تحلیلی بر روش‌های تشخیص همبستگی‌های همپوشان در شبکه‌های اجتماعی

معصومه زجاجی[1]، محمد جواد کارگر[2]

[1] دانشجوی دکترا، گروه کامپیوتر، دانشگاه آزاد اسلامی واحد میبد، میبد
Masoumeh.zojaji@gmail.com

[2] استادیار، گروه کامپیوتر، دانشگاه علم و فرهنگ، تهران
kargar@usc.ac.ir

چکیده

گراف‌ها نقش مهمی را در سیستم‌های پیچیده ایفا می‌کنند از شبکه‌های کامپیوتری گرفته تا حوزه‌های بیولوژی و جامعه شناسی. در واقع هر رابطه M:N در اصطلاح پایگاه داده‌ای می‌تواند به عنوان گراف ارائه شود. با توجه به محبوبیت و گسترش روز افزون شبکه‌های اجتماعی آنلاین، تجزیه و تحلیل آن‌ها مورد توجه بسیاری از پژوهشگران در حوزه‌های مختلف علمی قرار گرفته است اعم از جامعه شناسی، بازاریابی و غیره. در این تحقیق ابتدا گراف کاوی در حوزه‌های مختلف نرم افزار، تصویر، سبد کالا و... بررسی و گراف کاوی در شبکه‌های اجتماعی انتخاب شده است. سپس روش‌های مطرح شده در حوزه‌های مختلف گراف کاوی شبکه‌های اجتماعی شامل شناسایی میزان نفوذ پذیری، مدل‌ها، متریک‌ها و پویایی‌ها، شناخت روابط و تعاملات گروهی، انتشار اطلاعات، روش‌های مختلف تشخیص همبستگی‌ها در گراف کاوی در نظر گرفته شده است.

شبکه‌های اجتماعی شامل چندین همبستگی می‌باشند که هر همبستگی با مجموعه‌ای از نودهایی تعریف می‌شود که ارتباطات در داخل همبستگی فشرده تر از همبستگی‌های دیگر می‌باشد. روش‌های متفاوتی در این زمینه ارائه شده‌اند اما آنچه حائز اهمیت است نادیده گرفتن همپوشانی همبستگی‌های می‌باشد. هدف از ارائه این مقاله، تحلیل روش‌های مطرح شده در حوزه تشخیص همبستگی‌های همپوشان می‌باشد. یکی از مشکلات این الگوریتم‌ها پیچیدگی زمانی آن‌هاست که در گراف‌هایی با مقیاس بزرگ نمود پیدا می‌کند. با توجه به اینکه برای گراف کاوی از ماتریس مجاورتی استفاده می‌شود بنابراین نمی‌توان به زمان کمتر از $O(n^2)$ دست یافت. اگرچه با ظهور الگوریتم‌های معنایی شاید بتوان تا حدودی این زمان را تقلیل داد.

کلمات کلیدی

شبکه‌های اجتماعی، گراف کاوی، همبستگی، همبستگی‌های همپوشان.

Archive of SID

www.SID.ir

۱- مقدمه

امروزه تجزیه و تحلیل شبکه‌های اجتماعی مورد توجه بسیاری از پژوهشگران در حوزه‌های مختلف قرار گرفته است که یکی از عوامل آن محبوبیت روز افزون شبکه‌های اجتماعی در فضای مجازی می‌باشد. در دسترس بودن داده‌های شبکه‌های اجتماعی آنلاین و ساختار گراف گونه آن‌ها روند تحقیقات در این زمینه را نیز تسهیل نموده است.

شبکه‌های اجتماعی را می‌توان از جنبه‌های متفاوت مانند روانشناسی، جامعه شناسی، اهداف تجاری و تبلیغاتی تجزیه و تحلیل نمود. یکی از روش‌های تحلیل این شبکه‌ها داده کاوی است. علم داده کاوی انسان را قادر می‌سازد که حجم عظیمی از داده‌ها را مورد پردازش عمیق قرار دهد و کلیه نظم‌هایی که در عمق داده وجود دارند را تشخیص نموده و جهت استفاده عرضه نماید.

گراف کاوی یکی از تکنیک های داده کاوی می باشد که امروزه به دلیل کاربرد در حوزه‌های مختلف اعم از زیست شناسی، شبکه‌های کامپیوتری، وب، اشکال‌زدایی نرم افزاری و غیره بسیار مورد استفاده قرار گرفته است. در ضمن بسیاری از انواع جدید داده‌ها مانند XML را نیز می‌توان در قالب گراف ارائه نمود[۹-۱]. گراف مجموعه‌ای از گره‌هایی می‌باشد که توسط یال‌ها به یکدیگر متصل می‌شوند. در خیلی از زمینه‌ها داده‌ها ساختاری گراف گونه دارند اعم از شبکه‌های کامپیوتری که شامل تعدادی مسیریاب یا کامپیوتر می‌باشند که توسط رسانه‌های انتقال به یکدیگر متصل شده و تبادل داده بین آن‌ها وجود دارد در چنین ساختاری مسیریاب یا کامپیوترها نودهای یک گراف و اتصال بین آن‌ها یال‌های گراف را تشکیل می‌دهند، شبکه‌های اجتماعی که شامل افرادی می‌باشند که به دلایل کاری، خویشاوندی یا موارد دیگر با یکدیگر ارتباط دارند یا شبکه‌های زیستی که پروتئین‌ها باید برای انجام برخی از کنش‌های زیستی با یکدیگر همکاری نمایند. در این زمینه‌ها و خیلی موارد دیگر گراف‌ها ساختار اصلی داده‌ها را تشکیل می‌دهند.

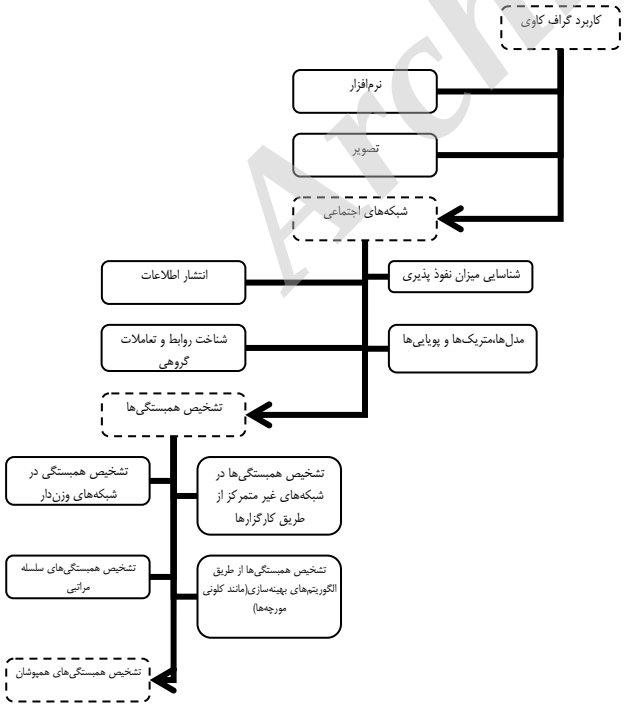
اگرچه در گراف کاوی شبکه‌های اجتماعی، مباحث اصلی و پایه مانند بهبود الگوریتم‌های جستجو در گراف، کاهش هزینه‌های محاسباتی و مواردی از این قبیل مورد بررسی قرار گرفته‌اند اما آنچه محققان علوم مختلف را به این حوزه ترغیب نموده است مفاهیم جامعه شناسی می‌باشد.

موضوعات اصلی مطرح شده در گراف کاوی شبکه‌های اجتماعی[۹]:

- تشخیص همبستگی‌ها: یکی از مهمترین ویژگی‌ها در هر شبکه اجتماعی تشخیص وابستگی‌هاست. انسان‌ها موجوداتی اجتماعی هستند و بر حسب علایق خود به گروه‌هایی اجتماعی ملحق می‌شوند. از این رو تشخیص همبستگی‌ها در یک گروه در شناخت علایق یک جامعه بسیار موثر است.
- شناسایی میزان نفوذ پذیری: در این موضوع محققان به دنبال پاسخ دادن به چنین سوالاتی هستند، چه فردی در یک شبکه اجتماعی موثرتر است یا نفوذ کدام یک از صفحات وب در اینترنت بیشتر است.
- مدل‌ها، متریک‌ها و پویایی‌ها: این موضوع از این جنبه مورد علاقه می‌باشد؛ استفاده از اطلاعات جغرافیایی، چگونگی شکل‌گیری شبکه‌ها و چگونگی پردازش مجموعه حجیم داده‌ها بطور موثر.
- شناخت روابط و تعاملات گروهی: این موضوع که مبتنی بر تحلیل محل و موقعیت می‌باشد از سه دیدگاه مطرح می‌شود: شناخت روابط از این جهت که کاربران چگونه تعامل دارند، چگونه تیم‌های مناسب را شکل می‌دهند و کاربران یک شبکه اجتماعی چه کسانی هستند.
- انتشار اطلاعات: این موضوع قابل توجه افرادی است که به دلایل تجاری قصد ارسال تبلیغات درست به مردم را دارند. یکی دیگر از موارد بررسی شبکه‌های اجتماعی از این دیدگاه، دنبال کردن این است که چه نوع اطلاعاتی از طریق یک شبکه مشخص و به چه کسی ارسال می‌شود.

یکی از ساختارهای کلیدی در شبکه‌های اجتماعی آنلاین و هر شبکه اجتماعی، همبستگی است. در سال‌های اخیر مطالعه این ساختار مورد توجه محققان علوم مختلف قرار گرفته است. شبکه‌های اجتماعی شامل چندین همبستگی می‌باشند که هر همبستگی با مجموعه‌ای از نودهایی تعریف می‌شود که تعاملات بین آنها در داخل همبستگی فشرده تر از سایر همبستگی‌ها می‌باشد.

اکثر روش‌های ارائه شده برای تشخیص همبستگی در شبکه‌های اجتماعی بر روی خوشه‌های مجزا کار می‌کنند یعنی هر نود به یک همبستگی واحد نسبت داده می‌شود[۱۵-۱۰]. همانطور که می‌دانیم افراد در شبکه‌های اجتماعی مختلف فعالیت دارند از این رو همپوشانی همبستگی‌ها باید مورد توجه قرار گیرد. بطور مثال یک شخص ممکن است با چندین گروه اجتماعی در ارتباط باشد مانند گروه دوستان، اقوام، همکاران و در عین حال با گروه‌های دیگر مانند گروه‌های علمی هم در ارتباط باشد. این مسئله هم در شبکه‌های زیستی مطرح است بطوریکه یک ژن ممکن است متعلق به گروه‌های متفاوتی باشد[۱۶]. در این زمینه روش‌های متفاوتی مطرح شده‌اندکه کار آن‌ها شناسایی خوشه‌هایی است که الزاما جدا از یکدیگر نمی‌باشند زیرا نودهایی وجود دارند که متعلق به بیش از یک خوشه هستند. هدف از ارائه این تحقیق بررسی روش‌های موجود در زمینه همبستگی کاوی و تشخیص همپوشانی این همبستگی‌ها است. روند انجام تحقیق در شکل (۱) نمایش داده شده است.



شکل(۱):روند انجام تحقیق

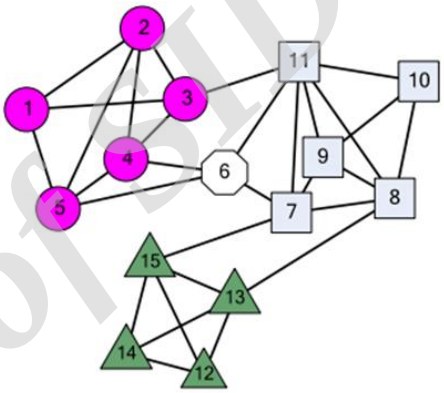
Archive of SID

www.SID.ir

۲- گراف کاوی

گراف کاوی نوع خاصی از داده‌کاوی است که برای حوزه‌هایی که داده‌های آن‌ها بصورت گراف ارائه می‌شوند، مناسب می‌باشد. یکی از کاربردهای گراف کاوی در تجزیه و تحلیل شبکه‌های اجتماعی تشخیص گروه‌هایی با ویژگی‌های مشابه می‌باشد. که در این تحقیق چنین گروه‌هایی با نام همبستگی مطرح شده‌اند. منظور از همبستگی گروه‌هایی شامل چندین نود می‌باشد بطوریکه اعضای یک گروه دارای ارتباط بیشتری نسبت به سایر گروه‌ها می‌باشند. در واقع چگالی یا تراکم پیوندها در داخل همبستگی‌ها بیشتر از خارج همبستگی‌ها می‌باشد.

تشخیص همبستگی که جزء روش‌های خوشه بندی محسوب می‌شود یکی از مباحث مهم در علوم کامپیوتر می‌باشد. در شکل (۲) سه همبستگی با اشکال هندسی دایره، مربع و مثلث آورده شده است که نود ۶ یک نود مشترک است و متعلق به همبستگی دایره شکل و مربع شکل می‌باشد.



شکل(۲): نمونه‌ای از همبستگی‌ها[۱۷]

در ارتباط با موضوع تشخیص همبستگی یا خوشه‌ها به مجموعه‌ای از خوشه‌های یافت شده یک پوشش گفته می‌شود شامل $C = \{C_1, C_2, ..., C_n\}$ که در آن هر نود ممکن است متعلق به بیش از یک خوشه باشد. برای هر نود i یک فاکتور تعلق در ارتباط با همبستگی ارائه می‌شود $[a_{i1}, a_{i2}, ..., a_{ik}]$ بطوریکه a_{ic} معیاری است برای ارزیابی میزان تعلق نود i و خوشه c و باید شرایط فرمول(۱) و (۲) برقرار باشد:

$$0 \le a_{ic} \le 1, \forall_i \in V, \forall_c \in C \qquad (۱)$$

$$\sum_{c=1}^{|c|} a_{ic} = 1 \qquad (۲)$$

و |c| بیانگر تعداد خوشه‌ها یا همبستگی‌هاست. همپوشانی بین دو همبستگی مجموعه‌ای از نودهایی است که بین آن‌ها به اشتراک گذاشته می‌شود(فرمول(۳)).

$$over(C_i, C_j) = \{V \mid V \in C_i \; and \; V \in c_j\} \qquad (۳)$$

بطورکلی نتایجی که الگوریتم‌های تشخیص همبستگی‌های همپوشان ایجاد می‌کنند، می‌تواند در زمره یکی از دو نوع غیرفازی و فازی قرار گیرند. در نوع غیرفازی رابطه بین یک نود و یک خوشه باینری است. یعنی نود i یا متعلق به خوشه c می‌باشد یا نمی‌باشد. در نوع فازی، هر نود با یک درجه تعلق با یک همبستگی در ارتباط می‌باشد. البته با داشتن یک حد آستانه می‌توان نوع فازی را به راحتی به نوع غیر فازی تبدیل نمود. خروجی اکثر الگوریتم‌ها از نوع غیر فازی می‌باشد.

در بخش بعدی الگوریتم‌های تشخیص همبستگی‌های همپوشان بررسی شده‌اند.

۳- الگوریتم‌ها

ساختارهای همبستگی یا ماژولار به عنوان یکی از ویژگی‌های مهم شبکه‌های اجتماعی در دنیای واقعی محسوب می‌شوند. گذشته از تعریف مبهم همبستگی، تکنیک‌های بیشماری برای تشخیص کارا و موثر همبستگی‌ها ارائه شده‌اند. کار عمده این روش‌ها تشخیص همبستگی‌های جدا از هم و بدون همپوشانی است. اگرچه همانطور که در شبکه‌های اجتماعی در دنیای واقعی مشهود است، افراد در گروه‌های متفاوتی عضو هستند. بطور مثال، یک شخص معمولا با چندین گروه اجتماعی مانند خانواده، دوستان و همکاران در ارتباط می‌باشد. یک محقق می‌تواند در حوزه‌های مختلف اجتماعی فعالیت داشته باشد. در ضمن در شبکه‌های اجتماعی آنلاین، تعداد گروه‌هایی که یک فرد می‌تواند به آن‌ها تعلق داشته باشد، نامحدود است.

بنابراین همپوشانی در شبکه‌های اجتماعی یکی از ویژگی‌های مهم محسوب می‌شود. به همین دلیل، رشد قابل توجهی در ارائه الگوریتم‌های تشخیص همبستگی‌های همپوشان اتفاق افتاده است که کار عمده آن‌ها شناسایی مجموعه‌ای از خوشه‌هایی است که لزوما جدا از هم نمی‌باشند و نودهایی وجود دارند که متعلق به بیش از یک خوشه می‌باشند.

همانطور که در شکل(۳) مشخص شده است روش‌های تشخیص همبستگی‌های همپوشان در شش گروه مطالعه و بررسی شده‌اند[۱۸]:

- روش‌های مبتنی بر گروه‌های کامل: در روش‌های مبتنی گروه کامل فرض بر این است که یک همبستگی شامل مجموعه‌های همپوشانی از زیرگراف‌های کاملا متصل است و تشخیص همبستگی‌ها با جستجو برای گروه‌های مجاور انجام می‌شود. روش کار بدین ترتیب است که در ابتدا همه گروه‌ها با اندازه k در یک شبکه شناسایی می‌شوند و سپس یک گراف جدید ساخته می‌شود. در گراف ایجاد شده هر نود بیانگر یکی از گروه‌ها با اندازه k می‌باشد. دو نود در صورتی با هم ارتباط دارند اگر گروه‌ها k-1 عضو را به اشتراک بگذارند. مولفه‌های متصل در گراف شناسایی می‌شوند بطوریکه گروه‌ها همبستگی‌ها را تشکیل می‌دهند. از آنجایی که یک نود می‌تواند بطور همزمان در چندین گروه با اندازه k باشد، همپوشانی بین همبستگی‌ها امکان پذیر است.CMP معروفترین الگوریتم این گروه می‌باشد[۱۹]. یکی از محدودیت‌های CMP این است که فقط می‌تواند گروه‌های کامل بزرگ را پیدا کند و به سختی نوع کوچک آن پیدا می‌شود. پالا[۱۹] نرم‌افزاری به نام CFinder را برای الگوریتم CMP طراحی نموده است. CMPw [۲۰] نمونه توسعه یافته CMP می‌باشد که از یک حد آستانه چگالی برای در نظر گرفتن گروه‌های کامل در یک همبستگی بهره برده است. گذشته از سادگی مفهومی الگوریتم‌هایی مانند CMP آن‌ها بیشتر شبیه تطبیق الگو می‌باشند تا یافتن همبستگی، زیرا هدف آن‌ها یافتن

Archive of SID

www.SID.ir

ساختارهایی مشخص و متمرکز در یک شبکه می‌باشد. لی و همکارانش[۲۱] به جای یافتن گروه‌های کامل با اندازه مشخص، بواسطه جستجوی عمقی و سطحی از گروه‌های کامل بیشینه بهره برده‌اند. در الگوریتم پیشنهادی گره‌هایی که از درجه یک می‌باشند حذف می‌شوند و گره‌هایی با بیشترین درجه و همسایگان آن پیدا شده تا یک گروه کامل بیشینه شکل گیرد. سپس از معیارهای مشابهت بین دو گروه برای الحالق زیر گروه‌ها و تشکیل یک گروه بزرگتر استفاده شده است. الگوریتم MOHCC [۲۲] نمونه دیگری از الگوریتم‌هایی است که در ابتدا گروه‌های کامل بیشینه را استخراج و گروه‌هایی که بر اساس یک معیار بیشترین شباهت را دارند با یکدیگر ادغام شده و یک زیرگراف کامل بزرگتر را شکل می‌دهند. اما از آنجایی که این الگوریتم برای تشخیص همبستگی‌های سلسله مراتبی نیز توسعه داده شده است، از گروه‌های کامل استخراج شده برای ایجاد دندوگرام استفاده شده است. الگوریتم ACC[۲۳] در ابتدا تمامی زیرگراف‌های بیشینه را از شبکه اصلی استخراج و سپس آن‌ها را با در نظر گرفتن ضریب خوشه‌بندی بین دو زیر گراف بیشینه و همسایه ادغام می‌نماید.

- روش‌های مبتنی بر افراز بندی پیوند: روش‌هایی که در این گروه قرار می‌گیرند به جای گره‌ها از پیوندها به عنوان واحدهای اولیه بهره می‌برند. یک گره در گراف اصلی در صورتی همپوشان نامیده می‌شود اگر پیوندهای متصل به آن متعلق به بیش از یک خوشه باشند. اگرچه افرازبندی پیوند برای تشخیص همپوشانی از نظرمفهومی طبیعی است اما هیچ تضمینی وجود ندارد که تشخیص همبستگی‌ها با کیفیت بالاتری نسبت به تشخیص مبتنی بر گره انجام شود زیرا این الگوریتم‌ها به یک تعریف مبهم همبستگی متکی می‌باشند. از جمله الگوریتم‌های شاخص این گروه می‌توان به ELC[۲۴] و کار آنه و همکاران [۲۵] اشاره نمود. ایوانز و لامبیوته [۲۶،۲۷] شبکه را در یک گراف وزن‌دار طرح ریزی نموده‌اند که در آن هر گره بیانگر پیوندهای گراف اصلی می‌باشد. سپس الگوریتم‌های تشخیص همبستگی‌های مجزا بکارگرفته می‌شوند. افرازبندی گره یک گراف خطی منجر به افرازبندی یال در گراف اصلی می‌شود. کیم و همکاران[۲۸] برای تشخیص همبستگی‌ها در شبکه‌های اجتماعی متدی را ارائه نمودند که از خوشه بندی یال‌های مبتنی بر تعاملات بهره می‌برد. در روش پیشنهادی در هنگام شناسایی همبستگی‌های همپوشان و سلسله مراتبی، هم توپولوژی شبکه مورد نظر و هم چگالی تعاملی در نظر گرفته می‌شود. این روش شامل دو فرآیند می‌باشد. در ابتدا ساختارهای همبستگی از طریق خوشه بندی یال که مبتنی بر خوشه بندی سلسله مراتبی تک پیوندی می‌باشد، شناسایی می‌شوند. در مرحله دوم کیفیت ساختارهای همبستگی شناسایی شده با استفاده از یک تابع کیفیت ارزیابی می‌شود.

  GaoCD [۲۹] از الگوریتم ژنتیک برای تشخیص همبستگی‌های همپوشان که مبتنی بر خوشه‌بندی پیوند می‌باشد، بهره برده است. از جمله الگوریتم‌های دیگری که در این گروه قرار می‌گیرند می‌توان به کار سیو و همکاران [۳۰]، جین و همکارن [۳۱] اشاره نمود.
- گسترش محلی و بهینه سازی: الگوریتم‌هایی که از گسترش محلی و بهینه سازی بهره می‌برند مبتنی بر روش‌هایی می‌باشند که با رشد و گسترش یک همبستگی طبیعی یا یک همبستگی جزئی توام می‌باشند. اکثر آن‌ها وابسته به یک تابع سودمند محلی می‌باشند که کیفیت گروهی از نودهای متصل متراکم را مشخص می‌کند. LFM [۳۲] از جمله قدیمی‌ترین الگوریتم‌های این گروه می‌باشد. این الگوریتم گسترش یک همبستگی را یک گره آغاز می‌کند که بطور تصادفی انتخاب می‌شود، اتمام روند گسترش همبستگی با یک تابع برازندگی مشخص می‌شود. پس از یافتن یک همبستگی، LFM بطور تصادفی گره‌های دیگری که هنوز به هیچ همبستگی انتساب داده نشده‌اند، را انتخاب می‌نماید تا یک همبستگی جدید رشد پیدا نماید. این روش هم برای تشخیص همبستگی‌های همپوشان و هم برای نوع سلسله مراتبی آن مناسب می‌باشد. برای کنترل اندازه گروه‌های شکل گرفته از یک پارامتر کنترلی استفاده می‌شود. MONC [۳۳] از تابع برازندگی تغییر یافته LFM استفاده نموده است. که باعث می‌شود یک گره واحد، خودش به عنوان یک همبستگی در نظر گرفته شود. تابع برازندگی پیشنهاد شده باعث می‌شود که MONC محدوده‌ای از پارامترهای کنترلی را برای مجموعه‌ای از گره‌هایی پیدا نماید که بطور محلی بهینه می‌باشند. چن و همکاران [۳۴] هم متدی شامل دو مرحله ۱) یافتن همبستگی اولیه و ۲) گسترش همبستگی را پیشنهاد داده‌اند. در این متد برای تشخیص ساختارهای همبستگی همپوشان، نودها با یک برچسب مناسب علامت گذاری می‌شوند. راموئا و همکاران [۱۷] الگوریتمی به نام DOCNet ارائه نمودند که استراتژی اصلی آن یافتن یک هسته اولیه و سپس افزودن گره‌های مناسب به این هسته و گسترش آن تا هنگامی که معیارهای توقف برآورده شود.

  EM-BOARD [۳۵] پس از استخراج تمامی همبستگی‌های ورودی، از یک تابع برای محاسبه درجه جذب بین دو گره مجاور در همبستگی ورودی استفاده می‌نماید. با توجه به مقدار درجه جذب، همبستگی‌های اولیه بطور ثابت رشد می‌کنند. CONA [۳۶] از جمله الگوریتم‌های این گروه می‌باشد که در شبکه‌های اجتماعی بزرگ بکار گرفته شده است. این الگوریتم گره‌های همپوشان را از طریق مجموعه گره‌های لبه شناسایی می‌کند.
- تشخیص فازی: الگوریتم‌های تشخیص همبستگی فازی، کیفیت استحکام ارتباطی بین همه جفت گره‌ها و همبستگی را مشخص می‌کند. در این الگوریتم‌ها، یک بردار عضویت یا فاکتور تعلق برای هر گره محاسبه می‌شود[۳۷]. مشکل چنین الگوریتم‌هایی تعیین بعد بردار عضویت می‌باشد. C-means فازی [۳۸] و NMF [۳۹] جزء الگوریتم‌های مشهور این گروه می‌باشند. الگوریتم NRATIO[۴۰] از نسبت مجاورتی همپوشان برای نمایش رابطه بین گره‌ها استفاده می‌کند. در واقع این نسبت اتصال بین گره‌هایی که انتظار می‌رود در یک همبستگی یکسان باشند را اندازه گیری می‌نماید. در روش فازی مطرح شده توسط سان و همکاران [۴۱] ابتدا ماتریس مجاورتی برگرفته از گراف اصلی به یک ماتریس رابطه‌ای فازی تبدیل و سپس یک رابطه معادل فازی از ان گرفته می‌شود و درنهایت برای تشخیص همبستگی‌های همپوشان براساس یک حد آستانه قوانین خوشه بندی بکار گرفته می‌شوند.
- الگوریتم‌های پویا و مبتنی بر کارگزار: روش‌های مبتنی بر کارگزار از برچسب‌هایی برای شناسایی میزان عضویت رئوس به همبستگی‌ها بهره

Archive of SID

www.SID.ir

می‌برند، سپس این برچسب‌ها در بین همسایه‌های آنها منتشر می‌شوند. ازآنجایی که هدف از ارائه این الگوریتم‌ها تشخیص همبستگی‌های همپوشان می‌باشد بنابراین یک نود می‌تواند بیش از یک برچسب داشته باشد. COPRA [۴۲] از الگوریتم‌های معروف این گروه به شمار می‌رود. بدیعی و همکاران [۴۳] یک الگوریتم به نام SATOCI را ارائه نموده‌اند که در واقع از گروهی از کارگزارها برای شناسایی همبستگی‌های همپوشان بکار گرفته شده است. این الگوریتم هم قادر به شناسایی همبستگی‌های ناهمپوشان و هم همبستگی‌های همپوشان می‌باشد. الگوریتم AntCBO[۴۴] از کلونی مورچه‌های به عنوان کارگزار برای تشخیص گره‌های همپوشان استفاده نموده است.

- تشخیص معنایی همبستگی: هدف از تشخیص همبستگی معنایی، خوشه بندی گره‌ها با محتوای معنایی مشابه به همبستگی یکسان است. از آنجایی که همبستگی‌های معنایی، هم توسط محتوا و هم رابطه‌ی بین گره‌ها تشخیص داده می‌شوند بنابراین نتایج بهتری را به دنبال خواهند داشت. از آنجایی که داده‌کاوی معنایی باید مبتنی بر آنالیز متن باشد از این رو اکثر الگوریتمهای تشخیص همبستگی معنایی از مدل LDA [۴۵،۴۶] به عنوان مدل اصلی استفاده می‌نمایند.

شکل(۳): رده‌بندی الگوریتم‌های تشخیص همبستگی‌های همپوشان

٤- نتیجه

در این مقاله هدف از تحلیل تعداد بسیاری از الگوریتم‌های تشخیص همبستگی همپوشان رده بندی آن‌ها در شش گروه: روش‌های مبتنی بر گروه‌های کامل، روش‌های مبتنی بر افراز بندی پیوند، گسترش محلی و بهینه سازی، تشخیص فازی، الگوریتم‌های پویا و مبتنی بر کارگزار و تشخیص معنایی همبستگی بوده است. اگرچه برخی از الگوریتم‌ها ممکن است در این حوزه قرار نگیرند[۴۷]. نایج تحلیل این الگوریتم‌های حاکی از آن است که این الگوریتم‌ها بر روی شبکه‌های اجتماعی با مقیاس کوچک آزمایش شده‌اند مانند شبکه اجتماعی دلفین‌ها با ۳۴ گره، شبکه اجتماعی کالج فوتبال با ۱۱۵ گره، شبکه اجتماعی باشگاه کاراته با ۳۴ گره یا شبکه‌های اجتماعی آنلاین مانند فیسبوک با ۱۰ هزار گره. بنابراین با توجه به گسترش روز افزون شبکه‌های اجتماعی آنلاین وعضویت نامحدود افراد در گروه‌های مختلف، هنوز باید کارهای بیشتری در این حوزه انجام شود و این الگوریتم‌ها برای بکارگیری در شبکه‌های اجتماعی با مقیاس بزرگ بهبود داده شوند. یکی از مشکلات این الگوریتم‌ها پیچیدگی زمانی آن‌هاست که در گراف‌هایی با مقیاس بزرگ نمود پیدا می‌کند. با توجه به اینکه برای گراف کاوی از ماتریس مجاورتی استفاده می‌شود بنابراین نمی‌توان به زمان کمتر از $O(n^2)$ دست یافت. اگرچه با ظهور الگوریتم‌های معنایی[۴۵، ۴۶] بتوان تا حدودی این زمان را تقلیل داد. گذشته از کارهای زیادی که در این حوزه انجام شده هنوز این سوال در ذهن باقی می‌ماند که چه موقع روش‌های همپوشان باید بکار گرفته شوند و ضرورت همپوشانی تا چه حد است.

٥- مراجع

[1] *Chakrabarti, Deepayan, and Christos Faloutsos. "Graph mining: Laws, generators, and algorithms." ACM Computing Surveys (CSUR) 38.1 (2006): 2.*

[2] *Tekin, Umut, and Feza Buzluca. "A graph mining approach for detecting identical design structures in object-oriented design models." Science of Computer Programming, vol 95, pp. 406-425, 2014.*

[3] *Pears, Russel, Songwut Pisalpanus, and Yun Sing Koh. "A graph based approach to inferring item weights for pattern mining." Expert Systems with Applications, vol 42.1, pp. 451-461, 2015.*

[4] *Shelokar, Prakash, Arnaud Quirin, and Óscar Cordón. "A multiobjective evolutionary programming framework for graph-based data mining." Information Sciences, vol 237, pp. 118-136, 2013.*

[5] *Ríos, Sebastián A., and Ivan F. Videla–Cavieres. "Generating groups of products using graph mining techniques." Procedia Computer Science, vol 35, pp.730-738, 2014 .*

[6] *Morales-González, Annette, et al. "A new proposal for graph-based image classification using frequent approximate subgraphs." Pattern Recognition, vol 47.1, pp. 169-177, 2014.*

[7] *Videla-Cavieres, Ivan F., and Sebastián A. Ríos. "Extending market basket analysis with graph mining techniques: A real case." Expert Systems with Applications, vol 41.4, pp. 1928-1936, 2014.*

[8] *Jiang, Xing, et al. "Mining globally distributed frequent subgraphs in a single labeled graph." Data & Knowledge Engineering, vol 68.10, pp. 1034-1058, 2009.*

[9] *Nettleton, David F. "Data mining of social networks represented as graphs." Computer Science Review, vol 7, pp. 1-34, 2013.*

[10] *Newman, Mark EJ, and Michelle Girvan. "Finding and evaluating community structure in networks." Physical review E, vol 69.2, pp. 026113, 2004.*

[11] *Blondel, Vincent D., et al. "Fast unfolding of communities in large networks." Journal of Statistical Mechanics: Theory and Experiment, vol 2008.10, pp. P10008, 2008.*

[12] *Yang, Bo, et al. "Hierarchical community detection with applications to real-world network analysis." Data & Knowledge Engineering, vol 83, pp. 20-38, 2013.*

[13] *Romdhane, L. Ben, et al. "A robust ant colony optimization-based algorithm for community mining in large scale oriented social graphs." Expert Systems with Applications, vol 40.14, pp. 5709-5718, 2013.*

Archive of SID

www.SID.ir

[14] Liu, Ruifang, et al. "Weighted Graph Clustering for Community Detection of Large Social Networks." Procedia Computer Science, vol 31, pp.85-94, 2014 .

[15] Zachary, Wayne W. "An information flow model for conflict and fission in small groups." Journal of anthropological research, pp. 452-473, 1977.

[16] Huang, Jing, et al. "Decentralized mining social network communities with agents." Mathematical and Computer Modelling, vol 57.11, pp. 2998-3008, 2013.

[17] Rhouma, Delel, and Lotfi Ben Romdhane. "An efficient algorithm for community mining with overlap in social networks." Expert Systems with Applications, vol 41.9, pp. 4309-4321, 2014.

[18] Xie, Jierui, Stephen Kelley, and Boleslaw K. Szymanski. "Overlapping community detection in networks: The state-of-the-art and comparative study", ACM Computing Surveys (CSUR), vol 45.4, pp. 43, 2013.

[19] Palla, Gergely, et al. "Uncovering the overlapping community structure of complex networks in nature and society." Nature 435.7043 (2005): 814-818.

[20] Farkas, Illés, et al. "Weighted network modules." New Journal of Physics, vol 9.6, pp. 180, 2007.

[21] Li, Junqiu, Xingyuan Wang, and Yaozu Cui. "Uncovering the overlapping community structure of complex networks by maximal cliques." Physica A: Statistical Mechanics and its Applications, vol 415, pp. 398-406, 2014.

[22] Zhang, Zhiwei, and Zhenyu Wang. "Mining overlapping and hierarchical communities in complex networks." Physica A: Statistical Mechanics and its Applications, vol 421, pp. 25-33, 2015.

[23] Cui, Yaozu, Xingyuan Wang, and Junqiu Li. "Detecting overlapping communities in networks using the maximal sub-graph and the clustering coefficient." Physica A: Statistical Mechanics and its Applications, vol 405, pp. 85-91, 2014.

[24] Huang, LAN, et al. "Link Clustering with Extended Link Similarity and EQ Evaluation Division." PloS one, vol 8.6, pp. e66005, 2013.

[25] Ahn, Yong-Yeol, James P. Bagrow, and Sune Lehmann. "Link communities reveal multi scale complexity in networks." Nature, vol 466.7307, pp. 761-764, 2010.

[26] Evans, T. S., and R. Lambiotte. "Line graphs of weighted networks for overlapping communities." The European Physical Journal B-Condensed Matter and Complex Systems, vol 77.2, pp. 265-272, 2010.

[27] Evans, T. S., and R. Lambiotte. "Line graphs, link partitions, and overlapping communities." Physical Review E, vol 80.1, pp. 016105, 2009.

[28] Kim, Paul, and Sangwook Kim. "Detecting overlapping and hierarchical communities in complex network using interaction-based edge clustering." Physica A: Statistical Mechanics and its Applications, vol 417, pp. 46-56., 2015.

[29] Shi, Chuan, et al. "A link clustering based overlapping community detection algorithm." Data & Knowledge Engineering, vol 87, pp. 394-404, 2013 .

[30] Cui, Yaozu, and Xingyuan 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.

[31] Jin, Di, Bogdan Gabrys, and Jianwu Dang. "Combined node and link partitions method for finding overlapping communities in complex networks." Scientific reports, vol 5, 2015.

[32] Lancichinetti, Andrea, Santo Fortunato, and János Kertész. "Detecting the overlapping and hierarchical community structure in complex networks." New Journal of Physics, vol 11.3, pp. 033015, 2009 .

[33] Havemann, Frank, et al. "Identification of overlapping communities and their hierarchy by locally calculating community-changing resolution levels." Journal of Statistical Mechanics: Theory and Experiment, vol 2011.01, pp. P01023, 2011.

[34] Chen, Duanbing, et al. "Detecting overlapping communities of weighted networks via a local algorithm." Physica A: Statistical Mechanics and its Applications, vol 389.19, pp. 4177-4187, 2010 .

[35] Li, Junqiu, Xingyuan Wang, and Justine Eustace. "Detecting overlapping communities by seed community in weighted complex networks." Physica A: Statistical Mechanics and its Applications, vol 392.23, pp. 6125-6134, 2013 .

[36] Wu, Zhihao, et al. "Efficient overlapping community detection in huge real-world networks." Physica A: Statistical Mechanics and its Applications, vol 391.7, pp. 2475-2490, 2012 .

[37] Nepusz, Tamás, et al. "Fuzzy communities and the concept of bridgeness in complex networks." Physical Review E, vol 77. , pp. 016107, 2008 .

[38] Zhang, Shihua, Rui-Sheng Wang, and Xiang-Sun Zhang. "Identification of overlapping community structure in complex networks using fuzzy c-means clustering." Physica A: Statistical Mechanics and its Applications, vol 374.1, pp. 483-490, 2007.

[39] Psorakis, Ioannis, et al. "Overlapping community detection using bayesian non-negative matrix factorization." Physical Review E, vol 83.6, pp. 066114, 2011.

[40] Eustace, Justine, Xingyuan Wang, and Yaozu Cui. "Overlapping community detection using neighborhood ratio matrix." Physica A: Statistical Mechanics and its Applications, vol 421, pp. 510-521, 2015.

[41] Sun, Peng Gang, Lin GAO, and Shan Shan Han. "Identification of overlapping and non-overlapping community structure by fuzzy clustering in complex networks." Information Sciences, vol 181.6, pp. 1060-1071, 2011.

[42] Gregory, Steve. "Finding overlapping communities in networks by label propagation." New Journal of Physics, vol 12.10, pp. 103018, 2010 .

[43] Badie, Reza, et al. "An efficient agent-based algorithm for overlapping community detection using nodes' closeness." Physica A: Statistical Mechanics and its Applications, vol 392.20, pp. 5231-5247, 2013.

[44] Zhou, Xu, et al. "An ant colony based algorithm for overlapping community detection in complex networks." Physica A: Statistical Mechanics and its Applications, vol 427, pp. 289-301, 2015.

[45] Yu, Xin, Jing Yang, and Zhi-Qiang Xie. "A semantic overlapping community detection algorithm based on field sampling." Expert Systems with Applications, vol 42.1, pp. 366-375, 2015.

[46] Xin, Yu, et al. "An overlapping semantic community detection algorithm base on the ARTs multiple sampling models." Expert Systems with Applications, vol 42.7, pp. 3420-3432, 2015.

Zhang, Xuewu, et al. "Overlapping community identification approach in online social networks." Physica A: Statistical Mechanics and its Applications 421 (2015): 233-248.

Archive of SID

www.SID.ir