE), 2014 IEEE 30th International Conference on*, pp. 292-303.
14. A. Singh and M. D. Humphries, (2015),"Finding communities in sparse networks," *Scientific reports,* vol. 5.
15. M. Burgess, E. Adar, and M. Cafarella, (2015), "Link-Prediction Enhanced Consensus Clustering for Complex Networks," *arXiv preprint arXiv:1506.01461*.
16. C. M. Le and E. Levina, (2015), "Estimating the number of communities in networks by spectral methods," *arXiv preprint arXiv:1507.00827*.
17. Y. Xing, F. Meng, Y. Zhou, and R. Zhou, (2015),"Overlapping Community Detection by Local Community Expansion," *JOURNAL OF INFORMATION SCIENCE AND ENGINEERING,* vol. 31, pp. 1213-1232, 2015.
18. M. Newman, "Community detection and graph partitioning, (2013),"*EPL (Europhysics Letters),* vol. 103, p. 28003.
19. B. Ø. Fruergaard and T. Herlau, (2014),"Efficient inference of overlapping communities in complex networks," *arXiv preprint arXiv:1411.7864*.
20. X. F. Wang and G. Chen, (2003),"Complex networks: small-world, scale-free and beyond," *Circuits and Systems Magazine, IEEE,* vol. 3, pp. 6-20.
21. Q. K. Telesford, K. E. Joyce, S. Hayasaka, J. H. Burdette, and P. J. Laurienti, (2011), "The ubiquity of small-world networks," *Brain connectivity,* vol. 1, pp. 367-375.
22. D. Lusseau, K. Schneider, O. J. Boisseau, P. Haase, E. Slooten, and S. M. Dawson, (2003),"The bottlenose dolphin community of Doubtful Sound features a large proportion of long-lasting associations," *Behavioral Ecology and Sociobiology,* vol. 54, pp. 396-405, 2003.
23. R. Guimera, L. Danon, A. Diaz-Guilera, F. Giralt, and A. Arenas, (2003),"Self-similar community structure in a network of human interactions," *Physical review E,* vol. 68, p. 065103.

24. W. W. Zachary, (1977),"An information flow model for conflict and fission in small groups," *Journal of anthropological research*, pp. 452-473, 1977.
25. L. A. Adamic and N. Glance, (2005),"The political blogosphere and the 2004 US election: divided they blog," in *Proceedings of the 3rd international workshop on Link discovery*, pp. 36-43.
26. B. Viswanath, A. Mislove, M. Cha, and K. P. Gummadi, (2009),"On the evolution of user interaction in facebook," in *Proceedings of the 2nd ACM workshop on Online social networks*, pp. 37-42.
27. A. L. Traud, P. J. Mucha, and M. A. Porter, (2012),"Social structure of Facebook networks," *Physica A: Statistical Mechanics and its Applications*, vol. 391, pp. 4165-4180.
28. J. Leskovec and J. J. Mcauley, (2012),"Learning to discover social circles in ego networks," in *Advances in neural information processing systems*, pp. 539-547.
29. M. Girvan and M. E. Newman, (2002),"Community structure in social and biological networks," *Proceedings of the national academy of sciences*, vol. 99, pp. 7821-7826.
30. C. Backes, A. Rurainski, G. W. Klau, O. Müller, D. Stöckel, A. Gerasch, *et al.*, (2012), "An integer linear programming approach for finding deregulated subgraphs in regulatory networks," *Nucleic acids research*, vol. 40, pp. e43-e43.
31. M. E. Newman, (2006),"Modularity and community structure in networks," *Proceedings of the National Academy of Sciences*, vol. 103, pp. 8577-8582, 2006.
32. M. E. Newman, (2006),"Finding community structure in networks using the eigenvectors of matrices," *Physical review E*, vol. 74, p. 036104.
33. M. E. Newman, (2001),"The structure of scientific collaboration networks," *Proceedings of the National Academy of Sciences*, vol. 98, pp. 404-409.
34. J. Leskovec and C. Faloutsos, (2006),"Sampling from large graphs," in *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 631-636.
35. T. Herlau, M. N. Schmidt, and M. Mørup, (2014),"Infinite-degree-corrected stochastic block model," *Physical Review E*, vol. 90, p. 032819.
36. P. Hagmann, L. Cammoun, X. Gigandet, R. Meuli, C. J. Honey, V. J. Wedeen, *et al.*, (2008),"Mapping the structural core of human cerebral cortex," *PLoS Biol*, vol. 6, p. e159.
37. X. Sun, "Small-world and Scale-free Properties of Urban Street Networks: A Tutorial."
38. A. Clauset, C. R. Shalizi, and M. E. Newman, (2009),"Power-law distributions in empirical data," *SIAM review*, vol. 51, pp. 661-703.
39. W. Lai, "Fitting Power Law Distributions to Data."