4th International Conference on Web Research
April 25-26, 2018; Tehran, IRAN
ICWR2018
کنفرانس بین المللی وب پژوهی
چهارمین
۵ و ۶ اردیبهشت ۱۳۹۷؛ تهران، ایران

Network Nodes Community Detection Based on Deep Learning

[1] Mohammad Mehdi Keikha, [2]Maseud Rahgozar

[1] PhD Candidate, University of Tehran, Tehran, Iran
Mehdi.keikha@ut.ac.ir

[1] Faculty member, University of Sistan and Baluchestan, Zahedan, Iran

[2] Associate Professor, University of Tehran, Tehran, Iran
rahgozar@ut.ac.ir

Abstract

Network clustering is one of the problems that has attracted many researchers in recent years. In this issue, each user is associated with a specific community based on the various features of the network, including the structure. In the recent years, deep learning is widely used to extract the feature vector of nodes then the vectors are used to find the community of each node. In this paper, a network representation learning algorithm is presented based on the information of the neighbors of each node and communities on the network. The results show that our nodes' representation method offers a better quality clustering of social networking users than the previous network representation learning methods.

Keywords: Network Clustering, Deep Learning, Node Representation Vector, Network Communities.

Archive of SID

www.SID.ir

4th International Conference on Web Research
April 25-26, 2018; Tehran, IRAN
ICWR2018
کنفرانس بین المللی وب پژوهی
چهارمین
۵ و ۶ اردیبهشت ۱۳۹۷؛ تهران، ایران

خوشه بندی گره های شبکه با استفاده از یادگیری عمیق

محمد مهدی کیخا[۱]، مسعود رهگذر[۲]

[۱] دانشجوی دکترا، دانشکده مهندسی برق و کامپیوتر ، دانشگاه تهران
تهران، ایران
عضو هیئت علمی، دانشگاه سیستان و بلوچستان
زاهدان، ایران
mehdi.keikha@ut.ac.ir

[۲] دانشیار، دانشکده مهندسی برق و کامپیوتر ، دانشگاه تهران
تهران، ایران
rahgozar@ut.ac.ir

Archive of SID

چکیده

خوشه بندی گره های شبکه، یکی از مسائلی است که در سال‌های اخیر مورد توجه بسیاری از محققین قرار گرفته است. در این مساله هر کاربر بر اساس ویژگی های مختلف موجود در شبکه از جمله ساختار، در یک جامعه خاص قرار می گیرد. یکی از جدیدترین رویکردهای این مسئله، استفاده از یادگیری عمیق جهت استخراج بردار ویژگی های هر گره و یافتن جامعه هر گره بر اساس بردار ویژگی آنهاست. در این مقاله روشی برای یادگیری بردار نمایش گره های شبکه بر اساس اطلاعات همسایه¬های هر گره و جوامع مختلف موجود در شبکه ارائه شده است. نتایج بدست آمده حاکی از آن است که روش بردار نمایش ارائه شده در این مقاله در مقایسه با سایر روش های نمایش ساختار شبکه، خوشه بندی با کیفیت تری از افراد موجود در شبکه اجتماعی ارائه می نماید.

کلمات کلیدی

خوشه بندی شبکه، یادگیری عمیق، بردار نمایش گره، جوامع گراف.

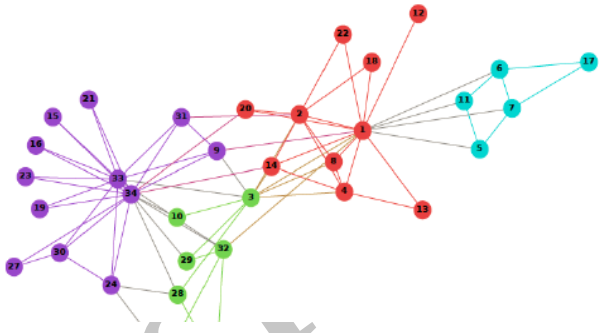
۱- مقدمه

خوشه بندی یکی از مسائل بسیار مهم و قدیمی یادگیری ماشین است که سعی در شناسایی گروه های مختف یک شبکه اجتماعی دارد. روش های بسیاری برای خوشه بندی شبکه های اجتماعی بر اساس ارتباطات موجود [1] و یا به کمک اطلاعات منتشر شده توسط افراد در یک شبکه[2] ، ارائه شده است. مهمترین چالش موجود در مساله خوشه بندی گره ها شبکه، نحوه استخراج و استفاده از اطلاعات نهفته در آن است. در مساله خوشه بندی باید همواره گره ها در خوشه ها به نحوی دسته بندی شوند که گره های موجود در داخل یک خوشه بیشترین ارتباط را با گره های درون همان خوشه داشته باشند و ارتباط کمتری با گره های سایر خوشه ها داشته باشند.

از جمله مهمترین روش های ارائه شده برای مساله خوشه بندی استفاده از یادگیری عمیق است که در سال های اخیر بسیار مورد توجه قرار گرفته است. روش های مبتنی بر یادگیری عمیق یکی از روش های بسیار موثر کاهش ابعاد و استخراج ویژگی ها می باشند که در سال های اخیر در کاربردهای گوناگون بسیار مورد استفاده قرار گرفته است[3, 4]. در واقع روش های مبتنی بر یادگیری عمیق تلاش می نمایند که یک نگاشت از گره های شبکه بیابند که نشان دهنده تمام ویژگی های ساختاری گراف اصلی باشد. این روش ها ابتدا گراف ارتباطات نمونه ها را به عنوان ورودی دریافت می

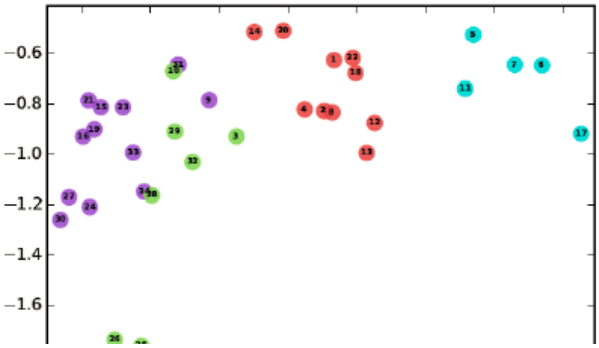
www.SID.ir

نمایند و سپس با استخراج ویژگی های مختلف توسط یادگیری عمیق، سعی در یادگیری بردار نمایش گره های شبکه می نمایند. هنگامی که هر گره شبکه توسط یک بردار با ابعاد بسیار کمتر نمایش داده شد، از این بردارها جهت حل مساله خوشه بندی گره ها استفاده می نمایند[5, 6, 7, 8]. در شکل ۱ گراف شبکه کاراته مشاهده می شود. در این گراف، گره هایی که متعلق به یک خوشه هستند توسط یک رنگ نمایش داده شده اند.



شکل ۱ – گراف کاراته به همراه خوشه ها[5]

در شکل ۲ هر گره توسط یک نقطه در فضای دو بعدی به نمایش در آمده است که مختصات این نقاط همان بردار ویژگی گره ها می باشد که توسط یادگیری عمیق استخراج شده است.



شکل ۲- نمایش گره های شبکه در یک فضای دو بعدی جدید[5]

همانگونه که در شکل ۲ مشاهده می نمایید، گره هایی که در شکل ۱ در یک خوشه قرار داشته اند، در فضای جدید نیز در نزدیکی یکدیگر قرار گرفته اند. حتی گره هایی که در دو خوشه مختلف قرار دارند نیز در فضای جدید ممکن است نزدیک یکدیگر قرار گیرند زیرا مسیرهایی با طول کم بین این گره ها ممکن است وجود داشته باشد.

در این مقاله، با استفاده از یک روش بردار نمایش گره های شبکه، یک روش جدید برای حل مساله خوشه بندی افراد در شبکه ارائه شده است. در روش ارائه شده هم از اطلاعات ساختار محلی یک گره مانند همسایه¬های مستقیم در شبکه و هم از اطلاعات ساختاری کلی شبکه استفاده شده است. روش ارائه شده در این مقاله که CARE[1] نامیده می شود ابتدا توسط یکی از روش های بسیار ساده تعدادی خوشه برای گراف ایجاد می نماید. پس از آن اقدام به تولید تعدادی مسیر ویژه برای هر گره در گراف می نماید و در نهایت از این مسیرهای ویژه برای یادگیری بردار ویژگی گره ها استفاده می نماید. همانگونه که پیش از این نیز به آن اشاره شده مهمترین بخش از روش پیشنهادی نحوه ایجاد مسیرهای ویژه است. ادامه ی مقاله به صورت زیر سازماندهی شده است:

در بخش ۲ به مرور کارهای انجام شده جهت حل مساله خوشه بندی گره ها توسط روش های یادگیری بردار نمایش گره های شبکه خواهیم پرداخت. روش پیشنهادی برای حل مساله خوشه بندی گره ها در بخش سوم و نتایج آزمایشات صورت گرفته و تحلیل آن‌ها در بخش چهارم ارائه شده است.. در نهایت در بخش ۵ نیز به نتیجه گیری و ارائه مسیرهای تحقیقاتی مرتبط با این مساله می پردازیم.

۲- تحقیقات پیشین

خوشه بندی گراف یکی از مسائل بسیار قدیمی در حوزه تحلیل گراف محسوب می شود. نخستین روش هایی که برای خوشه بندی شبکه ارائه شد، مبتنی بر شناسایی خوشه‌های گراف توسط معیارهای مرکزیت در یک شبکه اجتماعی بود [9]. هستینگ و همکاران خوشه بندی گراف را به عنوان یک مساله استقراء در نظر گرفتند که در آن به دنبال بهترین ترتیب از خوشه‌های گراف هستیم و این مساله را به کمک انتشار عقاید[2] حل نموده اند [10]. در سال های اخیر، شبکه های عصبی عمیق مانند انواع اتوانکدرها[3] در مساله خوشه بندی گراف بسیار مورد استفاده قرار گرفته اند که از آن جمله می توان به اتوانکدر خلوت[4] [11] و اتوانکدرهای متغیر[5] [12] اشاره نمود.

دسته ای دیگر از روش های خوشه بندی گراف ابتدا یک نمایش جدید از گراف را با در نظر گرفتن اطلاعات ساختاری موجود در آن استخراج می نمایند و بردار نمایش هر گره از گراف را به عنوان بردار ویژگی آن در نظر می گیرند. سپس خوشه های گراف را بر مبنای نزدیکی بردار ویژگی گره ها، استخراج می نمایند. در واقع این روش ها به کمک یادگیری بردار نمایش گره های گراف، اقدام به شناسایی خوشه های آن می نمایند. از آنجایی که روش ارائه شده در این مقاله نیز جزء این روش‌ها می باشد، در ادامه به بررسی برخی از مهمترین روش های یادگیری بردار نمایش گره ها می پردازیم.

الگوریتم Deepwalk یکی از اولین تحقیقات صورت گرفته در زمینه نمایش گراف توسط شبکه های عصبی عمیق می باشد [5] . در این روش از قدم زنی های تصادفی متوالی به عنوان دانش زمینه ای برای یافتن بردار نمایش یک گره استفاده می گردد. در این روش تنها از قدم زنی های تصادفی در گراف استفاده می گردد و اطلاعات خوشه های گراف توسط Deepwalk پوشش داده نمی شود، در نتیجه قادر نیست تمامی اطلاعات ساختاری گراف را منعکس نماید و در نتیجه یک نمایش کامل برای تمامی گره های گراف ارائه نخواهد شد. از جمله دیگر معایب این روش می توان به عدم کارایی آن برای گراف های خلوت اشاره نمود.

در [6] از همسایه¬های مستقیم[6] و درجه دوم (همسایه همسایه)[7] گره های گراف برای استخراج بردار نمایش گره های آنها استفاده می شود. روش ارائه شده در این مقاله اگرچه توانسته است خاصیت غیر خطی بودن و خلوت بودن گراف های واقعی را در نظر بگیرد اما ارتباط هر گره را فقط محدود به همسایه های نزدیک به آن نموده است حال آنکه دو گره می توانند عضو یک جامعه باشند اما در گراف فاصله بیشتر از ۲ با یکدیگر داشته باشند. در واقع این مقاله، هر دو گرهی که با یکدیگر ارتباط مستقیم داشته باشند یا چندین همسایه مشترک داشته باشند، را به عنوان دو گره مشابه از نظر ساختار در نظر می گیرد و تنها از این اطلاعات جهت ایجاد بردار نمایش هر گره استفاده می

Archive of SID
www.SID.ir

نماید. حال آنکه ممکن است دو گره در قسمت های مختلف گراف حضور داشته باشند اما از لحاظ نقش ساختاری و یا همویلی (جوامع) مشابه یکدیگر باشند.

روش Node2vec که به تازگی توسط لسکوک[7] ارائه شده است از قدم زنی های تصادفی استفاده نموده است. این روش نیز اگرچه مقیاس پذیری بالایی دارد و از دو استراتژی جستجوی BFS[۸] و DFS[۹] برای قدم زنی های تصادفی استفاده می نماید، اما نمی تواند تمامی اطلاعات ساختاری محلی و کلی گراف را حفظ نماید زیرا فقط از روابط همسایگی مستقیم و روابط همسایه درجه دوم استفاده می نماید و از اطلاعات خوشه های گراف به عنوان اطلاعات ساختاری کلی گره ها استفاده نمی کند. به همین دلیل برای نمایش گره‌ها، اطلاعات زیادی در رابطه با ساختار کلی گراف را در نظر نمی-گیرد، همچنین این روش نیز مانند Deepwalk برای گراف‌های خلوت کارایی بالایی ندارد.

اخیرا Wang و همکاران در [8] از روش های فاکتورگیری ماتریسی جهت یادگیری بردار نمایش گره ها استفاده نموده‌اند. آنها دو تابع مختلف جهت استخراج اطلاعات جوامع و همسایه¬های تعریف می نمایند که این دو تابع به طور مستقل از یکدیگر بردار نمایش یک گره را می سازند. از آنجایی که دو تابع به طور مستقل از یکدیگر عمل می کنند و عملیات روی ماتریس ها در فاکتورگیری ماتریسی بسیار هزینه بر است، این روش مقیاس پذیری بالایی نداشته و برای گراف های بزرگ اصلا کارایی ندارد.

۳- روش پیشنهادی

همانگونه که در بخش قبل مشاهده نمودیم، تمامی روش های یادگیری بردار نمایش گره های شبکه به کمک یادگیری عمیق، ابتدا تعدادی از گره های مرتبط با هر گره را توسط استراتژی های مختلف استخراج می نمایند و سپس از این گره ها به عنوان دانش زمینه برای یادگیری بردار نمایش هر گره استفاده می کنند. تفاوت روش های مختلف ارائه شده، در نحوه استخراج گره های مرتبط با یک گره خاص می‌باشد. اکثر روش های ارائه شده فقط از همسایه¬های مستقیم و همسایه¬های درجه دوم به عنوان گره های مرتبط استفاده می نمایند. حال آنکه روش پیشنهادی ارائه شده که به اختصار CARE نامیده می شود، ابتدا برای هر گره تعدادی مسیر ویژه[۱۰] متشکل از گره های مرتبط با آن ایجاد می نماید. هنگام تشکیل مسیرها، از اطلاعات جوامع موجود در گراف نیز استفاده می گردد و بر اساس یک پارامتر، در مسیر ایجاد شده برای هر گره از برخی گره ها که در همسایگی گره نمی باشند اما با گره مورد نظر در یک جامعه قرار دارند، نیز به عنوان دانش زمینه ای برای یادگیری نمایش برداری آن گره استفاده می شود. با توجه به اینکه اطلاعات جوامع یک گراف، حاوی ساختار کلی گراف است و اکثر روش های پیشین از این اطلاعات هیچگونه استفاده ای ننموده اند، استفاده از این اطلاعات در هنگام یادگیری بردار نمایش گره های شبکه می تواند نتایج خوبی داشته باشد.

در ابتدا برای تعیین جوامع کلی گراف، از الگوریتم لوین برای تشخیص جوامع مختلف شبکه ورودی استفاده شده است[13] . پس از اینکه جوامع مختلف گراف ورودی شناسایی شد، اقدام به تولید مسیرهای ویژه ای می شود که علاوه بر اطلاعات همسایه¬های یک گره، اطلاعات گره هایی که با گره مبدا مسیر در یک جامعه قرار دارند را نیز شامل شود. برای تولید مسیرها، برای هر گره مشاهده شده در طی روند ایجاد مسیر، با احتمال α یکی از همسایه¬های مستقیم گره مشاهده شده را انتخاب و با احتمال 1-α نیز یکی از گره هایی برگزیده می شود که با گره مشاهده شده در یک جامعه قرار گرفته باشد. به این ترتیب CARE قادر خواهد بود هم ساختار محلی یک گره و هم ساختار کلی گراف را هنگام استخراج ویژگی های یک گره مد نظر قرار دهد. برای یادگیری بردار نمایش هر گره با استفاده از مسیرهای ویژه ایجاد شده از الگوریتم Skip-Gram استفاده می شود که پیش از این در پردازش زبان طبیعی از آن استفاده شده است [۱۸].

ایده اصلی الگوریتم skip-gram این است که کلماتی که در جملاتی بیشتری کنار یکدیگر قرار میگیرند، از لحاظ معنایی به یکدیگر مرتبط هستند. حال اگر الگوریتم skip-gram را در شبکه اعمال نماییم، گره هایی که در مسیرهای بیشتری در نزدیکی یکدیگر قرار گرفته اند، احتمالا از نظر ویژگی های ساختاری به یکدیگر مشابهت بیشتری دارند که این خاصیت باید در بردار نمایش نهایی گره ها نیز وجود داشته باشد.

الگوریتم ۱ مراحل انجام این روش را تشریح می نماید:

Algorithm 1: CARE (G, w, d, μ, l)

Input:
Graph G (V, E)
Window length w
Representation size d
Number of random walks per node μ
Random walk max length l

Output:
Matrix of node representations $f \in \mathbb{R}^{|V| \times d}$

1: Com = ModularityMaximization(G)
2: sample f from $\mathbb{R}^{|V| \times d}$
3: while (i < μ)
4: s = shuffle (V)
5: for each $v_i \in s$ do
6: $\mathcal{W}_{v_i}$ = CommunityawareRW (G, v_i , Com, l)
7: SkipGram (f, $\mathcal{W}_{v_i}$, w)
8: end for
9: end while

در الگوریتم ۱، G گراف ورودی است و w پنجره مفهومی در نظر گرفته شده برای هر گره می باشد. در این الگوریتم، d نشان دهنده طول بردار ویژگی استخراج شده برای هر گره می باشد و l طول مسیر تصادفی ایجاد شده است. همانگونه که در الگوریتم CARE مشاهده می شود، در گام ۱ ابتدا جوامع مختلف گراف ورودی تشخیص داده می شوند. سپس در گام ۲ بردار نمایش هر گره مقدار دهی اولیه می شود. سپس برای هر گره در شبکه به تعداد μ مسیر ویژه ایجاد می گردد تا اطمینان حاصل شود که تمامی اطلاعات ساختار محلی و ساختار کلی گراف برای هر گره حفظ خواهد شد. برای جلوگیری از وابستگی روش به ترتیب مشاهده گره های شبکه، در گام ۴ ترتیب آنها بصورت تصادفی تغییر داده می شود. در گام ۶ به ازای هر گره یک مسیر تصادفی ایجاد می شود که این مسیر حاوی اطلاعات همسایه¬های و اطلاعات گره هایی است که با گره های مشاهده شده در یک جامعه قرار دارند. سپس مسیر ایجاد شده برای یک گره به الگوریتم skip-gram تحویل داده می شود تا به کمک این روش، بردار نمایش گره‌ها یادگرفته شود.

Archive of SID
www.SID.ir

پس از آنکه بردار نمایش تمامی گره ها توسط الگوریتم CARE استخراج شد، هر کدام از این بردارها به عنوان یک نقطه در فضای n بعدی جدید در نظر گرفته می شود که n تعداد ویژگی های اتخراج شده برای هر یک از گره ها است. در انتها برای استخراج خوشه های نهایی گراف، از الگوریتم k همسایه نزدیک استفاده شده است. در این الگوریتم ابتدا فاصله دو گره از شبکه توسط محاسبه فاصله بردار ویژگی‌های آنها محاسبه می شود و در نهایت گره هایی که نزدیکترین فاصله با یکدیگر را دارند به عنوان یک خوشه در نظر می گیریم. در بخش ۴ به تشریح آزمایشات انجام شده خواهیم پرداخت.

٤- نتایج عملی و ارزیابی

جهت ارزیابی روش پیشنهادی برای مساله خوشه بندی گراف، از چند الگوریتم بردار نمایش گره های شبکه که پیش از این برای مسائل دیگر از جمله دسته بندی چند برچسبی[11] مورد استفاده قرار گرفته اند، استفاده شده است که این الگوریتم ها عبارتند از:

الگوریتم Spectral clustering [14] : این الگوریتم با استفاده از برش های مختلف گراف، سعی در شناسایی جوامع مختلف یک گره می کند. برای یافتن برش های مختلف گراف، ابتدا ماتریس لاپلاسین نرمال شده[12] گراف را محاسبه می نماید و سپس d کوچکترین مقدار بردار ویژه را به عنوان بردار نمایش هر گره انتخاب می نماید.

الگوریتم DeepWalk [5] : این روش نخستین الگوریتمی است که از بردار نمایش گره ها که توسط یادگیری عمیق بدست می آورد، برای حل مساله دسته بندی چند برچسبی استفاده می نماید. این روش در ابتدا به ازای هر گره، تعدادی قدم زنی تصادفی ایجاد می کند و از این مسیرهای تصادفی به عنوان اطلاعات زمینه برای یادگیری بردار نمایش گره ها در فضای d بعدی استفاده می کند.

الگوریتم LINE [6] : این روش دو تابع جداگانه برای جمع آوری اطلاعات همسایه¬های مستقیم و همسایه¬های درجه دوم گره استفاده می کند و از اطلاعات جمع آوری شده توسط این دو تابع جهت یادگیری بردار نمایش گرهه ها استفاده می کند.

الگوریتم Node2vec [7] : این روش نیز مانند DeepWalk از قدم زنی های تصادفی برای جمع آوری اطلاعات گره ها استفاده می نماید. تفاوت اصلی آن با DeepWalk در این است که این روش فقط از همسایه¬های مستقیم و همسایه¬های درجه دوم گره به منظور ایجاد مسیرهای تصادفی استفاده می کند.

برای ارزیابی کامل CARE از دو مجموعه داده استاندارد استفاده شده است که اطلاعات آنها در جدول زیر ارائه شده است:

جدول ۱: مشخصات مجموعه داده های مورد استفاده در آزمایشات

نام مجموعه داده	تعداد گره ها	تعداد یالها
Protein-Protein Interactions (PPI)	3890	76584
Wikipedia	4777	184812

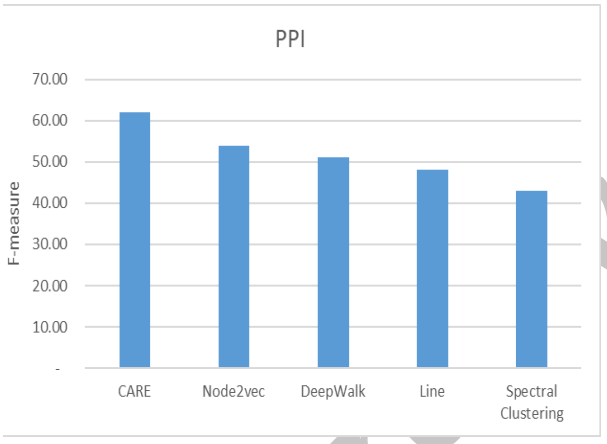
جهت مقایسه CARE با الگوریتم های ذکر شده، پس از آنکه بردار نمایش هر گره در هر کدام از مجموعه داده های ذکر شده توسط یکی از روش‌ها استخراج شد، هر نقطه را به عنوان یک نقطه در فضای جدید در نظر می گیریم و این نقاط را توسط الگوریتم k نزدیکترین همسایه خوشه بندی می نماییم. برای مقایسه میزان کارایی روش های مختلف خوشه بندی از معیار f-measure استفاده می شود که بر اساس روابط ۱ تا ۳ محاسبه می شود:

$$Precision = \frac{\sum_{A \in C} TP\ (A)}{\sum_{A \in C}(TP(A) + FP(A))} \quad (1)$$

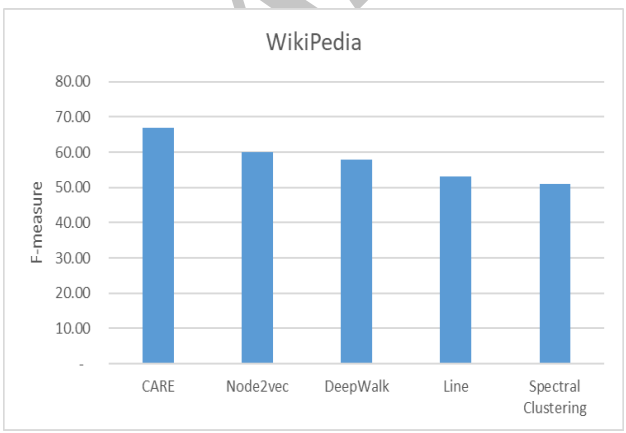
$$Recall = \frac{\sum_{A \in C} TP\ (A)}{\sum_{A \in C}(TP(A) + FN(A))} \quad (2)$$

$$f - measure = \frac{2 * Precision * Recall}{Precision + Recall} \quad (3)$$

در ادامه مقدار f-measure روش های ذکر شده در مجموعه داده های مختلف در شکل های ۳ و ۴ نمایش داده شده است.



شکل ۳- مقایسه روش های مختلف برای خوشه بندی توسط PPI



شکل ٤- مقایسه روش های مختلف برای خوشه بندی توسط WikiPedia

همانگونه که مشاهده می گردد، در شکل ۳ روش پیشنهادی حدود ۲۰ درصد نسبت به بهترین روش یعنی Node2vec بهبود عملکرد داشته است زیرا قادر است نمایش بهتری از گره های گراف ایجاد نماید. در شکل ۴ نیز نتایج مشابهی حاصل شده است. یکی دیگر از دلایل بهبودی عملکرد

Archive of SID
www.SID.ir

CARE در شکل های ۳ و ۴ این است که این روش قادر است برای گراف های با چگالی پایین یال نیز نتایج خوبی ارائه دهد زیرا چگالی گراف PPI بسیار کمتر از گراف Wikipedia است. در نتیجه یکی از مهمترین ویژگی های الگوریتم CARE کارایی آن برای گراف های خلوت است که اتفاقا بسیاری از شبکه های دنیای واقعی نیز خلوت هستند و ضریب بخش بندی[۱۳] پایینی دارند.

با توجه به اینکه تفاوت روش های مختلف یادگیری نمایش گره ها، در استراتژی جستجو در گراف به منظور استخراج اطلاعات زمینه ای است، همانطور که در شکل های ۳ و ۴ به وضوح مشخص است استراتژی جستجو در الگوریتم CARE بسیار مناسب تر از سایر روش های یادگیری بردار نمایش گره ها است زیرا قادر است تمامی اطلاعات ساختاری محلی و سراسری گراف را در فضای جدید نیز حفظ نماید.

۵- نتیجه‌گیری

در این مقاله روشی برای حل مساله دسته بندی خوشه بندی گراف مبتنی بر یادگیری بردار نمایش گره های شبکه ارائه شده است. بر خلاف روش های قبلی که فقط ساختار محلی گره ها به عنوان ویژگی های هر گره در هنگام نمایش ساختار مورد استفاده قرار می گیرفت، روش پیشنهادی هنگام یادگیری بردار نمایش گره های شبکه، هم ساختار محلی شامل همسایه‌ها گره ها و هم ساختار کلی گراف شامل جوامع موجود در شبکه را در نظر می گیرد. نتایج ارزیابی های انجام شده بر روی دو مجموعه داده استاندارد نشان از عملکرد مطلوب روش پیشنهادی در مقایسه با سایر روش های نمایش برداری برای حل مساله خوشه بندی گراف است.

مراجع

[1] Luxburg, U., "*A tutorial on spectral clustering*", Statistical Computing, Vol. 17, No. 4, pp. 395–426, 2007.
[2] Sun, Y., Han, J., Gao, J., and Yu, Y., "iTopicModel: Information network integrated topic modeling", In Proc. 9th IEEE Int. Conf. Data Mining, pp. 493–502, 2009.
[3] Bengio, Y., Courville, A., and Vincent, P., "*Representation learning: A review and new perspectives*" IEEE TPAMI, Vol. 35, No. 8, pp.1798–1828, 2013.
[4] Yu, X., Ren, X., Sun, Y., Gu, Q., Sturt, B., Khandelwal, U., Norick, B., and Han, J., "Personalized entity recommendation: A heterogeneous information network approach", In WSDM, pp. 283-292, 2014.
[5] Perozzi, B., Al-Rfou, R., and Skiena, S., "DeepWalk: Online learning of social representations", In KDD, 2014.
[6] Tang, J., Qu, M., Wang, M., Zhang, M., Yan, J., and Mei, Q., "LINE: Large-scale Information Network Embedding", In WWW, 2015.
[7] Grover, A., and Leskovec, J., "Node2vec: Scalable feature learning for networks", In Proc. of the 22nd ACM SIGKDD, pp. 1225–1234, 2016.
[8] Wang, X., Cui, P., Wang, J., Pei, J., Zhu, W., and Yang, S., "Community preserving network embedding", in AAAI, pp. 203–209, 2017.
[9] Girvan, M., and Newman, M. E, "Community structure in social and biological networks", in Proc. of NAS, Vol. 12, pp. 7821–7826, 2002.
[10] Hew, M., and Hastings, B., "*Community detection as an inference problem*", Physical Review E, Vol. 74, No. 3, 2006.
[11] Hu, P., Chan, K. C., and He, T., "Deep Graph Clustering in Social Network" In Proc. of WWW, pp. 1425–1426, 2017.
[12] Kipf, O. N., and Welling, M., "*Variational Graph Auto-Encoders*", arXiv preprint arXiv:1611.07308, 2016.
[13] D-Blondel, V., Guillaume, J., Lambiotte, R., and Lefebvre, E., "*Fast unfolding of communities in large networks*" Journal of Statistical Mechanics: Theory and Experiment, Vol. 10, 2008.
[14] Breitkreutz, B. J., Stark, C., Reguly, T., Boucher, L., Breitkreutz, A., Livstone, M., Oughtred, R., Lackner, D. H., Bähler, J., Wood, V., "*The BioGRID interaction database*" Nucleic acids research, Vol. 36, pp. 637–640, 2008.

زیرنویس

[1] Community aware random walk for network embedding
[2] Belief propagation
[3] Auto encoders
[4] Sparse auto encoders
[5] Variational auto encoders
[6] First order proximity
[7] Second order proximity
[8] Breadth first search
[9] Depth first search
[10] Customized path
[11] Multi label classification
[12] Normalized laplacian matrix
[13] Clustering coefficient

Archive of SID
www.SID.ir