

Community Detection in Bipartite Networks Using HellRank Centrality Measure

Ali Khosrozadeh¹, Ali Movaghar^{2*}, Mohammad Mehdi Gilanian Sadeghi¹, and Hamidreza Mahyar³

¹ Department of Computer and Information Technology Engineering, Qazvin Branch, Islamic Azad University, Qazvin, Iran

² Department of Computer Engineering, Sharif University of Technology, Tehran, Iran

³ Department of Computer Engineering, McMaster University, Hamilton, Ontario, Canada

Received: 06 April 2023, Revised: 21 April 2023, Accepted: 11 June 2023

Paper type: Research

Abstract

Community structure is a common and important feature in many complex networks, including bipartite networks. In recent years, community detection has received attention in many fields and many methods have been proposed for this purpose, but the heavy consumption of time in some methods limits their use in large-scale networks. There are methods with lower time complexity, but they are mostly non-deterministic, which greatly reduces their applicability in the real world. The usual approach that is adopted to community detection in bipartite networks is to first construct a unipartite projection of the network and then communities detect in that projection using methods related to unipartite networks, but these projections inherently lose information. In this paper, based on the bipartite modularity measure that quantifies the strength of partitions in bipartite networks and using the HellRank centrality measure, a quick and deterministic method for community detection from bipartite networks directly and without need to projection, proposed. The proposed method is inspired by the voting process in election activities in the social society and simulates it.

Keywords: Social Networks, Bipartite Graphs, Centrality Measure, Community Detection, Voting.

* Corresponding Author's email: movaghar@sharif.edu

شناسائی انجمن در شبکه‌های دوبخشی با استفاده از معیار مرکزیت هلرنک

علی خسروزاده^۱، علی موقر^۲، محمد مهدی گیلانیان صادقی^۱، حمیدرضا ماهیار^۳
^۱ گروه مهندسی کامپیوتر و فناوری اطلاعات، واحد قزوین، دانشگاه آزاد اسلامی، قزوین، ایران
^۲ گروه مهندسی کامپیوتر، دانشگاه صنعتی شریف، تهران، ایران
^۳ گروه مهندسی کامپیوتر، دانشگاه مک‌مستر، همیلتون، انتاریو، کانادا

تاریخ دریافت: ۱۴۰۲/۰۱/۱۷ تاریخ بازبینی: ۱۴۰۲/۰۲/۰۱ تاریخ پذیرش: ۱۴۰۲/۰۳/۲۱
نوع مقاله: پژوهشی

چکیده

ساختار انجمن ویژگی مشترک و مهمی در بسیاری از شبکه‌های پیچیده از جمله شبکه‌های دوبخشی است. شناسائی انجمن‌ها در سال‌های اخیر در بسیاری زمینه‌ها مورد توجه قرار گرفته و روش‌های زیادی برای این منظور پیشنهاد شده است، اما مصرف سنگین زمان در برخی روش‌ها، استفاده از آنها را در شبکه‌های بزرگ مقیاس محدود می‌کند. روش‌هایی با پیچیدگی کمتر وجود دارند اما اکثراً غیرقطعی هستند که کاربرد آنها در دنیای واقعی را کاهش می‌دهد. رویکرد معمول اتخاذ شده برای شناسائی انجمن‌ها در شبکه‌های دوبخشی این است که ابتدا یک طرح‌ریزی یک‌بخشی از شبکه ساخته شود و سپس انجمن‌ها در آن طرح‌ریزی با استفاده از روش‌های مربوط به شبکه‌های یک‌بخشی شناسائی شوند. این طرح‌ریزی‌ها به طور ذاتی اطلاعات را از دست می‌دهند. در این مقاله بر اساس معیار ماژولاریتی دوبخشی که قدرت تقسیم‌بندی‌ها را در شبکه‌های دوبخشی محاسبه می‌کند و با استفاده از معیار مرکزیت هلرنک، روشی سریع و قطعی برای شناسائی انجمن‌ها از شبکه‌های دوبخشی به طور مستقیم و بی‌نیاز از طرح‌ریزی ارائه گردیده است. روش پیشنهادی از فرآیند رأی‌گیری در فعالیت‌های انتخاباتی در جامعه انسانی الهام گرفته و آن را شبیه‌سازی می‌کند. نتایج آزمایشات نشان می‌دهد، مقدار ماژولاریتی انجمن‌های حاصل و دقت شناسائی تعداد آنها در روش پیشنهادی بهبود یافته است.

کلیدواژگان: شبکه‌های اجتماعی، گراف‌های دوبخشی، معیار مرکزیت، شناسائی انجمن، رأی‌گیری.

۱- مقدمه

شبکه‌های اجتماعی و برای تعیین اهمیت نسبی رأس‌های شبکه که نقش مهمی در تعیین ساختار انجمن‌ها دارند استفاده می‌شوند. یکی از این معیارهای مرکزیت، معیاری دقیق برای شناسایی رأس‌های مرکزی در شبکه‌های دوبخشی به نام معیار هلرنک^۷ است که مهم ترین مزیت این معیار جهت استفاده از آن در این مقاله، عدم نیاز به تبدیل شبکه دوبخشی مورد بررسی به معادل یک‌بخشی آن است. برخی روش‌های مطرح در زمینه شناسایی انجمن‌ها دارای مشکلاتی شامل عدم قطعیت، مصرف بالای زمان و حافظه، عدم امکان اعمال در شبکه‌های بزرگ مقیاس و محدودیت تفکیک^۸ هستند. روش پیشنهادی با استفاده از مفاهیم ماژولاریتی دوبخشی و معیار مرکزیت هلرنک توانسته به طور مستقیم و بدون نیاز به تبدیل به معادل یک‌بخشی، در شبکه‌های دوبخشی به ویژه شبکه‌های دوبخشی بزرگ مقیاس اعمال گردد. نتایج حاصل از پیاده‌سازی روش پیشنهادی بر روی تعدادی شبکه مصنوعی^۹ و واقعی^{۱۰} و مقایسه آنها با دو روش شناخته شده برای شناسایی انجمن در شبکه‌های دوبخشی این امر را اثبات می‌کند.

ساختار انجمن

ساختار انجمن نه تنها اتصال‌های دانه‌درشت^{۱۱} را در شبکه‌ها آشکار می‌کند، بلکه نقش مهمی در عملکرد شبکه‌ها ایفا می‌نماید [۲]. شناسایی انجمن، امکان بررسی کمی شبکه‌های فرعی را فراهم می‌کند که ممکن است ویژگی‌های متفاوتی از ویژگی‌های کل شبکه داشته باشند. به طور غیر رسمی یک انجمن شبکه، زیرگرافی است که احتمال اتصال رأس‌های آن به یکدیگر بیشتر از رأس‌های خارج از زیرگراف است. یکی از ویژگی‌های کلیدی اکثر شبکه‌های دوبخشی، ساختار انجمن است که در آن، شبکه‌ها به گروه‌هایی از رأس‌ها و یال‌ها تقسیم می‌شوند. انجمن، ساختار ماژولار شبکه‌های زیرین^{۱۲} است که در آن یال‌هایی متراکم بین رأس‌های همان گروه و در عین حال یال‌هایی سست بین گروه‌های مختلف وجود دارد. شناسایی انجمن‌ها می‌تواند زیرشبکه‌های دانه‌درشت را از شبکه‌های زیرین شناسایی کند. از سوی دیگر، انجمن‌های مختلف دارای ویژگی‌های ساختاری متمایزی هستند، بنابراین ویژگی‌های متوسط عمومی^{۱۳} برای ارائه ویژگی‌های ساختارهای انجمن کافی نیست [۴]. انجمن همچنین راه بهتری برای درک ساختارهای ماژولار شبکه‌های

شبکه‌ها، سیستم‌های پیچیده مختلفی را در رشته‌های گوناگون توصیف می‌کنند. شبکه‌های استنباط شده از سیستم‌های پیچیده از رأس‌ها و یال‌هایی تشکیل شده‌اند که نشان دهنده موجودیت‌ها و رابطه‌ها هستند. تحقیقات زیادی بر روی شبکه‌های یک‌بخشی^۱ یا یک‌حالتی که تنها شامل یک نوع رأس هستند متمرکز شده‌اند. با این حال، شبکه‌های دنیای واقعی معمولاً شامل چندین نوع رأس هستند که ساده‌ترین حالت آنها، شبکه دوبخشی^۲ یا دوحالتی است و شامل دو نوع رأس مختلف می‌باشد. یک کلاس خاص و فراگیر از شبکه‌ها، شبکه‌های دوبخشی هستند که در آن رأس‌ها به دو کلاس مجزا تقسیم شده و یال‌های درون شبکه‌های دوبخشی فقط در بین کلاس‌های مختلف رأس‌ها رخ داده و هیچ یالی بین یک نوع از رأس‌ها وجود ندارد [۱]. شبکه‌های دوبخشی مدل‌های طبیعی بسیاری از سیستم‌های دنیای واقعی می‌باشند. شبکه‌های اجتماعی به ساختار اجتماعی بسیار مهمی در جامعه مدرن ما تبدیل شدند و ساختار انجمن^۳ یک ویژگی رایج بین این شبکه‌های اجتماعی است که نشان دهنده خوشه‌بندی رأس‌های شبکه با یال‌های متراکم در داخل گروه‌ها اما یال‌های پراکنده بین آنهاست. مفهوم انجمن راه بهتری برای درک ساختارهای ماژولار در شبکه‌های دوبخشی ارائه می‌دهد، لذا شناسایی انجمن^۴ در شبکه‌های پیچیده می‌تواند به درک ویژگی‌های ساختاری این شبکه‌ها و پیش‌بینی بیشتر ویژگی‌های کاربردی با توجه به ساختار آنها کمک کند. تا کنون روش‌های بسیار متنوعی برای شناسایی انجمن در شبکه‌های دوبخشی معرفی گردید. یک ایده ساده برای شناسایی انجمن این بود که ابتدا شبکه‌های دوبخشی به معادل یک‌بخشی خود تبدیل شده و پس از آن روش‌های معمول شناسایی انجمن در شبکه‌های یک‌بخشی بکار گرفته شود که این ایده همواره با مشکل از دست رفتن اطلاعات در این فرآیند مواجه است. بعد از ارائه مفهوم ماژولاریتی^۵ دوبخشی که در آن دو بخش مستقل از رأس‌ها به طور مستقل در ساختارهای ماژولار قرار می‌گیرند، ایده شناسایی انجمن‌ها به صورت مستقیم و بدون نیاز به تبدیل شبکه‌های دوبخشی به یک‌بخشی مطرح شد. یکی از ابزارهای مفید برای شناسایی انجمن‌ها استفاده از معیار مرکزیت^۶ است. معیارهای مرکزیت به شکل گسترده در تجزیه و تحلیل

⁸ Resolution Issue⁹ Synthetic Network¹⁰ Real Network¹¹ Coarse-Grained¹² Underlying Network¹³ Global Average¹ Unipartite² Bipartite³ Community Structure⁴ Community Detection⁵ Modularity⁶ Centrality Measure⁷ HellRank

شناسائی انجمن‌ها در شبکه‌های دوبرخی استفاده کرد بدون اینکه مجبور باشیم ابتدا یک طرح‌ریزی^۷ یک‌بخشی از شبکه دوبرخی در اختیار بسازیم. در واقع یکی از دلایل استفاده از این معیار هدف مهمی در روش پیشنهادی است که همان شناسائی انجمن‌ها در شبکه‌های دوبرخی به صورت مستقیم و بی نیاز از طرح‌ریزی یک‌بخشی معادل آنهاست.

۱-۱-۱- معیار مرکزیت هلرنک

با مد نظر قرار دادن شاخص‌های اهمیت کاربران برای شناسائی رأس‌های مرکزی، معیار مرکزیت جدیدی به نام هلرنک معرفی شد [۱۰]. معیار هلرنک، یک معیار مرکزیت دقیق جهت شناسائی رأس‌های مرکزی در شبکه‌های اجتماعی دوبرخی است که بر اساس فاصله هلینگر^۸ بین دو رأس در یک بخش شبکه دوبرخی تعریف شد [۱۱]. محاسبه این معیار می‌تواند با اجازه به هر رأس برای استفاده از اطلاعات محلی‌اش، توزیع شده باشد، پس نیازی به یک موجودیت مرکزی که دانش کاملی از ساختار توپولوژیکی شبکه داشته باشد نیست. فاصله هلینگر یک معیار f-divergence است که بیانگر شباهت ساختاری هر رأس با دیگر رأس‌های شبکه است. از این‌رو، این معیار مبتنی بر فاصله برای شناسائی رأس‌های با بیشترین نمایندگی رفتاری، دقیق است. نشان داده شده که رأس با مرکزیت هلرنک بالا، مرکزیت درجه و بینابینی به نسبت بالاتری دارد. در واقع، با وجود اهداف متفاوت در شناسائی رأس‌های مرکزی، یک همبستگی جزئی بین هلرنک و دیگر معیارهای مرسوم وجود دارد. از آنجا که معیارهای شباهت مثل هلرنک معمولاً معکوس معیارهای فاصله هستند، ابتدا معیار فاصله مناسب مثل هلینگر انتخاب و به شبکه دوبرخی اعمال می‌شود. به شکل تئوری تأثیر معیار فاصله در شبکه‌های دوبرخی آنالیز و یک ماتریس فاصله روی یک بخش شبکه تولید و امتیاز هلرنک هر رأس با توجه به این ماتریس محاسبه می‌شود. به عنوان نتیجه، رأس‌های با هلرنک بالا رأس‌های با بیشترین نمایندگی رفتاری در شبکه اجتماعی دوبرخی هستند. در معیار هلرنک، هدف ارائه معیاری مبتنی بر شباهت برای هر زوج از رأس‌ها در شبکه است، بنابراین، یک معیار فاصله خوش تعریف به عنوان معیار پایه انتخاب می‌شود.

تعریف ۱: یک فضای معیار، یک مجموعه X است که مفهوم تابع فاصله $d(x, y)$ بین هر زوج از نقاط $x, y \in X$ را دارد. یک معیار

دوبرخی ارائه می‌دهد. دو نوع ایده در رابطه با انجمن دوبرخی وجود دارد که اعضاء از یک نوع یا از نوع متفاوت در یک گروه در نظر گرفته شوند. ساختار انجمن در شبکه‌های دوبرخی نه تنها میزان اتصال رأس‌ها در یک بخش با رأس‌های بخش دیگر را نشان می‌دهد، بلکه رأس‌های مشابه را در همان بخش به گروه‌هایی خوشه‌بندی می‌کند. برای شناسائی ساختار انجمن در شبکه‌های دوبرخی، معیار مازولاریتی خاصی برای شبکه‌های دوبرخی توسط آقای باربر [۱] پیشنهاد شد و الگوریتم‌هایی مانند BRIM [۱] و LP [۵] توسعه یافتند.

معیار مرکزیت

مرکزی‌ترین رأس اغلب رأس‌هایی هستند که وزن بیشتری هم از نظر تعداد تعامل اجتماعی^۱ و هم از نظر تعداد یال با دیگر رأس‌ها دارند [۶]. در تجزیه و تحلیل شبکه‌های اجتماعی، چنین مفهوم مرکزیتی برای شناسائی کاربران تأثیرگذار استفاده می‌شود زیرا تأثیر یک کاربر، توانائی عمومی کردن محتوائی خاص در شبکه است [۷]. بدین منظور، معیارهای مرکزیت مختلفی در طول سال‌ها برای رتبه‌بندی رأس‌های شبکه بر اساس ویژگی‌های توپولوژیکی و ساختاری آنها پیشنهاد شده‌اند. این معیارها را می‌توان از چند دیدگاه با مقدار پیچیدگی محاسباتی متفاوت در نظر گرفت، از معیارهای کم‌هزینه مثل مرکزیت درجه^۲ تا معیارهای پرهزینه مثل مرکزیت بینابینی^۳ و نزدیکی^۴. این معیارهای ساختاری نیازمند داشتن دانش کاملی از توپولوژی شبکه هستند [۸]. یک مشاهده جالب این است که بسیاری از شبکه‌های اجتماعی واقعی ماهیتی دووجهی دارند که اجازه می‌دهد به عنوان یک گراف دوبرخی مدل شوند. سیستم توصیه‌کننده اجتماعی^۵ یکی از سیستم‌هایی است که می‌توان آن را به عنوان یک گراف دوبرخی مدل‌سازی کرد. در این سیستم‌ها، معیارهای مرکزیت می‌تواند تفسیر متفاوتی از معیارهای مرکزیت مرسوم مانند بینابینی، نزدیکی، درجه و رتبه‌صفحه^۶ داشته باشد [۹]. معیارهای بینابینی و نزدیکی، به عنوان رایج‌ترین شناسه رأس‌های مرکزی در شبکه‌های یک‌بخشی شناخته می‌شوند، گرچه در شبکه‌های اجتماعی دوبرخی معمولاً در شناسائی کاربران مرکزی که نماینده کاملی برای ساختار شبکه‌اند مناسب نیستند. یکی از معیارهای مناسب برای شبکه‌های اجتماعی دوبرخی که در این مقاله از آن برای شناسائی انجمن‌ها بهره بردیم معیار مرکزیت هلرنک است [۱۰]. معیار هلرنک معیاری است که می‌توان از آن در

⁵ Social Recommender System

⁶ PageRank

⁷ Projection

⁸ Hellinger

¹ Social Interaction

² Degree

³ Betweenness

⁴ Closeness

۱-۱-۳- پیش‌بینی رتبه با هلرنک

بعد از یافتن فاصله‌های هلینگر بین هر زوج رأس در هر بخش شبکه دوبخشی، ماتریس فاصله $n_1 \times n_1$ تولید می‌شود که n_1 تعداد رأس‌ها در آن بخش شبکه است. با توجه به ویژگی‌های معیار خوش‌تعریف و قابلیت نگاشت به فضای اقلیدسی، می‌توان رأس‌ها را بر اساس فاصله‌های آنها خوشه‌بندی کرد. یعنی هر زوج از رأس‌های با فاصله کمتر می‌تواند به وسیله شعاع همسایگی خاصی در یک خوشه^۴ قرار گیرد. با میانگین‌گیری معکوس درایه‌های هر سطر در ماتریس، امتیاز شباهت نهائی هلرنک برای هر رأس x در یک بخش از شبکه (با مجموعه رأس‌های V_1) به صورت زیر به دست می‌آید:

$$HellRank(x) = \frac{n_1}{\sum_{z \in V_1} d(x,z)} \quad (5)$$

فرض می‌شود $HellRank^*(x)$ مقدار نرمال شده هلرنک رأس x به صورت زیر به دست آید که علامت \cdot نشان دهنده ضرب و $\min_{z \in V_1}(HellRank(z))$ حداقل ممکن هلرنک برای آن است:

$$HellRank^*(x) = HellRank(x) \cdot \min_{z \in V_1}(HellRank(z)) \quad (6)$$

یک معیار شباهت، مقادیر کوچک را برای رأس‌های نامشابه و مقادیر بزرگ را برای رأس‌های مشابه می‌گیرد. رأس‌های یک بخش از شبکه با امتیازات شباهت بالاتر، نشان دهنده بیشترین نمایندگی رفتاری آن بخش شبکه هستند، یعنی این رأس‌ها بیشتر از دیگر رأس‌ها به آن بخش شباهت دارند. معیار هلرنک در واقع شباهت ساختاری هر رأس به دیگر رأس‌های شبکه را بیان می‌کند. بنابراین مرتب‌سازی رأس‌ها بر اساس این معیار یک پیش‌بینی رتبه بهتر برای رأس‌هاست. رأس‌های با امتیاز هلرنک بالا بیشتر شبیه دیگر رأس‌ها بوده و رأس‌های با امتیاز کمتر برای شناسائی رأس‌هایی که احتمالاً بسیار متفاوت از دیگر رأس‌ها هستند، بکار می‌روند.

ماژولاریتی

معیار ماژولاریتی آقای نیومن [۳]، گستره‌ای را نسبت به یک شبکه مدل تهی^۵ نشان می‌دهد که یال‌ها به جای بین انجمن‌ها در داخل انجمن‌ها شکل می‌گیرند. با استفاده از این ماژولاریتی، می‌توان کیفیت تخصیص رأس‌ها به انجمن‌ها را ارزیابی کرد. به حداکثر رساندن دقیق ماژولاریتی یک مشکل حل نشدنی است. برای یافتن تقسیم‌بندی‌های^۶ دارای ماژولاریتی بالا از رأس‌ها، الگوریتم‌هایی معرفی شد. نظر به وابستگی صریح ماژولاریتی به یک مدل تهی،

فاصله خوش‌تعریف d روی مجموعه X یک تابع $d: X \times X \rightarrow \mathbb{R}$ است که برای همه $x, y, z \in X$ ، سه ویژگی برقرار باشد [۱۰]:

- قطعیت مثبت^۱: $d(x, y) \geq 0$ و $d(x, y) = 0$ اگر و تنها اگر $x = y$
- تقارن^۲: $d(x, y) = d(y, x)$
- نامساوی مثلثی^۳: $d(x, y) \leq d(x, z) + d(z, y)$

تابع فاصله به صورت تفاوت بین توزیع احتمال هر زوج از رأس‌ها و بر اساس تابع f-divergence به شکل زیر تعریف می‌شود.

تعریف ۲: یک f-divergence یک تابع $D_f(P \parallel Q)$ است که تفاوت بین دو توزیع احتمال P و Q را اندازه‌گیری می‌کند. برای یک تابع محدب f با $f(1) = 0$ ، f-divergence توزیع احتمال Q از P به صورت زیر تعریف می‌شود، که در آن Ω یک فضای نمونه است که مجموعه همه خروجی‌های ممکن می‌باشد [۱۰]:

$$D_f(P \parallel Q) = \int_{\Omega} f\left(\frac{dP}{dQ}\right) dQ \quad (1)$$

برای دو توزیع احتمال گسسته $P = (p_1, \dots, p_m)$ و $Q = (q_1, \dots, q_m)$ (طول بردارها)، فاصله هلینگر برابر است با:

$$D_H(P \parallel Q) = \frac{1}{\sqrt{2}} \sqrt{\sum_{i=1}^m (\sqrt{p_i} - \sqrt{q_i})^2} \quad (2)$$

که آشکارا با قاعده اقلیدسی تفاوت جذر بردارها مرتبط است:

$$D_H(P \parallel Q) = \frac{1}{\sqrt{2}} \|\sqrt{P} - \sqrt{Q}\|_2 \quad (3)$$

۱-۱-۲- اعمال فاصله هلینگر

هدف اعمال فاصله هلینگر به یک بخش شبکه دوبخشی اندازه‌گیری شباهت رأس‌های آن بخش شبکه است. فرض می‌شود x رأسی در شبکه دوبخشی با همسایگی $N(x)$ و درجه $deg(x) = |N(x)|$ ، Δ بزرگترین درجه رأس شبکه، l_i تعداد همسایه‌های x با درجه i و بردار $L_x = (l_1, \dots, l_{\Delta})$ توزیع نرمال‌نشده l_i برای همه همسایه‌های مجاور x باشد. فاصله هلینگر بین دو رأس x و y در یک بخش شبکه دوبخشی به صورت زیر تعریف می‌شود که تابع $d(x, y)$ تفاوت بین دو توزیع احتمال L_x و L_y را نشان می‌دهد:

$$d(x, y) = \sqrt{2} D_H(L_x \parallel L_y) \quad (4)$$

⁴ Cluster

⁵ Null Model

⁶ Partition

¹ Positive Definiteness

² Symmetry

³ Triangle Inequality

برابر صفر است. ماتریس B دربرگیرنده اطلاعات یال‌ها در یک شبکه دوبخشی است. از $d_i = \sum_j^q B_{ij}$ و $g_j = \sum_i^p B_{ij}$ به ترتیب برای نشان دادن درجه رأس‌های $i \in X$ و $j \in Y$ استفاده می‌شود. بدیهی است که $\sum_i d_i = \sum_j g_j$. ماژولاریتی دوبخشی یک تقسیم^۱ به صورت زیر تعریف می‌شود:

$$Q = \frac{1}{m} \sum_{i=1}^p \sum_{j=1}^q (B_{ij} - \frac{d_i g_j}{m}) \delta(r_i, s_j) \quad (8)$$

که d_i درجه رأس قرمز i -ام، g_j درجه رأس آبی j -ام و r_i و s_j انجمن‌هائی هستند که i و j به آنها اختصاص دارند. مقدار m تعداد یال‌ها و تابع $\delta = (u, v)$ اگر $u = v$ برابر با ۱ و وگرنه برابر است با صفر. این تعریف برای شبکه دوبخشی منطقی است چون مانع از یال بین رأس‌های هم‌رنگ می‌شود. افزایش ماژولاریتی^۲ دوبخشی، وقتی یک رأس ایزوله i به انجمن c منتقل می‌شود، طبق رابطه ۹ (بازنویسی رابطه ۸) محاسبه می‌گردد. افزایش حاصل از انتقال یک رأس آبی می‌تواند به همین شکل به دست آید.

$$Q = \frac{1}{m} \sum_{c \in C} (\Psi_c^{in} - \frac{\Psi_c^r \Psi_c^b}{m}) \quad (9)$$

که مقادیر Ψ_c^{in} ، Ψ_c^r و Ψ_c^b به ترتیب نشان دهنده تعداد کل یال‌ها و جمع درجه‌های همه رأس‌های قرمز و آبی درون انجمن c هستند. اگر یک رأس ایزوله قرمز i و یک انجمن c به صورت یک انجمن تنها در نظر گرفته شوند، ماژولاریتی ترکیبی برابر است با:

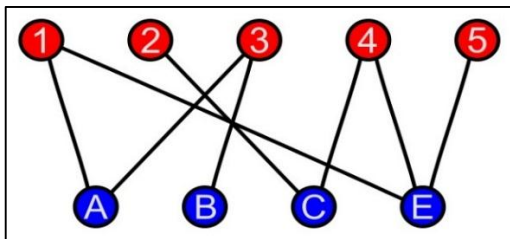
$$Q_{comb} = \frac{\Psi_c^{in} + k_{i, in}^b}{m} - \frac{(\Psi_c^r + k_i) \Psi_c^b}{m^2} \quad (10)$$

که $k_{i, in}^b$ تعداد یال‌های بین رأس قرمز i و همه رأس‌های آبی انجمن c است. ماژولاریتی انجمن جزئی متشکل از رأس قرمز تنها i برابر است با صفر، چون هر دو مقدار Ψ_c^{in} و Ψ_c^b صفر هستند. لذا، ماژولاریتی وقتی c و i جدا از هم هستند برابر است با:

$$Q_{sep} = \frac{\Psi_c^{in}}{m} - \frac{\Psi_c^r \Psi_c^b}{m^2} \quad (11)$$

از این رو ΔQ به صورت زیر به دست می‌آید:

$$\Delta Q = Q_{comb} - Q_{sep} = \frac{k_{i, in}^b}{m} - \frac{k_i \Psi_c^b}{m^2} \quad (12)$$



شکل ۱. نمونه‌ای از یک گراف دوبخشی با دو مجموعه رأس مجزا

واضح است که انتخاب خاص مدل تهی تأثیر عمیقی بر ماژولاریتی دارد. به نظر می‌رسد، تا کتون تنها یک مدل تهی به طور مفصل مورد بررسی قرار گرفت، یعنی شبکه‌هائی با یال‌هائی که به طور تصادفی تخصیص داده شده‌اند به طوری که درجه‌های مورد انتظار رأس‌های شبکه مدل برابر با درجه‌های واقعی رأس‌های شبکه واقعی متناظر است [۱۲]. رویکرد معمول برای شناسائی انجمن در شبکه‌های دوبخشی، این است که ابتدا یک طرح‌ریزی یک‌بخشی از یک بخش شبکه ساخته و سپس انجمن‌ها در آن طرح‌ریزی با استفاده از روش‌های موجود برای شبکه‌های یک‌بخشی شناسائی شوند. این طرح‌ریزی یک‌بخشی می‌تواند روشن‌کننده باشد، اما به طور ذاتی اطلاعات را از دست می‌دهد. نشان داده شده که تجزیه و تحلیل یک طرح‌ریزی بدون وزن و یک‌بخشی می‌تواند نتایج غیرقابل اعتماد یا نادرستی ارائه دهد [۴].

۱-۱-۴- ماژولاریتی در شبکه‌های دوبخشی

برای شناسائی ساختار انجمن در شبکه‌های دوبخشی، باربر [۱] تعریف خاصی از ماژولاریتی در این شبکه‌ها را برای ارزیابی همزمان درستی تقسیم‌بندی هر دو نوع رأس‌های داخل انجمن‌ها پیشنهاد داد. انجمن‌ها بوسیله بهینه‌سازی این تعریف از ماژولاریتی شناسائی می‌شوند. محدودیت‌های خاصی باید در تعریف ماژولاریتی دوبخشی در نظر گرفته شود، به عنوان مثال رأس‌های از یک نوع اجازه ندارند به هم متصل شوند. باربر مدل تهی را بر اساس ماتریس مجاورت دوبخشی یک شبکه تعریف کرد. گراف دوبخشی H گرافی است که در آن مجموعه رأس‌ها به دو زیرمجموعه مجزای X و Y تقسیم شده به طوری که هر یال تنها اجازه دارد یک رأس در X را به رأسی در Y متصل کند.

رأس‌ها با دو رنگ قرمز و آبی که نشان دهنده مجموعه‌هائی است که متعلق به آن هستند از هم جدا می‌شوند. اگر فرض شود گراف H از p و q رأس قرمز و آبی تشکیل شده، ماتریس مجاورت H با ابعاد $(p+q) \times (p+q)$ به شکل زیر نمایش داده می‌شود:

$$A_{(p+q) \times (p+q)} = \begin{bmatrix} O_{p \times p} & B_{p \times q} \\ B_{q \times p}^T & O_{q \times q} \end{bmatrix} \quad (7)$$

که O یک بلوک ماتریسی با درایه‌های همه صفر را نشان می‌دهد و بلوک ماتریسی B به عنوان ماتریس مجاورت دوبخشی، نشان دهنده یال‌های بین رأس‌های قرمز و آبی است. این یک ماتریس $p \times q$ است که به شکل $B := (B_{ij})$ تعریف می‌شود و B_{ij} برابر با ۱ است اگر یک یال، رأس‌های $i \in X$ و $j \in Y$ را به هم متصل کند و گرنه

² Modularity Gain

¹ Partition

ضریب خوشه‌بندی در شبکه‌های دوبخشی

این معیار، گرایش رأس‌ها برای تشکیل خوشه را نشان می‌دهد. در بسیاری از شبکه‌های دنیای واقعی، رأس‌ها متمایل‌اند در انجمن‌های مترکم خوشه‌بندی شوند. معیارهای زیادی برای آزمودن این گرایش پیشنهاد شده‌اند. آقای اوپسال، فاکتور $C^*(i)$ را به عنوان ضریب خوشه‌بندی رأس i ، به صورت زیر تعریف کرد [۱۳]:

$$C^*(i) = \frac{\text{closed 4-paths centered on node } i}{\text{4-paths centered on node } i} = \frac{\tau_{i,\Delta}^*}{\tau_i^*} \quad (13)$$

که τ_i^* تعداد ۴-مسیرهای^۱ متمرکز بر رأس کانونی i و $\tau_{i,\Delta}^*$ زیرمجموعه‌ای از این مقدار است به طوری که ابتدا و انتهای مسیر، رأسی را به اشتراک دارند که بخشی از ۴-مسیر نیست و بخشی از حداقل یک ۶-دور^۲ است، به عبارتی، تعدادی از این ۴-مسیرهاست که وقتی به صورت بخشی از حداقل یک ۶-دور یعنی حلقه‌ای از شش یال با پنج رأس باشند، بسته می‌شوند. این ضریب دو ویژگی دارد. اول، مقدار آن اگر صورت و مخرج کسر اعداد مثبت باشند و صورت کسر زیرمجموعه‌ای از مخرج کسر باشد، بین صفر تا ۱ متغیر است. دوم، در یک شبکه کاملاً متصل این مقدار برابر با ۱ است زیرا همه ۴-مسیرها بسته هستند.

پیشینه تحقیق

ساختار انجمن ویژگی ساختاری مهمی در بسیاری از شبکه‌های پیچیده است که در آن انجمن‌ها همیشه با ماژول‌های کاربردی در سیستم‌های دنیای واقعی مرتبط هستند. وجود ساختارهای انجمن در شبکه‌ها می‌تواند تأثیر قابل توجهی بر فرآیندهای پویایی مانند پراکندگی اطلاعات^۳ [۱۴] و همگام‌سازی^۴ [۱۵] داشته باشد. بنابراین، شناسایی انجمن توجه زیادی به خود جلب و روش‌های شناسایی زیادی پیشنهاد شد، شامل روش‌های شناسایی انجمن‌ها با مفروضات مختلف مثل تجزیه ماتریس غیرمنفی^۵ [۱۶]، انتشار برچسب^۶ [۱۷]، گسترش مجموعه‌های بذر^۷ [۱۸]، روش تکاملی^۸ [۱۹]، رویکردهای نظریه‌بازی^۹ [۲۰]، گراف‌های خطی^{۱۰} [۲۱] و بهینه‌سازی ماژولاریتی^{۱۱} [۲۲]. یک ایده ساده، یک روش طرح‌ریزی است که در آن شبکه‌های دوبخشی به شبکه‌های یک‌بخشی تبدیل

شده و پس از آن می‌توان روش‌های شناسایی انجمن در شبکه‌های یک‌بخشی را به طور طبیعی به کار برد [۲۳]. با این حال، استدلال شده که روش‌های طرح‌ریزی ممکن است به مشکل اطلاعات ناقص منجر شوند زیرا تنها یک نوع از رأس‌ها اعمال شده و نوع دیگر آنها پس از طرح‌ریزی از بین می‌رود [۲۴]. بنابراین روش‌های مختلفی برای حفظ دو نوع رأس پس از تقسیم‌بندی انجمن‌ها معرفی گردید. برابر ابتدا یک ماژولاریتی دوبخشی را پیشنهاد داد [۱] که از ماژولاریتی یک‌بخشی [۲۵] گسترش می‌یابد و سپس الگوریتم BRIM را برای القای دوبخش مستقل از رأس‌ها به ساختارهای ماژولار توسعه داد. با این حال، ماژولاریتی شبکه‌های دوبخشی دارای محدودیت‌هایی در مسئله تفکیک [۲۶] است زیرا نمی‌توان انجمن‌های کوچک را با امتیاز ماژولاریتی بالا به دقت شناسایی کرد. برای شناسایی انجمن دو-دسته^{۱۲} بر اساس توسعه الگوریتم شناسایی انجمن k -دسته روشی ارائه شد که تمام مزایای الگوریتم k -دسته را دارد و سطحی از انعطاف‌پذیری را با ترکیب آستانه‌های دسته مستقل فراهم می‌کند [۲۷]. روش‌های دیگری با مفروضات متمایز نیز پیشنهاد شد، مانند بردارهای ویژه ماتریس‌ها^{۱۳} [۱۲]، بهینه‌سازی ماژولاریتی [۴]، ضریب خوشه‌بندی [۲۸]، درجه صمیمیت^{۱۴} [۲۹]، مدل بلوک تصادفی^{۱۵} [۳۰]، ماژولاریتی مبتنی بر چگالی^{۱۶} [۳۱] و صمیمیت نامتقارن^{۱۷} [۲۴]. یک روش یکپارچه شناسایی انجمن مبتنی بر احتمال تشابه رأس^{۱۸} [۳۲] برای مقابله با شبکه‌های یک‌بخشی و دوبخشی با هم پیشنهاد گردید [۳۳]. همانطور که اندازه شبکه به سرعت افزایش می‌یابد، ایجاد تعادل بین دقت و عملکرد در عمل مهم بوده است. روش‌های کنونی، جز در شبکه‌های یک‌بخشی، به شناسایی انجمن در شبکه‌های دوبخشی بزرگ با صدها هزار رأس و یال توجه چندانی نکرده‌اند. روش‌های دقیق و کارآمدی برای شناسایی انجمن در شبکه‌های یک‌بخشی بزرگ با استفاده از شباهت رأس پیشنهاد شد [۳۲]. با این حال، روش‌های بیشتری برای مقابله موثر با شبکه‌های دوبخشی بزرگ مورد نیاز است. یکی از دلایل احتمالی این وضع، حداقل پیچیدگی زمانی از درجه دوم است که باعث می‌شود ساعت‌ها طول بکشد تا با شبکه‌های دوبخشی بزرگ سروکار داشته باشیم و مشکل محدودیت تفکیک نیز وجود دارد که منجر به شناسایی نادرست انجمن‌های

¹⁰ Line Graph

¹¹ Modularity Optimization

¹² Bi-Clique

¹³ Eigen-Vector of Matrix

¹⁴ Intimate Degree

¹⁵ Stochastic Block Model

¹⁶ Density-Based Modularity

¹⁷ Asymmetric Intimacy

¹⁸ Vertex Similarity Probability

¹ 4-Path

² 6-Cycle

³ Information Dispersion

⁴ Synchronization

⁵ Non-Negative Matrix Decomposition

⁶ Label Propagation

⁷ Expansion from Seed Set

⁸ Evolutionary Method

⁹ Game Theoretic Approache

کوچک می‌شود [۲۶].

Single LP تشکیل شده است. الگوریتم BiMLPA بر مسئله ناپایدار LPA و محدودیت‌های رویکردهای قبلی غلبه می‌کند و چندین ویژگی مطلوب مانند سرعت و پایداری را داراست. نتایج تجربی نشان می‌دهد که BiMLPA از رویکردهای قبلی مانند LPA+, biSBM و ComSim بهتر عمل می‌کند.

اکثر رویکردهای شناسائی انجمن پیشنهادی برای شبکه‌های دویخشی از محدودیت‌هایی رنج می‌برند، مانند شناسائی انجمن‌ها در حضور بسیاری رأس‌های غیرمتمايز با یال‌های غیرمعمول که ساختارهای انجمن را پنهان می‌کنند، از دست دادن اطلاعات توپولوژیکی مرتبط به دلیل تبدیل شبکه دویخشی به گراف‌های ساده استاندارد و تعیین دستی چندین پارامتر ورودی از جمله تعداد انجمن‌هایی که باید شناسائی شوند. برای کاهش این مشکلات، در مرجع [۳۸] یک الگوریتم شناسائی انجمن بدون پارامتر در شبکه‌های دویخشی به نام BiNeTClus پیشنهاد شد که در دو فاز عمل می‌کند. فاز اول بر شناسائی یک گروه‌بندی اولیه از رأس‌ها از طریق یک مدل داده تراکنشی که قادر به مقابله با وضعیتی است که شبکه‌هایی با بسیاری از یال‌های غیر معمول را در بر می‌گیرد، تمرکز دارد. هدف فاز دوم، پالایش نتایج خوشه‌بندی فاز اول از طریق یک استراتژی بهینه‌سازی ماژولاریتی دویخشی برای شناسائی ساختارهای انجمن نهائی است. در شبکه‌های دویخشی، ساختارهای انجمن محدود به ناسازگاری‌اند، زیرا رأس‌های از یک نوع بر اساس الگوهای مشترک اتصال با رأس‌های نوع دیگر گروه‌بندی شده و این امر باعث می‌شود که مدل بلوک تصادفی که مدل تولیدی بسیار انعطاف‌پذیری برای شبکه‌های با ساختار بلوک است، انتخابی شهودی برای شناسائی انجمن دویخشی باشد. با این حال، فرمول‌های معمولی مدل بلوک تصادفی از ساختار ویژه شبکه‌های دویخشی استفاده نمی‌کنند. در مرجع [۳۹] یک فرمول ناپارامتریکی بیزین^۵ از مدل بلوک تصادفی و یک الگوریتم متناظر برای شناسائی موثر انجمن‌ها در شبکه‌های دویخشی معرفی شد. BiSBM، نتایج شناسائی انجمن را نسبت به SBM‌های عمومی هنگامی که داده‌ها نویز دارند و نیز محدودیت تفکیک مدل را بهبود داده و درک ما از چشم انداز بهینه-سازی پیچیده مرتبط با وظایف شناسائی انجمن را گسترش می‌دهد. یک مقایسه مستقیم بین توزیع‌های قبلی در BiSBM و یک SBM سلسله‌مراتبی با وضوح بالا، یک ساختار غیرشهودی از مشکلات شناسائی انجمن را نشان می‌دهد که توسط شبکه‌های کوچک‌تر و پراکنده‌تر پر شده و در آن مدل‌های غیر سلسله‌مراتبی بهتر از

الگوریتم BiAttractor [۳۴]، روشی برای شناسائی انجمن‌های دوحالتی در شبکه‌های دویخشی بزرگ با استفاده از دینامیک فاصله^۱ است که در آن از تعامل‌های جامعه انسانی الهام گرفته شده، به طوری که تعامل‌های بیشتر در همان انجمن و تعامل‌های کمتر بین انجمن‌های متفاوت وجود دارد. این روش در شبکه‌های پراکنده پیچیدگی $O(|E|)$ دارد و تقسیم‌بندی دقیقی از انجمن‌ها به دست می‌آورد. یک تقسیم‌بندی خوب مقادیری بهینه یا نزدیک به بهینه برخی معیارها را تولید و سازماندهی سلسله‌مراتبی ساختار انجمن را در وضوح‌های مختلف نشان می‌دهد. تقسیم‌بندی گراف و الگوریتم‌های خوشه‌بندی سلسله‌مراتبی^۲ مثل الگوریتم تجمعی^۳، بطور سنتی برای حل مشکل شناسائی انجمن پیاده‌سازی می‌شوند. از وقتی نیومن تعریف ماژولاریتی را ارائه کرد [۳]، به سرعت به معیاری محبوب که کیفیت تقسیم‌بندی را اندازه می‌گیرد تبدیل شد. ماژولاریتی نسبت یال‌های تشکیل شده در انجمن‌ها به یال‌های تشکیل شده بین انجمن‌هاست. نیومن یک مدل تهی تعریف کرد که درجه رأس‌ها را حفظ تا تصادفی بودن یال‌های تشکیل شده بین رأس‌ها را اندازه‌گیری کند. از آنجا که ماژولاریتی بالا نشان از تقسیم‌بندی خوب است، حداکثرسانی این معیار با تکنیک‌های بهینه‌سازی موضوعی محبوب است. با این حال، یافتن راه حل بهینه دقیق برای این مشکل، غیرقابل حل است، لذا الگوریتم‌های تقریبی زیادی توسعه یافتند. رویکرد حرصانه‌ای به نام Louvain، یکی از کارآمدترین این الگوریتم‌ها با عملکردی عالی است [۲۲]. الگوریتم BiLouvain نیز روش دو مرحله‌ای شناسائی ساختار انجمن در شبکه‌های دویخشی است [۳۵] که در آن الگوریتم Louvain به شبکه‌های دویخشی گسترش داده شد. این الگوریتم در شبکه‌های دویخشی یک ساختار شبکه متعادل با تعداد مساوی از دو نوع رأس تولید و سپس برای شبکه متعادل حاصل از الگوریتم اول، از یک روش خوشه‌بندی تجمعی برای خوشه‌بندی بیشتر استفاده می‌کند. در شبکه‌های دویخشی، دو راه برای شناسائی انجمن‌ها وجود دارد، انجمن‌های متناظر^۴ یک به یک و انجمن‌های متناظر چند به چند. دومی به طور طبیعی ساختارهای خوشه‌ای را در شبکه دویخشی نشان می‌دهد. با این حال، روش‌های کمی، شناسائی انجمن‌های متناظر چند به چند را هدف قرار داده‌اند. در مرجع [۳۷] یک الگوریتم انتشار چند برچسبی به نام BiMLPA برای این منظور پیشنهاد شد که از دو نوع انتشار برچسب Multi-LP و Multi-Multi-

⁴ Correspondence

⁵ Bayesian

¹ Distance Dynamic

² Hierarchical Clustering

³ Agglomerative Algorithm

فرمول‌بندی کرد. این یعنی روش ماژولاریتی و روش پارتیشن‌بندی گراف اساساً معادل هستند. در اینجا، ماژولاریتی دوبخشی بابر تحلیل و مشخص شد که حداکثرسانی ماژولاریتی دوبخشی بابر را می‌توان با رفع محدودیت‌های گسستگی به عنوان یک مسئله طیفی دوباره فرمول‌بندی کرد. همچنین، یک الگوریتم طیفی ماژولاریتی برای شناسایی ساختارهای انجمن در شبکه‌های دوبخشی طراحی شد و نتایج نشان داد این الگوریتم طیفی بهتر از الگوریتم‌های حداکثرسانی محلی ماژولاریتی مثل BRIM و BLP عمل می‌کند. به‌طور کلی، الگوریتم بهینه‌سازی عمومی بهتر از الگوریتم بهینه‌سازی محلی است. با این حال، بیشتر الگوریتم‌های فعلی حداکثرسانی ماژولاریتی دوبخشی به صورت محلی اختصاص داده شده‌اند و تعداد کمی، خوشه‌بندی طیفی را مطالعه می‌کنند. شناسایی انجمن در شبکه‌های دوبخشی زمانی (در حال تکامل) چالش برانگیز است زیرا می‌توان آن را با بر روی شبکه دوبخشی زمانی انجام داد یا روی شبکه‌های مختلف طرح‌ریزی شده و از طریق الگوریتم‌های شناسایی انجمن متنوع. سه روش پیشنهادی در مرجع [۴۳]، ساختارهای انجمن مشهود را شناسایی می‌کنند. تحلیل شبکه‌های ارزیابی سه روش، شباهت و تفاوت بین آنها در شناسایی زوج رأس‌های مشترک یا گروه‌هایی از رأس‌ها که اغلب متعلق به یک انجمن هستند را آشکار می‌کند. دو روشی مبتنی بر شبکه طرح‌ریزی شده یکسان، ساختارهای انجمن منسجم را شناسایی کرده، در حالی که روش مبتنی بر شبکه دوبخشی اولیه زمانی این دیدگاه از ساختار انجمن را تکمیل می‌کند. در واقع دو روشی که شبکه را بر اساس طرح‌ریزی زمانی یکسان تقسیم کردند به ترتیب با استفاده از Louvain و Louvain تثبیت شده، ساختارهای انجمن منسجم را شناسایی کرده، در حالی که روش سوم، بر اساس شبکه دوبخشی زمانی اولیه، چشم‌انداز مکملی از ساختار انجمن ارائه می‌دهد.

۲- روش پیشنهادی

استخراج سریع، قطعی^{۱۱} و با کیفیت ساختار انجمن از شبکه‌ها مشکلی چالشی است. در این مقاله، یک روش شناسایی انجمن پیشنهاد شده که روشی سریع و قطعی است و می‌تواند ساختارهای انجمن با کیفیت و قطعی را آشکار کند. روش پیشنهادی از فرآیند رأی‌گیری در فعالیت‌های انتخاباتی یک جامعه انسانی الهام گرفته،

همتای انعطاف‌پذیرتر خود عمل می‌کنند. در مرجع [۴۰]، الگوریتم کلونی زنبور مصنوعی^۱ بهبودیافته IABC-BN پیشنهاد شد که برای شناسایی انجمن در شبکه دوبخشی استفاده می‌شود. در مرحله اول، یک فرآیند اولیه‌سازی جمعیت^۲ جدید از روش کلونی زنبور مصنوعی برای شناسایی خوشه در گراف دوبخشی پیشنهاد شد. در مرحله دوم یعنی مرحله جستجوی زنبور شاغل^۳، یک معادله جستجوی ترکیبی جدید پیشنهاد گردید. در مرحله سوم یعنی مرحله زنبورهای تماشاگر^۴، معادله جستجوی ترکیبی جدید دیگری پیشنهاد شد. در مرحله چهارم یعنی مرحله زنبور پیشاهنگ^۵، یک آستانه احتمال برای افزایش توانایی اکتشاف و بهبود تنوع جمعیتی الگوریتم معرفی گردید. طبق دانش نویسنده، IABC-BN اولین روش کلونی زنبور مصنوعی برای شناسایی خوشه در گراف‌های دوبخشی است.

شناسایی ساختار انجمن به‌ویژه در تحلیل شبکه‌های بیولوژیکی بزرگ مقیاس با ماهیت دوبخشی، مانند آنهایی که نشان‌دهنده فعل و انفعالات نظارتی^۶، دارو-ژن^۷، یا ارتباط ژن-بیماری^۸ هستند، مفید است. انجمن‌های شناسایی شده در شبکه‌های بیولوژیکی برای فرآیندهای بیولوژیکی خاص غنی می‌شوند، بنابراین به فرد اجازه می‌دهند داروها، مولکول‌های تنظیم‌کننده یا بیماری‌ها را به چنین فرآیندهایی اختصاص دهد. در مرجع [۴۱] یک مبنای نظری از روش‌های مختلفی که می‌توانند برای شناسایی انجمن در شبکه‌های بیولوژیکی دوبخشی استفاده شوند، ارائه شد. چندین استراتژی شناسایی انجمن‌ها و نقاط قوت و ضعف آنها در زمینه کاربردهای شبکه‌های دوبخشی ژنومی^۹ مورد بررسی قرار گرفته و سپس آنها را در یک شبکه تعامل دارو-ژن گسترده ژنومی به کار می‌برند. امتیاز Murata+ همه‌کاره است و انجمن‌های شناسایی شده توسط این روش به ساختار دوبخشی شبکه احترام می‌گذارند. با این حال، تنها روشی که آن را پیاده‌سازی می‌کند BiLouvain بوده که اجرای آن در شبکه‌های گسترده ژنومی بسیار کند است. بنابراین روشی که از یک بهینه‌ساز طیفی مانند BRIM برای به حداکثر رساندن امتیاز Murata+ استفاده می‌کند، بسیار مفید خواهد بود. دو روش پرکاربرد شناسایی ساختار انجمن در شبکه‌های دوبخشی، روش ماژولاریتی و روش پارتیشن‌بندی گراف است. نتایج تحلیلی مرجع [۴۲] نشان می‌دهد که مسئله حداکثرسانی ماژولاریتی را می‌توان پس از رفع محدودیت‌های گسستگی^{۱۰}، به عنوان یک مسئله طیفی دوباره

⁷ Drug-Gene

⁸ Gene-Disease

⁹ Genomic

¹⁰ Discreteness Constraints

¹¹ Deterministic

¹ Artificial Bee Colony

² Population Initialization

³ Employed Bee

⁴ Onlooker Bee

⁵ Scout Bee

⁶ Regulatory Interactions

و یال $e \in E$ انواع مختلف رأس‌ها را به هم متصل می‌کند. هیچ یالی بین رأس‌های از همان مجموعه U یا I وجود ندارد. روش پیشنهادی، انجمن‌ها را با شبیه‌سازی فرآیند رأی‌گیری در انتخاباتی که اجازه نامزدی آزادانه را می‌دهد شناسائی می‌کند. قوانین رأی‌گیری برای رویه شناسائی اهمیت زیادی دارد. رأس $u \in U$ طبق قوانین زیر رأی می‌دهد:

۱. اگر رأس u به عنوان کاندیدا نامزد شود یا رأس u بین همسایه‌های مرتبه دوم خود بزرگترین درجه را داشته باشد، آنگاه به خودش رأی می‌دهد.
۲. وگرنه، رأسی از همسایه‌های مرتبه دوم رأس u که درجه آن بزرگتر از درجه رأس u باشد انتخاب و به عنوان رأس v در نظر گرفته می‌شود. اگر بیشتر از یک رأس از این رأس‌ها وجود داشته باشد، آن رأسی که بیشترین شباهت را با رأس u دارد انتخاب و به عنوان رأس v در نظر گرفته می‌شود. شباهت بین رأس‌های u و v به صورت $sim(u, v)$ تعریف می‌شود. اگر $sim(u, v) = 0$ باشد، رأس u خود را به عنوان کاندیدا نامزد کرده و به خودش رأی می‌دهد.
۳. وگرنه، اگر رأس v به رأس‌های دیگر رأی نداده باشد، رأس u رأی v را به عنوان کاندیدا نامزد کرده و به آن رأی می‌دهد.
۴. اگر رأس v به رأس دیگری مثل رأس w رأی داده باشد، این بدان معناست که رأس v از حق خود برای نامزد شدن به عنوان یک کاندیدا دست برداشته و در نتیجه رأس u به رأس w رأی می‌دهد.

طبق این قوانین رأی‌گیری، ترتیب رأی‌گیری رأس‌ها می‌تواند بر ترتیب رأس‌های معرفی شده به عنوان نامزد تأثیر داشته باشد، به طوری که ترتیب‌های مختلف رأی‌گیری ممکن است منجر به نتایج متفاوتی شود.

مراحل کار

در گام اول، ضریب خوشه‌بندی برای هر رأس مجموعه U در شبکه دوبخشی محاسبه شده و رأس‌ها به ترتیب صعودی ضرایب خوشه‌بندی آنها رأی می‌دهند. برای هر رأس $x \in U$ ، ضریب خوشه‌بندی به صورت زیر محاسبه می‌گردد:

$$CC(x) = \frac{\text{closed 4-paths centered on node } x}{\text{4-paths centered on node } x} \quad (14)$$

برای رأس x ، هرچه تعداد یال‌های بین همسایه‌های آن بیشتر باشد، ضریب خوشه‌بندی آن بزرگتر است. اگر زیرگراف متشکل از همسایه‌های مرتبه دوم رأس x ، گرافی کامل باشد، $CC(x)$ به

طوری که هر رأس بر اساس قوانین رأی‌گیری، به نامزدهای معرفی شده رأی داده و گروه‌های رأس به هم متصل می‌توانند به سرعت در مورد نامزدهای خود به اجماع برسند. در پایان رویه، نامزدها و رأی‌دهندگان، گروهی از خوشه‌ها را تشکیل و سپس این خوشه‌ها به عنوان انجمن‌های اولیه در نظر گرفته شده و برخی از آنها برای به دست آوردن ساختار انجمن نهائی، در قالب انجمن‌های بزرگتر با کارائی بالا تجمیع می‌شوند. تحلیل ساختارهای انجمن استخراجی از بسیاری شبکه‌ها نشان می‌دهد هر رأس و بیشتر همسایه‌های آن همیشه به یک انجمن تعلق دارند و آن رأس و همسایه‌های در یک انجمن یک خوشه کوچک را تشکیل و هر انجمن از چندین خوشه کوچک تشکیل شده است. در هر خوشه همیشه رأسی وجود دارد که درجه به نسبت بزرگتری از دیگران دارد و خوشه گروهی از رأس‌هاست که با آن رأس با درجه بزرگتر مرتبطند. این پدیده مشابه فرآیند رأی‌گیری در جامعه انسانی است. در شبکه‌های اجتماعی، همیشه افرادی با نفوذ در حوزه محلی وجود دارند که اگر در انتخاباتی با امکان نامزدی آزادانه شرکت کنند، هر رأی دهنده به فردی که نفوذ بیشتری در اطرافش دارد رأی می‌دهد. این روش منجر به ایجاد خوشه‌های کوچک زیادی شده که یک گروه رأس است و رأس تأثیرگذار را با دیگران احاطه می‌کند. بر اساس این واقعیت‌ها، روشی جهت شبیه‌سازی فرآیند رأی‌گیری برای شناسائی ساختارهای انجمن پیشنهاد شد که در آن اصولی به عنوان قوانین رأی‌گیری فرض شده، از درجه برای انعکاس تأثیر هر رأس و از شباهت بین رأس‌ها برای نشان دادن نزدیکی آنها استفاده شد. هر رأس به یکی از همسایه‌های خود رأی می‌دهد که درجه آن بزرگتر است. اگر بیش از یک رأس با درجه بزرگتر وجود دارد که تأثیر یکسانی بین همسایه‌ها داشته باشد، رأس رأی دهنده به رأسی رأی می‌دهد که بیشترین شباهت را با آن دارد. بررسی شباهت بین رأس‌ها با معیار مرکزیت هلرنک صورت می‌گیرد. به این ترتیب، هر رأس به سرعت رأی داده و در پایان فرآیند رأی‌گیری، خوشه‌های شامل رأس کوچک زیادی به دست می‌آید. طبق این تحلیل، هر انجمن از چند خوشه تشکیل شده و ما خوشه‌ها را به عنوان انجمن‌های اولیه در نظر گرفته و برخی از آنها را ادغام می‌کنیم تا در نهایت ساختار انجمن نهائی را بسازیم. روند کلی مراحل مختلف روش پیشنهادی در شکل ۲ نشان داده شده است.

قوانین رأی‌گیری

گراف مورد استفاده در روش پیشنهادی گراف دوبخشی، بدون جهت و بدون وزن است و می‌توان آن را به صورت $G = (U, I, E)$ نشان داد که در آن U و I دو مجموعه از انواع مختلف رأس را نشان داده

افزایش ماژولاریتی در هر تکرار منجر گردد بکار می‌رود. مقدار ماژولاریتی ساختار حاصل از خوشه‌های اولیه طبق رابطه ۱۶ و مقدار افزایش ماژولاریتی حاصل از هر ادغام جهت تصمیم‌گیری برای انجام آن طبق رابطه ۱۷ محاسبه می‌گردد:

$$Q = \frac{1}{m} \sum_{i=1}^{|CS|} (\Psi_{C_i} - \frac{\Psi_{C_i}^U \Psi_{C_i}^L}{m}) \quad (16)$$

$$\Delta Q = \sum_{User \in C_j \& Item \in C_i} \left(\frac{k_{C_i,p}^L}{m} - \frac{k_p \Psi_{C_i}^L}{m^2} \right) \quad (17)$$

در رابطه ۱۶، m تعداد یال‌های گراف، Ψ_{C_i} تعداد یال‌های داخل انجمن C_i ، $\Psi_{C_i}^U$ جمع درجه رأس‌های نوع اول و $\Psi_{C_i}^L$ جمع درجه رأس‌های نوع دوم داخل انجمن C_i است. در رابطه ۱۷، افزایش ماژولاریتی حاصل از افزودن انجمن C_j به انجمن C_i بدست می‌آید. در اینجا، m تعداد یال‌های گراف، k_p درجه رأس p -ام از نوع اول در انجمن C_j ، $\Psi_{C_i}^L$ جمع درجه رأس‌های نوع دوم داخل انجمن C_i و $k_{C_i,p}^L$ تعداد یال‌های بین رأس p -ام از نوع اول در انجمن C_j با همه رأس‌های نوع دوم داخل انجمن C_i است. برای دستیابی به کیفیت بالاتر، فقط جفت خوشه‌هایی در نظر گرفته می‌شوند که شباهت آنها بزرگتر از یک آستانه معین باشد. شباهت بین یک جفت انجمن اولیه C_i و C_j طبق رابطه ۱۸ محاسبه می‌گردد. اگر این شباهت بزرگتر از آستانه تعریفی باشد، افزایش ماژولاریتی حاصل از این ادغام محاسبه می‌گردد. در هر تکرار انجمن‌هایی با هم ادغام می‌شوند که ادغام آنها بالاترین افزایش ماژولاریتی را نتیجه دهد. در پایان حلقه تکرار، ساختاری به عنوان ساختار نهایی به خروجی می‌رود که در بین همه تکرارها بالاترین ماژولاریتی حاصل از کل ساختار را به دست آورد.

$$sim(C_i, C_j) = \sum_{x \in U_{C_i}} \sum_{y \in U_{C_j}} \frac{1}{hellinger\ distance(x,y)} \quad (18)$$

روند ادغام از وضعیتی شروع می‌شود که هر خوشه کوچک به عنوان یک انجمن اولیه در نظر گرفته شود، نه از وضعیتی که هر رأس یک انجمن است، یعنی زمان ادغام بسیار کمتر است. علاوه بر این، چون ادغام انجمن‌های غیرمشابه ممکن است کیفیت ساختار انجمن حاصل را تضعیف کند، فقط ادغام انجمن‌هایی که شباهت آنها از یک آستانه معین بیشتر است مدنظر قرار می‌گیرد. همچنین، زمانی که بین هر جفت انجمن شباهتی بزرگتر از آستانه معین وجود نداشته باشد، بجای تکرار عملیات ادغام تا زمانی که همه رأس‌ها در یک انجمن قرار گیرند، می‌توان رویه ادغام را زودتر خاتمه داد و این یعنی فرآیند ادغام با کارایی بالا انجام می‌شود.

حداکثرش یعنی ۱ می‌رسد. با این حال، بعید است زیرگراف همسایگی برای هر رأس در یک شبکه پراکنده، گرافی کامل باشد، به خصوص برای رأس‌های با درجه بزرگتر. به عبارتی، ضریب خوشه‌بندی برای یک رأس با درجه‌های بزرگتر همیشه کوچک است. بنابراین، رأی دادن به ترتیب صعودی ضریب خوشه‌بندی می‌تواند رأس‌های با درجه بزرگتر را زودتر به عنوان کاندیدا معرفی کرده و آنها را با همسایه‌های خود احاطه کند تا خوشه‌ها را بسازند. در کنار محاسبه ضریب خوشه‌بندی رأس $x \in U$ ، فاصله هلینگر این رأس تا تمامی رأس‌های دیگر از روی ماتریس هلینگر جهت محاسبه مقدار مرکزیت هلرنک آن به دست می‌آید. در گام دوم و در یک حلقه تکرار با تعداد رأس‌ها، رأس u با کمترین ضریب خوشه‌بندی که احتمالاً درجه بالائی دارد انتخاب و مجموعه همسایه‌های مرتبه دوم با درجه بزرگتر از آن به دست می‌آید. اگر این مجموعه تهی باشد، رأس u به عنوان کاندیدا نامزد شده، به خودش رأی داده و اولین خوشه را تشکیل و اگر در همسایه‌های مرتبه دوم رأس u چند رأس وجود داشته باشد که درجه آنها بزرگتر از رأس u است، از شباهت بین آنها و رأس u استفاده شده تا مشخص شود رأس u باید به کدامیک رأی دهد. شباهت بین رأس‌ها نقش مهمی در روند رأی‌گیری دارد. شباهت بین رأس‌های u و v بر اساس مرکزیت هلرنک به صورت زیر محاسبه می‌شود:

$$sim(u, v) = \frac{1}{hellinger\ distance(u,v)} \quad (15)$$

رأس v با بالاترین شباهت انتخاب می‌شود، اگر این شباهت مقدار صفر داشته باشد، رأس u به عنوان کاندیدا نامزد شده، به خودش رأی داده و خوشه‌ای را تشکیل می‌دهد. اگر مقدار این شباهت بالاتر از صفر باشد، وضعیت رأی‌دهی رأس v تعیین کننده است. اگر رأس v به رأس دیگری رأی نداده باشد، رأس u ، رأس v را به عنوان کاندیدا نامزد کرده و به آن رأی می‌دهد. در اینجا رأس v خوشه‌ای را تشکیل داده و رأس u به آن خوشه اضافه می‌گردد. اگر رأس v به رأس دیگری مثل رأس w رأی داده باشد، بدان معناست که رأس v از حق خود برای نامزدی به عنوان یک کاندیدا دست برداشته، در نتیجه رأس u به رأس w رأی می‌دهد. در اینجا رأس u به خوشه‌ای که رأس w ایجاد کرده و رأس v هم در آن عضویت دارد اضافه خواهد شد. پس از رأی‌گیری، خوشه‌های کوچکی به دست می‌آیند که البته انجمن‌های نهایی نیستند. در گام سوم، این خوشه‌ها به عنوان انجمن‌های اولیه در نظر گرفته شده و در یک حلقه تکرار، برخی از آنها برای ایجاد انجمن‌های نهایی ادغام می‌شوند. برای افزایش کارایی، از ایده باربر [۱] بهره گرفته شد و استراتژی مشابهی برای پیوستن یک جفت انجمن که ادغام آنها می‌تواند به بزرگترین



شکل ۲. روند کلی مراحل مختلف روش پیشنهادی

۳- نتایج آزمایش‌های تجربی

در این بخش روش پیشنهادی در مقایسه با دو الگوریتم شناخته شده برای شناسائی انجمن‌ها در شبکه‌های دوبخشی مصنوعی و واقعی به نام AdaptiveBRIM و BiAttractor (به شرح زیر) ارزیابی می‌شود. کد روش پیشنهادی در محیط Colab نوشته شده و نتایج کار، با پیاده سازی این کد بر روی یک سرور با مشخصات Intel E5-2683 v4 Broadwell @ 2.1Ghz حاصل شد.

- الگوریتم AdaptiveBRIM: آقای باربر الگوریتم ماژول‌های

شبه کد بیانگر مراحل مختلف روش پیشنهادی به صورت زیر است:

00 Input: $G = (U, I, E)$, Bipartite Network; Output: CS , Detected Community Structure

01 for $\forall x, y \in U$ do

02 $cc(x) \leftarrow \frac{\text{closed 4-paths centered on node } x}{\text{4-paths centered on node } x}$

03 Calculate *hellinger dist.* (x, y) from *hellinger matrix* of all $u \in U$

04 $V_u \leftarrow U$;

05 while $V_u \neq \emptyset$ do

06 $u \leftarrow \text{argmin}_x \{cc(x) | x \in V_u\}$; $V_u \leftarrow V_u - \{u\}$

07 $LDN1(u) \leftarrow \{v | |N(v)| > |N(u)|, v \in NN(u)\}$

08 if $LDN(u) = \emptyset$ then

09 There is no vertex whose degree is larger than u in the neighborhood of v and u votes for itself

10 for $\forall z \in LDN(u)$ do

11 $sim(u, z) \leftarrow \frac{1}{\text{hellinger dist.}(u, z)}$

12 $v \leftarrow \text{argmax}_z \{sim(u, z) | z \in LDN(u)\}$

13 if $sim(u, v) = 0$ then

14 The similarity between u and v is 0, u nominates itself as a candidate and votes for itself

15 if v has not voted for other vertices then u nominates v as a candidate, and votes for v

16 else v has voted for other vertex, u votes for that vertex also

17 return CS_{init}

18 $CS \leftarrow CS_{init}$

19 $Q \leftarrow \frac{1}{m} \sum_{i=1}^{|CS|} (\Psi_{C_i}^U - \frac{\Psi_{C_i}^U \Psi_{C_i}^I}{m})$

20 while $(|CS| > 1)$ do

21 for $\forall C_i, C_j \in CS$ do

22 if $(\sum_{x \in U_{C_i}} \sum_{y \in U_{C_j}} \frac{1}{\text{hellinger dis.}(x, y)} \geq a \text{ threshold})$ then

23 $\Delta Q \leftarrow \sum_{\forall User \in C_j \& Item \in C_i} \left(\frac{k_{C_i, p}^I}{m} - \frac{k_p \Psi_{C_i}^I}{m^2} \right)$

24 Calculate the modularity after joining the selected pair of communities and record the current largest modularity and the corresponding community structure as a CS_{max}

25 $CS \leftarrow CS_{max}$

26 return CS

Q_B به دلیل مسئله محدودیت تفکیک نیز محدودیت دارد [۲۶].

۳-۱-۱- شبکه‌های مصنوعی

برخی روش‌های شناسائی انجمن وابسته به بهینه‌سازی مقدار ماژولاریتی هستند. این روش‌ها ممکن است مشکل محدودیت تفکیک و محدودیت شناسائی انجمن‌های کوچکتر از مقداری معین، بسته به اندازه شبکه و یال‌های داخلی داشته باشند. برای نشان دادن اثربخشی روش پیشنهادی حلقه‌ای از biclique با تعداد biclique مختلف طراحی شد. هر biclique از دو نوع مجموعه رأس دوتائی و سه‌تائی تشکیل شده که به طور کامل با هم متصل هستند. ویژگی‌های توپولوژیکی پایه آنها در جدول ۱ آمده است. آزمایش‌ها روی حلقه‌های با تعداد biclique مختلف انجام و نتایج تفصیلی در جدول ۲ نشان داده شده و بیانگر آن است که مقدار NMI انجمن‌های حاصل و دقت شناسائی تعداد آنها در روش پیشنهادی بهبود یافته است.

جدول ۱. ویژگی‌های توپولوژیکی پایه حلقه‌های biclique

	n_1	m_1	n	m	$\langle k \rangle$	C	r
4-bicliq.	۱۲	۸	۲۰	۲۸	۲.۸۰	۰.۴۸۲	-۰.۵
8-bicliq.	۲۴	۱۶	۴۰	۵۶	۲.۸۰	۰.۴۸۲	-۰.۵
16-bicliq.	۴۸	۳۲	۸۰	۱۱۲	۲.۸۰	۰.۴۸۲	-۰.۵

جدول ۲. مقایسه کارآئی الگوریتم AdaptiveBRIM، BiAttractor و روش پیشنهادی. NMI دقت شناسائی انجمن‌های دوحالته و N_C تعداد آنها.

	AdaptiveBRIM		BiAttractor		روش پیشنهادی	
	NMI	N_C	NMI	N_C	NMI	N_C
4-bicliq.	۱.۰۰۰	۴	۱.۰۰۰	۴	۱.۰۰۰	۴
8-bicliq.	۱.۰۰۰	۸	۱.۰۰۰	۸	۱.۰۰۰	۸
16-bicliq.	۰.۹۳۴	۱۵	۱.۰۰۰	۱۶	۱.۰۰۰	۱۶

۳-۱-۲- شبکه‌های واقعی

در این بخش، چند مطالعه تجربی در شبکه‌های واقعی شناخته شده بدون ساختار ماژولار، انجام و معیار Q_B برای تأیید صحت استفاده شد. شبکه‌های مورد استفاده SW، AR، SCI و CN (به شرح زیر) هستند که ویژگی‌های توپولوژیکی پایه آنها در جدول ۳ آمده است. نتایج تفصیلی مطالعات در جدول ۴ نشان می‌دهد که مقدار ماژولاریتی انجمن‌های حاصل و دقت شناسائی تعداد آنها در روش پیشنهادی بهبود یافته است. انجمن‌های شناسائی شده در شبکه دوبخشی SW در شکل ۳ قابل مشاهده است.

القائی بازگشتی دوبخشی^۱ را بر اساس ایده به حداکثررسانی مکرر ماژولاریتی Q_B در شبکه دوبخشی پیشنهاد کرد [۱] و [۳۶]. در هر تکرار، تضمین می‌شود که Q_B کاهش نیابد. با این حال، تقسیم‌بندی شناسائی شده شبکه‌های دوبخشی منجر به یک حداکثر محلی به جای حداکثر عمومی Q_B می‌شود. در همین حال تعداد ماژول‌ها نیز با به حداکثر رساندن ماژولاریتی Q_B تعیین می‌شود.

• الگوریتم BiAttractor: آقای سان روش جدیدی با استفاده از دینامیک فاصله برای شناسائی انجمن‌های دوحالته در شبکه‌های دوبخشی بزرگ پیشنهاد داد که در آن از تعامل‌های جامعه انسانی الهام گرفته شده، به طوری که تعامل‌های بیشتر داخل انجمن و تعامل‌های کمتر بین انجمن‌های متفاوت وجود دارد. الگوریتم در شبکه‌های پراکنده پیچیدگی زمانی $O(|E|)$ دارد و تقسیم‌بندی دقیقی از انجمن‌ها را به دست می‌آورد [۳۴].

پیچیدگی زمانی

در روش پیشنهادی، شناسائی انجمن‌های اولیه با شبیه‌سازی فرآیند رأی‌گیری را می‌توان با پیچیدگی زمانی $O(md + n \log n)$ انجام داد که در آن n و m به ترتیب تعداد رأس‌ها در یکی از دو مجموعه رأس و تعداد یال‌های شبکه بوده و d میانگین درجه رأس‌هاست. ادغام برخی انجمن‌های اولیه برای به دست آوردن انجمن‌های نهائی پیچیدگی زمانی $O(mk)$ را داراست، که در آن k دفعات تکرار ادغام بوده و در حالت کلی $k \ll n$ و $d < k$ است. پس کل زمان مصرفی $O(md + n \log n + mk) \sim O(mk)$ در شبکه‌های پراکنده است و نتیجه می‌گیریم که روش پیشنهادی می‌تواند به طور موثر در شبکه‌های بزرگ مقیاس اعمال شود.

پارامترهای ارزیابی

برای مقایسه روش‌های مختلف دو نوع اندازه‌گیری وجود دارد. اگر تقسیمات انجمن‌ها روی شبکه‌های زیرین از قبل داده شده باشد، معیار اطلاعات متقابل نرمال شده^۲ (NMI) برای دادن امتیاز از بازه [۰-۱] اعمال می‌شود [۴۴]، وگرنه معیار ماژولاریتی بکار گرفته می‌شود [۲۵]. ماژولاریتی اصل فقط برای شبکه‌های یک‌بخشی تعریف شده بود اما باربر تعریف ماژولاریتی در شبکه‌های دوبخشی را با عنوان Q_B گسترش داد [۱]. مقدار بالاتر از بازه [۰-۱] نشان دهنده یال درون انجمن بیشتر از یال مورد انتظار مدل تهی است،

² Normalized Mutual Information

¹ Bipartite Recursively Induced Modules

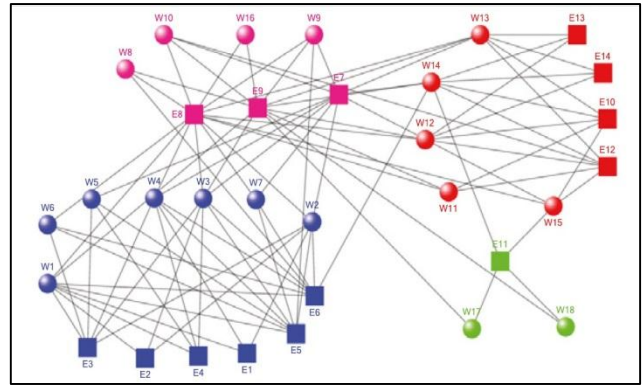
در این شبکه دوبخشی ۱۴۱ رأس و ۱۶۰ یال وجود دارد.

- Scotland Corporate Interlock (SCI): این مجموعه داده حاوی اطلاعاتی در مورد بهم پیوستگی‌های شرکتی در اوایل قرن بیستم در اسکاتلند است. این مجموعه شامل اعضای هیئت مدیره شرکت‌های اسکاتلندی است که در دوره ۱۹۰۴ تا ۱۹۰۵ چندین سمت مدیریتی داشتند. مسئولیت‌ها یک شبکه دوبخشی متشکل از ۱۳۶ فرد و ۱۰۸ را نگهداری می‌کنند.

- Crime Network (CN): این مجموعه داده شامل افرادی است که حداقل در یک پرونده جنائی به عنوان قربانی، شاهد یا مظنون ثبت شده‌اند. روابط بین افراد در ارتباط با جرم و پرونده‌های جرم، شبکه‌ای دوبخشی از ۱۴۷۶ یال را تشکیل می‌دهد که ۸۲۹ نفر و ۵۵۱ پرونده جرم را بهم متصل می‌کنند.

۴- نتیجه‌گیری

ساختار انجمن ویژگی مهمی در شبکه‌های دوبخشی است. برخی از روش‌های شناسائی انجمن دارای مصرف زمان زیاد، محدودیت‌های حل مسئله یا حالت غیرقطعی هستند که باعث محدود شدن اعمال آنها در شبکه‌های مقیاس بزرگ می‌شود. برخی از روش‌ها انجمن‌ها را در شبکه‌های دوبخشی با ساختن یک طرح‌ریزی یک‌بخشی از یک قسمت از شبکه و سپس استفاده از روش‌هایی برای شبکه‌های یک-بخشی شناسائی می‌کنند. نتایج تحقیقات نشان می‌دهد که تجزیه و تحلیل یک طرح‌ریزی یک‌بخشی می‌تواند نتایج غیرقابل اعتماد یا نادرستی بدهد. در این مقاله بر اساس معیار مرکزیت هلرنک، ما یک روش جدید شناسائی انجمن دوحالته را توسعه می‌دهیم، که یک روش شناسائی انجمن سریع و قطعی است. روش پیشنهادی الهام گرفته از رفتارهای رأی‌گیری در فعالیت‌های انتخاباتی در جامعه انسانی است. آزمایش‌های زیادی برای تأیید عملکرد روش پیشنهادی و مقایسه با AdaptiveBRIM و BiAttractor که الگوریتم‌های شناسائی انجمن شناخته شده‌ای در شبکه‌های دوبخشی هستند، انجام گردید. نتایج نشان می‌دهد روش پیشنهادی دقیق‌تر و کارآمدتر از الگوریتم‌های مورد مقایسه است. به عنوان کارهای آتی پیشنهاد می‌شود استفاده از معیارهای مرکزیت دیگر و همچنین الگوبرداری از دیگر رفتارهای موجود در طبیعت و جامعه به منظور بهبود عملکرد الگوریتم‌های شناسائی انجمن در شبکه‌های دوبخشی مورد توجه قرار گیرد.



شکل ۳. انجمن‌های شناسائی شده در شبکه دوبخشی SW

جدول ۳. ویژگی‌های توپولوژیکی پایه شبکه‌های واقعی دوبخشی

	n_1	m_1	n	m	$\langle k \rangle$	C	r
SW	۱۸	۱۴	۳۲	۸۹	۵,۵۶۳	۰,۳۲۸	-۰,۳۳۷
AR	۱۴۶	۵	۱۴۱	۱۶۰	۲,۲۷۰	۰,۷۸۱	-۰,۷۴۳
SCI	۱۰۸	۱۳۶	۲۴۴	۳۵۸	۳,۱۴۰	۰,۳۰۳	-۰,۱۷۱
CN	۸۲۹	۵۵۱	۱۳۸۰	۱۴۷۶	۲,۱۳۹	۰,۴۲۷	-۰,۱۶۶

جدول ۴. مقایسه کارآئی الگوریتم AdaptiveBRIM، BiAttractor و روش پیشنهادی. Q_b ماژولاریتی شبکه‌های دوبخشی و N_C تعداد انجمن‌ها.

	AdaptiveBRIM		BiAttractor		روش پیشنهادی	
	Q_b	N_C	Q_b	N_C	Q_b	N_C
SW	۰,۳۴۵	۴	۰,۳۴۵	۴	۰,۳۵۸	۴
AR	۰,۶۰۲	۵	۰,۶۰۱	۵	۰,۶۳۶	۵
SCI	۰,۶۶۰	۲۴	۰,۶۶۰	۳۹	۰,۸۹۹	۳۵
CN	۰,۷۹۸	۱۰۴	۰,۸۵۹	۱۳۲	۰,۹۳۸	۷۹

- Southern Women events participation (SW): یک مجموعه داده معیاری شناخته شده در زمینه شناسائی انجمن‌های شبکه‌های دوبخشی که دیویس در طول دهه ۱۹۳۰ و در می‌سی‌سی‌پی برای مطالعه طبقه و نژاد جمع‌آوری کرد. در این تقسیم‌بندی ۱۸ زن در ۱۴ رویداد اجتماعی توصیف می‌شوند. زنان و رویدادهای اجتماعی یک شبکه دوبخشی از ۳۲ رأس و ۸۹ یال هستند.
- American Revolution (AR): این مجموعه داده حاوی اطلاعات عضویت ۱۳۶ نفر از ۵ سازمان، قبل از انقلاب آمریکا است. روابط بین افراد شناخته شده و سازمان‌های آنها را می‌توان در یک شبکه دوبخشی توصیف کرد. یال بین یک فرد و یک سازمان نشان می‌دهد که آن فرد عضو این سازمان است.

سپاسگزاری

از خانم فاطمه سرشار تهرانی کمال تشکر و سپاسگزاری را دارم. ایشان در مراحل تهیه مقاله زحمات بسیاری متحمل و مشارکت کاملی داشتند. این مقاله برگرفته از رساله دکتری آقای علی خسروزاده با عنوان روشی چندهدفه برای شناسایی انجمن‌ها در شبکه‌های دویخشی در دانشگاه آزاد اسلامی واحد قزوین می‌باشد.

مراجع

- Statistical Mechanics and its Applications*, vol. 460, pp. 98-104, 2016.
- [18] J. X. Yang and X. D. Zhang, "Finding Overlapping Communities Using Seed Set," *Physica A: Statistical Mechanics and its Applications*, vol. 467, pp. 96-106, 2017.
- [19] S. Bilal and M. Abdelouahab, "Evolutionary Algorithm and Modularity for Detecting Communities in Networks," *Physica A: Statistical Mechanics and its Applications*, vol. 473, pp. 89-96, 2017.
- [20] H. L. Sun, E. Ch'ng, X. Yong, J. M. Garibaldi and S. See, "An Improved Game-Theoretic Approach to Uncover Overlapping Communities," *International Journal of Modern Physics C*, vol. 28, 2017.
- [21] T. S. Evans and R. Lambiotte, "Line Graphs, Link Partitions and Overlapping Communities," *Physical Review E*, vol. 80, 2009.
- [22] V. D. Blondel, J. L. Guillaume, R. Lambiotte and E. Lefebvre, "Fast Unfolding of Communities in Large Networks," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 10, 2008.
- [23] A. Mukherjee, M. Choudhury and N. Ganguly, "Understanding How Both the Partitions of a Bipartite Network Affect its One-Mode Projection," *Physica A: Statistical Mechanics and its Applications*, vol. 390, pp. 3602-3607, 2011.
- [24] X. Wang and X. Qin, "Asymmetric intimacy and algorithm for detecting communities in bipartite networks," *Physica A: Statistical Mechanics and its Applications*, vol. 462, pp. 569-578, 2016.
- [25] M. E. Newman, "Modularity and Community Structure in Networks," *In Proceedings of the National Academy of Sciences of USA*, 2006, vol. 103, pp. 8577-8582.
- [26] S. Fortunato and M. Barthelemy, "Resolution Limit in Community Detection," *In Proceedings of the National Academy of Sciences of USA*, 2007, vol. 104, pp. 36-41.
- [27] S. Lehmann, M. Schwartz and L. K. Hansen, "Biclique Communities," *Physical Review E*, vol. 78, 2008.
- [28] P. Zhang, J. Wang, X. Li, M. Li and Z. Di, "Clustering Coefficient and Community Structure of Bipartite Networks," *Physica A: Statistical Mechanics and its Applications*, vol. 387, pp. 6869-6875, 2008.
- [29] Y. Cui and X. Wang, "Uncovering Overlapping Community Structures by the Key Bi-Community and Intimate Degree in Bipartite Networks," *Physica A: Statistical Mechanics and its Applications*, vol. 407, pp. 7-14, 2014.
- [30] D. B. Larremore and A. Clauset, "Efficiently Inferring Community Structure in Bipartite Networks," *Physical Review E*, vol. 90, 2014.
- [31] Y. Xu, L. Chen, B. Li and W. liu, "Density-Based Modularity for Evaluating Community Structure in Bipartite Networks," *Information Sciences*, vol. 317, pp. 278-294, 2015.
- [32] J. G. Liu, L. Hou, X. Pan, Q. Guo and T. Zhou, "Stability of Similarity Measurements for Bipartite Networks," *Nature*, vol. 6, 2016.
- [33] K. Li and Y. Pang, "A Unified Community Detection Algorithm in Complex Network," *Neurocomputing*, vol. 130, pp. 36-43, 2014.
- [34] H. Liang Sun, E. Ch'ng, X. Yong, J. M. Garibaldi, S. See and D. Bing Chen, "A Fast Community Detection Method in Bipartite Networks by Distance Dynamics," *Physica A: Statistical Mechanics and its Applications*, vol. 496, pp. 108-120, 2018.
- [35] C. Zhou, L. Feng and Q. Zhao, "A Novel Community Detection Method in Bipartite Networks," *Physica A: Statistical Mechanics and its Applications*, vol. 492, pp. 1679-1693, 2018.
- [36] M. J. Barber, M. Faria and L. Streit, "Searching for Communities in Bipartite Networks," *AIP Conference Proceedings*, 2008, vol. 1021.
- [1] M. J. Barber, "Modularity and Community Detection in Bipartite Networks," *Physical Review E*, vol. 76, 2007.
- [2] S. Fortunato, "Community Detection in Graphs," *Physics Reports*, vol. 486, pp. 75-174, 2002.
- [3] M. Newman and M. Girvan, "Finding and Evaluating Community Structure in Networks," *Physical Review E*, vol. 69, 2004.
- [4] R. Guimera and M. Sales-Pardol, "Module Identification in Bipartite and Directed Networks," *Physical Review E*, vol. 76, 2007.
- [5] M. J. Barber, "Detecting Network Communities by Propagating Labels Under Constraints," *Physical Review E*, vol. 80, 2009.
- [6] A. Silva, S. Guimaraes, W. Meira and M. Zaki, "Profilerank: Finding Relevant Content and Influential Users," *In Proceedings of the 7th Workshop on Social Network Mining and Analysis*, 2013, pp. 1-9.
- [7] H. Mahyar, "Detection of Top-k Central Nodes in Social Networks: A Compressive Sensing Approach," *IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, 2015, pp. 902-909.
- [8] H. Mahyar, H. R. Rabiee, A. Movaghar, E. Ghalebi and A. Nazemian, "Cs-Comdet: A Compressive Sensing Approach for Intercommunity Detection in Social Networks," *IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, 2015, pp. 89-96.
- [9] M. Kitsak, L. K. Gallos, S. Havlin, F. Liljeros, L. Muchnik, and H. E. Stanley, "Identifying Influential Spreaders in Complex Networks," *Nature Physics*, vol. 6, pp. 888-893, 2010.
- [10] S. M. Taheri and H. Mahyar, "Hellrank: A Hellinger-Based Centrality Measure for Bipartite Social Networks," *Social Network Analysis and Mining*, vol. 7, 2017.
- [11] M. Nikulin, *Hellinger distance*, Encyclopedia of Mathematics, 2001.
- [12] M. E. Newman, "Finding Community Structure in Networks Using the Eigenvectors of Matrices," *Physical Review E*, vol. 74, 2006.
- [13] T. Opsahl, "Triadic Closure in Two-Mode Networks: Redefining the Global and Local Clustering Coefficients," *Social Networks*, vol. 35, pp. 159-167, 2013.
- [14] A. Nematzadeh, E. Ferrara, A. Flammini and Y. Y. Ahn, "Optimal Network Modularity for Information Diffusion," *Physical Review Letters*, vol. 113, 2014.
- [15] T. Zhou, M. Zhao, G. Chen, G. Yan and B. H. Wang, "Phase Synchronization on Scale-Free Networks with Community Structure," *Physics Letters A*, vol. 368, pp. 431-434, 2007.
- [16] L. Y. Tang, S. N. Li, J. H. Lin and Q. Guo, "Community Structure Detection Based on the Neighbor Node Degree Information," *International Journal of Modern Physics C*, vol. 27, 2016.
- [17] H. Peng, D. Zhao, L. Li and J. Lu, "An Improved Label Propagation Algorithm Using Average Node Energy in Complex," *Physica A:*

- [41] G. Calderer, M. L. Kuijjer, "Community Detection in Large-Scale Bipartite Biological Networks," *Frontiers in Genetics*, vol. 12, 2021.
- [42] G. Wu, C. Gu, H. Yang, "A Spectral Method of Modularity for Community Detection," *EPL Journal*, vol. 137, 2022.
- [43] O. F. Robledo, M. Klepper, E. Boven, H. Wang, "Community Detection for Temporal Weighted Bipartite Networks," *In Proceedings of the 11th International Conference on Complex Networks and their Applications*, 2023, pp. 245-257.
- [44] P. Zhang, "Evaluating Accuracy of Community Detection Using the Relative Normalized Mutual Information," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 11, 2015.
- [37] H. Taguchi, T. Murata, X. Liu, "BiMPLA: Community Detection in Bipartite Networks by Multi-Label Propagation," *In Proceedings of NetSciX 2020: 6th International Winter School and Conference on Network Science*, 2020, pp. 17-31.
- [38] M. Bouguessa, K. Nouri, "BiNeTClus: Bipartite Network Community Detection Based on Transactional Clustering," *ACM Transactions on Intelligent Systems and Technology*, vol. 12, pp. 1-26, 2020.
- [39] T. C. Yen, D. B. Larremore, "Community Detection in Bipartite Networks with Stochastic Blockmodels," *Physical Review E*, vol. 102, 2020.
- [40] S. Che, W. Yang, W. Wang, "An Improved Artificial Bee Colony Algorithm for Community Detection in Bipartite Networks," *IEEE Access*, vol. 9, pp. 10025-10040, 2021.