

A Similarity Measure for Link Prediction in Social Network

Ali Sarabadani^{1*}, Kheirollah RahseparFard², Seyyed Morteza Pournaghi³

1. PhD Student, IT Engineering, University of Qom, Qom, Iran. *Corresponding Author*,
alisarabadani14@gmail.com

2. Assistant Professor, Department of IT & Computer Engineering, University of Qom, Qom, Iran.

3. Assistant Professor, Faculty of Computer, Network and Communication, Imam Hussein University, Tehran, Iran.

Abstract

Introduction: A social network is a social structure made up of individuals or organizations. Social network analysis is an approach in which the network is considered as a set of nodes and relationships between them. Nodes are individuals and actually actors in the network and the relationships between them are displayed as connections between nodes.

Method: Among many social network analysis issues, link prediction has attracted much attention due to the growing number of social network users. Link prediction means predicting which new interaction is going to happen in the future. Traditional link prediction methods considered pairs of nodes as a unit and made decisions based on commonalities between them. In addition, we proposed a new similarity measure for link prediction in social networks.

Results: We compared this criterion with four prediction methods of Jaccard link, Salton Index, Salton Cosine, and resource allocation). Experimental runs in this article were carried out on five social network datasets. Our results showed that this criterion performed better than other link prediction techniques on all datasets.

Discussion: Social network analysis has recently attracted lots of attention among researchers due to its wide applicability in capturing social interactions. Link prediction, related to the likelihood of having a link between two nodes of the network that are not connected, is a key problem in social network analysis. Many methods have been proposed to solve the problem. Among these methods, similarity-based methods exhibit good efficiency by considering the network structure and using as a fundamental criterion the number of common neighbors between two nodes to establish structural similarity.

Keywords: Social network analysis, link prediction, Similarity measure, Co-authorship network.

ارائه یک معیار مشابهت جهت پیش‌بینی پیوند در شبکه‌های اجتماعی

دوره چهارم، بهار ۱۴۰۲
شماره اول، صص: ۱۷-۲۶

تاریخ دریافت: ۱۴۰۱/۱۲/۰۴
تاریخ پذیرش: ۱۴۰۲/۰۱/۳۱

علی سرآبادانی^{۱*}، خیرالله رهسپار فرد^۲، سید مرتضی پورنقی^۳

۱- دانشجوی دکتری، مهندسی فناوری اطلاعات، دانشگاه قم، قم، ایران. (نویسنده مسئول) alisarabadani14@gmail.com

۲- استادیار، گروه مهندسی کامپیوتر و فناوری اطلاعات، دانشکده فنی و مهندسی، دانشگاه قم، قم، ایران.

۳- استادیار، گروه رایانش امن، دانشکده رایانه، شبکه و ارتباطات، دانشگاه جامع امام حسین (ع)، تهران، ایران.

چکیده: یک شبکه اجتماعی، ساختاری متشکل از افراد یا سازمان‌ها است. تحلیل شبکه‌های اجتماعی، مبتنی بر رویکردی است که در آن شبکه را به صورت مجموعه‌ای از گره‌ها و روابط میان آن‌ها در نظر می‌گیرد. گره‌ها شامل اشخاص و موجودیت‌های درون شبکه هستند که با یکدیگر در تعاملند و در واقع بازیگران درون شبکه محسوب می‌شوند که روابط میان آن‌ها به صورت اتصالاتی بین گره‌ها نمایش داده می‌شود. باتوجه به تعداد روبه‌شد کاربران شبکه‌های اجتماعی، تحلیل روابط حاکم بر آن، پیش‌بینی پیوندها و تعامل‌های ناشی از ارتباط میان گره‌ها (پیش‌بینی لینک یا پیوند، یعنی پیش‌بینی تعامل جدیدی که قرار است در آینده رخ دهد) از چالش‌های جدی در شبکه‌های اجتماعی می‌باشد. ما در این مقاله یک معیار شباهت جدید برای پیش‌بینی لینک در شبکه‌های اجتماعی را پیشنهاد می‌دهیم. این معیار را با چهار روش پیش‌بینی لینک Salton Cosine, Salton Index, Jaccard و Resource Allocation مقایسه می‌کنیم. ما شبیه‌سازی معیار پیشنهادی خود را بر روی پنج مجموعه داده در شبکه‌های اجتماعی، انجام می‌دهیم. نتایج شبیه‌سازی نشان می‌دهد که معیار پیشنهادی ما عملکرد بهتری نسبت به سایر روش‌های پیش‌بینی پیوند در شبکه‌های اجتماعی بر روی همه دیتاست‌ها را دارد. تجزیه و تحلیل شبکه‌های اجتماعی به دلیل کاربرد گسترده آن در ثبت تعاملات اجتماعی اخیراً توجه بسیاری را در بین محققان به خود جلب کرده‌است. پیش‌بینی پیوند، مربوط به احتمال وجود پیوند بین دو گره شبکه که متصل نیستند، یک مشکل کلیدی در تحلیل شبکه‌های اجتماعی است. روش‌های زیادی برای حل مشکل پیش‌نهاد شده است. در میان این روش‌ها، روش‌های مبتنی بر شباهت با در نظر گرفتن ساختار شبکه و استفاده به عنوان معیاری اساسی از تعداد همسایه‌های مشترک بین دو گره برای ایجاد شباهت ساختاری، کارایی خوبی از خود نشان می‌دهند.

واژه‌های کلیدی: تحلیل شبکه اجتماعی، پیش‌بینی پیوند، معیار مشابهت، شبکه هم‌نویسندگی.

۱. مقدمه

روند روبه‌رشد کاربران شبکه‌سازی اجتماعی، زندگی‌ها و کسب‌وکار جهان را هر روز تغییر داده است [۱]، این موضوع در پژوهش‌های اخیر به بحث گذاشته شده است. شبکه اجتماعی، مکانی است که دو یا چند نفر که دارای رابطه مشترک هستند، می‌آیند و اطلاعات خود را به اشتراک می‌گذارند، دیدگاه خود را مبادله می‌کنند، بحث می‌کنند و همدیگر را در شبکه‌های اجتماعی دنبال می‌کنند. شبکه‌های اجتماعی می‌توانند از طریق تماس‌های رودررو یا برون‌خط باشند [۲]، مانند مدارس، دانشگاه‌ها، کنفرانس‌ها و دیگر مکان‌های عمومی، اما با استفاده از توییتر [۳]، فیس‌بوک [۴]، گوگل پلاس و لینکدین، این ارتباطات می‌توانند برخط باشند. شبکه اجتماعی، یک گراف است که در آن افراد به صورت گره نشان داده می‌شوند و تماس آن‌ها، یال بین آن‌ها است. یال بدان معناست که این گره‌ها با همدیگر در ارتباط یا تعامل هستند، حتی اگر از نظر جغرافیایی از همدیگر دور باشند. از چند سال گذشته، جذابیت مردم به سمت شبکه‌های اجتماعی، مانند فیس‌بوک و توییتر، فرصت‌های بسیاری برای محققان جهت مطالعه و تحلیل مشخصات شبکه اجتماعی و همچنین جنبه‌های مختلف رفتار انسانی در کل شبکه اجتماعی، به وجود آورده است. تحلیل شبکه اجتماعی، به مردم در شناسایی افراد دارای علائق مشترک و جوامع مربوطه آن‌ها، کمک می‌کند [۵]. علاوه بر آن، می‌توانیم بفهمیم که چگونه جوامع ساخته می‌شوند و در شبکه اجتماعی مردم چگونه با هم در تعامل هستند.

پژوهشگران متخصص در زمینه تحلیل شبکه اجتماعی، با مسائل بسیار زیادی مواجه هستند [۶]. پیش‌بینی ارتباط بین گره‌ها، به معنای پیش‌بینی پیوندهای مشاهده‌نشده یا ناموجود در شبکه اجتماعی از آن جمله است [۷]. ماهیت پویای شبکه اجتماعی، این چالش را جالب‌تر کرده است. [۸،۹]

پیش‌بینی پیوند بین گره‌ها، توجه پژوهشگران حوزه‌های مختلف را به خود جلب کرده است. پژوهشگران در حوزه‌های مختلف، مانند سیستم‌های توصیه‌کننده که دوستان جدید با علائق مشترک را توصیه می‌کنند [۱۰]، بیوانفورماتیک در شبکه‌های تعاملی پروتئین - پروتئین [۱۱] و نقل‌وقول در شبکه نقل‌وقول [۱۲]، در حال پژوهش هستند. علاوه بر آن، از پیش‌بینی لینک برای پیش‌بینی لینک‌های آینده در شبکه‌ها (فیس‌بوک)، پیش‌بینی لینک‌های ناموجود [۱۳]، اقلام جهت فروش در آمازون [۱۴]، مجرمان در شبکه‌های مجرمان [۱۵]، استفاده‌شده است؛ بنابراین، تحلیل پیش‌بینی لینک‌های آینده و همچنین تکمیل شبکه، بسیار مهم است.

روش‌های پیش‌بینی زیادی برای پیوند بین گره‌ها که از اطلاعات گره‌ها و توپولوژی شبکه اجتماعی برای محاسبه شباهت بین جفت گره‌ها استفاده می‌کنند؛ وجود دارد. شباهت محاسبه شده، به صورت یک نمره یا امتیاز نشان داده می‌شود که به یک جفت گره (x,y) ، این امتیاز

داده می‌شود. امتیاز بالا نشان‌دهنده شانس بالاتر پیوند داشتن گره x و y با هم در آینده است. از سوی دیگر، نمره کم، نشان‌دهنده آن است که گره x و y به احتمال کم در آینده با هم پیوند خواهند داشت. در ادبیات پژوهش، روش‌های پیش‌بینی پیوند جدید برای محاسبه نمره شباهت بین جفت گره‌ها، هر دو گره را به اندازه برابر مدنظر قرار می‌دهند لذا این روش‌ها گاهی قادر به پیش‌بینی پیوندهای آتی نیستند. ما در این مقاله، یک معیار شباهت جدید را پیشنهاد می‌دهیم که تک‌تک گره‌ها را جهت محاسبه شباهت مدنظر قرار می‌دهد.

۲. پیشینه پژوهش

در سال‌های اخیر، پیش‌بینی پیوند در شبکه‌های اجتماعی موضوعی مهم برای پژوهشگران علوم رایانه و شبکه‌های اجتماعی شده است. یکی از روش‌هایی که در این حوزه استفاده می‌شود، استفاده از معیارهای مشابهت بین گره‌ها به منظور پیش‌بینی پیوند است.

تحقیقات زیادی در این حوزه انجام شده است که برخی از آن‌ها عبارتند از:

کار ارائه‌شده توسط لادا و آدامیک [۱۶] که در آن از معیارهای مشابهت احتمالی بین گره‌ها استفاده شده است.

طرح ارائه‌شده توسط زاراپور و همکاران [۱۷]، که در آن بررسی گسترده‌ای از روش‌های پیش‌بینی پیوند با استفاده از معیارهای مشابهت ارائه شده است.

این تحقیقات نشان می‌دهد که استفاده از معیارهای مشابهت بین گره‌ها برای پیش‌بینی پیوند در شبکه‌های اجتماعی معیارهای مشابهت مختلفی وجود دارند که در این تحقیقات استفاده شده‌اند، از جمله:

معیارهای مبتنی بر فاصله: این معیارها بر اساس فاصله بین گره‌ها در شبکه، مانند فاصله کوتاهترین مسیر یا فاصله گرانشی، تعیین می‌شوند.

معیارهای مبتنی بر همسایگی: این معیارها بر اساس تعداد ویژگی‌های مشترک بین دو گره و همچنین تعداد مشترکین همسایگان آن‌ها تعیین می‌شوند.

معیارهای مبتنی بر گراف: این معیارها بر اساس ساختار گراف و ویژگی‌های آن، مانند تعداد گره‌های مشترک همسایگان یا تعداد مثلث‌های مشترک، تعیین می‌شوند.

معیارهای مبتنی بر زمان: این معیارها بر اساس زمانی که دو گره با یکدیگر متصل شده‌اند، مانند تعداد مشترکین بازدیدکنندگان یا تعداد مشترکین کامنت‌ها، تعیین می‌شوند.

در این تحقیقات، برای بهبود دقت پیش‌بینی پیوند، روش‌های مختلفی مانند ترکیب معیارها، استفاده از الگوریتم‌های یادگیری ماشین، و استفاده از ویژگی‌های بیشتر گره‌ها در نظر گرفته شده‌اند.

به طور کلی، معیارهای مشابهت بین گره‌ها به عنوان یکی از روش‌های پیش‌بینی پیوند در شبکه‌های اجتماعی مورد توجه

پژوهشگران علوم رایانه و شبکه‌های اجتماعی قرارداد. استفاده از معیارهای مشابهت مختلف و ترکیب آن‌ها با هم، می‌تواند دقت پیش‌بینی پیوند را افزایش دهد. همچنین، استفاده از الگوریتم‌های یادگیری ماشین و ویژگی‌های بیشتر گره‌ها می‌تواند به دقت پیش‌بینی کمک کند، به‌خصوص در شبکه‌های اجتماعی بزرگ و پیچیده.

به علاوه، روش‌های پیش‌بینی پیوند بر اساس معیار مشابهت در شبکه‌های اجتماعی در بسیاری از پروژه‌های کاربردی نیز مورد استفاده قرار می‌گیرد، مانند تو صیه‌گرهای فیلم و موسیقی، پیش‌بینی دوستان جدید در شبکه‌های اجتماعی و بهبود رتبه‌بندی صفحات وب.

برای پیش‌بینی پیوند در شبکه‌های اجتماعی، معیارهای مختلفی برای سنجش شباهت بین گره‌ها (نودها) وجود دارد. در ادامه به برخی از این معیارها اشاره می‌کنیم:

معیار مشابهت جاکارد: این معیار بر اساس تعداد مشترک بودن همسایگان دو گره است که به عنوان یک معیار ساده و کارآمد برای پیش‌بینی پیوند استفاده می‌شود.

معیار مشابهت عددی: این معیار برای محاسبه شباهت بین دو گره، فاصله یا اختلاف بین بردارهای مشخصه دو گره را محاسبه می‌کند. این بردارها معمولاً شامل ویژگی‌هایی مانند سن، جنسیت، شغل و... هستند.

معیار مشابهت کسینوسی: این معیار برای محاسبه شباهت بین دو گره از طریق زاویه بین بردارهای مشخصه آن‌ها استفاده می‌شود. این معیار در پیش‌بینی پیوند در شبکه‌های اجتماعی بسیار مؤثر است.

معیار مشابهت آدامیک-آدار: این معیار برای سنجش شباهت بین دو گره، مجموع وزن‌های پیوندهای مشترک بین آن‌ها را محاسبه می‌کند.

معیار مشابهت همسایگان مشترک: در این معیار، تعداد همسایگان مشترک دو گره با یکدیگر در نظر گرفته می‌شود. این معیار در برخی موارد می‌تواند به عنوان یک معیار کارآمد برای پیش‌بینی پیوند در شبکه‌های اجتماعی مورد استفاده قرار گیرد.

لازم به ذکر است که در بسیاری از موارد، استفاده از تنها یک معیار برای پیش‌بینی پیوند در شبکه‌های اجتماعی کافی نیست و باید از ترکیب چندین معیار با هم مثلاً ترکیب خطی از معیارهای مشابهت جاکارد، مشابهت کسینوسی و مشابهت آدامیک-آدار استفاده کرد.

در ادامه، چندین معیار دیگر نیز برای پیش‌بینی پیوند در شبکه‌های اجتماعی وجود دارد که به طور کلی، به دو دسته معیارهای مبتنی بر ساختار شبکه و معیارهای مبتنی بر مشخصات گره‌ها تقسیم می‌شوند.

در معیارهای مبتنی بر ساختار شبکه، از ویژگی‌هایی مانند درجه گره‌ها، میانگین مسیر کوتاه بین گره‌ها و میزان همبستگی بین گره‌ها استفاده می‌شود. از طرفی، در معیارهای مبتنی بر مشخصات گره‌ها، ویژگی‌هایی مانند سن، جنسیت، تحصیلات و شغل گره‌ها به عنوان معیارهای مشابهت مورد استفاده قرار می‌گیرند.

در نهایت، برای پیش‌بینی پیوند در شبکه‌های اجتماعی، لازم است از الگوریتم‌های یادگیری ماشینی مانند شبکه‌های عصبی و رگرسیون لجستیک استفاده کرد. این الگوریتم‌ها با استفاده از مجموعه‌ای از داده‌های شبکه، تلاش می‌کنند یک مدل پیش‌بینی برای پیوندهای جدید در شبکه تولید کنند.

یکی از کارهای آشنا در پیش‌بینی پیوند بین دو گره x و y سنجش شباهت بین آن‌ها می‌باشد. هرچه دو گره بیشتر به هم شبیه باشند، احتمال اینکه یک پیوند بین این دو گره به وجود بیاید، بیشتر است. با استفاده از این رویکرد، اکثر پژوهش‌ها به‌صورت زیرند:

وانگ و همکاران [۱۸]، مدل‌های گرافیکی احتمالاتی محلی را ارائه داده‌اند که می‌توانند تا گراف‌های مقیاس بزرگ گسترده شوند و شانس وجود پیوند بین دو گره را اندازه‌گیری کنند. تلیندا و همکاران

[۱۸]، مدل احتمال محلی را به وسیله جدیدترین معیارهای شباهت یعنی Root PageRank و Adamic Adar (AA) که به آن پیش‌بینی پیوند آگاه از زمان، گفته می‌شود، را تعمیم دادند. سه معیار پراستفاده، یعنی Katz, PageRank و احتمال گریز، توسط سانگ و همکاران [۲۰]، برای پیش‌بینی پیوند استفاده شده‌اند. موناسینگه و همکاران [۲۱]،

معیار «نمره زمانی» (TS) را برای پیش‌بینی پیوند بین جفت‌گره‌هایی که در زمان تعامل و تعداد گره به هم رسیده‌اند را مورد استفاده قرار می‌دهند تا قدرت رابطه بین پیوندهای قوی و برجسته زمانی مشغولیت به هم (engagement) متصل به آینده را به‌دست آورند. سوارز

و همکاران [۲۲]، در مقاله خود، نمره شباهت را برای هر جفت گره منفصل را از طریق معیارهای شباهت توپولوژیک، محاسبه کردند. ژانگ و فیلیپ [۲۳]، دو پرسش را برای محاسبه شباهت در نظر گرفتند که در آنجا یک روش جستجوی شباهت مبتنی بر یال را پیشنهاد کردند که شباهت بین دو گره را به‌صورت شباهت دو پرشی، محاسبه می‌کند.

ابراهیم و چن [۲۴]، از ساختار اجتماعی، مرکزیت گره و اطلاعات زمانی استفاده کردند تا پیوندهای احتمالی در شبکه‌های اجتماعی را پیش‌بینی کنند. هان و همکاران [۲۵] بر روی شباهت بین جوامع کار کردند و در مقاله خود، معیار «درجه شباهت اجتماعی» (CSD) را پیشنهاد دادند تا با آن درجه شباهت علائق بین چندین کاربر در یک

جامعه، به‌دست‌آید. موراتا و همکاران [۲۶] دو نمره شباهت وزنی یعنی WAA و WCN، تعمیم‌یافته از «همسایگان مشترک» [۲۷] و Adamic Adar [۲۸] را پیشنهاد دادند. بر اساس مقاله موراتا [۲۶]، اسماعیل و همکاران [۲۹]، ضریب Jaccard وزنی (WJC) را پیشنهاد کردند.

یافتن پیوندهای تماس جدید یا ناموجود بین انسان‌ها در یک شبکه اجتماعی، یک مسئله چالش‌برانگیز در نظر گرفته می‌شود. وانگ و همکاران [۳۰] نتایج امیدوارکننده‌ای برای مسئله استنباط گراف، ارائه می‌دهند. صمد و همکاران [۳۱]، پروفایل‌های اجتماعی از انسان‌ها و تأثیر پروفایل‌ها بر پیش‌بینی لینک را تحلیل کردند و ثابت کردند که همه ویژگی‌های اجتماعی انسان‌ها، اهمیت متفاوت دارند. در یک گراف، بین دو گره، بر روی یال‌ها، وزن نشان‌دهنده میزان قدرت پیوند

تماس‌های مختلف است. جانوتولا و همکاران [۳۲]، بر پیش‌بینی لینک برای شبکه اجتماعی آنلاین (OSN) با استفاده از توان پیش‌بینی‌کننده ترکیب شبکه‌های دوستی و تعامل، تمرکز کردند. با استفاده از شبکه‌های دوستی، آن‌ها پیش‌بینی یال‌های آتی در شبکه‌های تعامل را بهبود دادند. در یک شبکه دوستی، یک پیوند «فالو یا دنبال کردن» وجود دارد، اما در شبکه تعاملی، دو گره در یک روز خاص، با هم گفتگو دارند. برای پیش‌بینی شبکه‌های تعاملی آینده، پیش‌بینی‌کننده‌ها به سه دسته تقسیم می‌شوند، پیش‌بینی‌کننده‌هایی که از دوستی استفاده نمی‌کنند، پیش‌بینی‌کننده‌هایی که تنها از دوستی‌های فعلی استفاده می‌کنند و پیش‌بینی‌کننده‌هایی که از دوستی‌های پیش‌بینی شده، استفاده می‌کنند. نتایج شامل حاصل از دوستی‌های فعلی، واقعاً باعث به وجود آمدن یک پیش‌بینی‌کننده پیوند بسیار بهتر برای شبکه‌های تعاملی می‌شود. ژو و همکاران [۳۳] بر روی الگوریتم‌های مختلف برای مسئله پیش‌بینی پیوند مبتنی بر جذب شباهت برای گره‌های که پیوند حذف‌شده در شبکه دارند، استفاده کردند. برای این رویکرد، آن‌ها دو دسته گسترده را تعریف کردند، یکی اطلاعات محلی برای ایجاد حمله‌های بهینه برای معیارهای CND در مورد پیوندهای هدف در یک گروه و دیگری، اطلاعات شبکه جهانی برای حالت‌های خاص انگیزش یافته که از معیار NP-Hard استفاده می‌کند. برای ردیابی این پیوندهای ناموجود یا حذف شده، آن‌ها از الگوریتم‌های زمان چندجمله‌ای استفاده می‌کنند. لیم و همکاران [۳۴]، بر روی پیش‌بینی پیوند پنهان با استفاده از شبکه‌های مجرمان، کار کردند. آن‌ها بررسی کردند که معیارهای یادگیری ماشینی تحت نظارتی، نیاز به دیتاست‌های بزرگ برای آموزش دادن و تست کردن، داشتند؛ بنابراین، برای پیش‌بینی پیوندهای پنهان، آن‌ها از کاربرد «یادگیری تقویتی عمیق» (DRL) برای بازسازی شبکه‌های مجرمان، استفاده کردند. در آزمایش‌های تجربی، ایشان نتیجه‌گیری کردند که پیش‌بینی پیوند از طریق DRL، عملکرد بهتری نسبت به یادگیری ماشینی نظارتی دارند. علاوه بر آن، لیم و همکاران [۳۵]، عملکرد DRL با یادگیری ماشینی نظارتی از نظر دقت پیش‌بینی‌کنندگی و قدرت محاسبات را مقایسه کردند.

۳. روش تحقیق

ما در این مقاله با استفاده از پنج شبکه اجتماعی هم - نویسنده‌گی که به وسیله گراف بدون جهت $G=(V,E)$ نشان داده می‌شوند، طرح پیشنهادی خود را آمیخ می‌کنیم. در گراف اجتماعی، E نشان‌دهنده مجموعه پیوندهای بین نویسندگان است و V مجموعه گره‌های نشان‌دهنده نویسندگان است. مقاله ما، قرار است یک نمره $S(x,y)$ به هر جفت گره (x,y) ، بدهد. احتمال پیوند بین گره‌ها، با یک نمره معین، نشان‌دهنده می‌شود. هر چه مقدار $S(x,y)$ بیشتر باشد، احتمال اینکه بین گره‌های x و y برای جفت گره (x,y) یک پیوند به وجود بیاید، بیشتر می‌شود.

۱.۳. توضیح دیتاست

در این مقاله، از ۵ دیتاست مختلف [۳۶] (یعنی CondMat, AstroPh, GrQc, HepTh و HepPh) که حاوی شبکه‌های هم‌نویسنده‌گی می‌باشند، استفاده کردیم. این دیتاست‌ها، مشارکت پژوهشی بین نویسندگان را پوشش می‌دهد و از e-print arxiv، استخراج شده‌است. اگر یک نویسنده x ، در یک مقاله با نویسنده y همکاری داشته‌است، این گراف حاوی یک یال بدون جهت بین x و y می‌شود. اگر z عدد «هم‌نویسنده» در یک مقاله وجود داشته‌باشد، در نتیجه، (زیر) گراف کامل بر روی z گره تولید می‌کند. علاوه بر آن، معیارهای تفصیلی دیتاست‌ها در جدول ۱ بیان شده‌اند.

جدول ۱: معیارهای تفصیلی دیتاست‌ها

ویژگی‌ها		دیتاست
تعداد یال‌ها	تعداد گره‌ها	
198110	18772	AstroPh
93497	23133	CondMat
14496	5242	GrQc
118521	12008	HepPh
25998	9877	HepTh

۲.۳. ضریب جکارد (Jaccard)

یکی از معیارهای مشابهت معروف در شبکه‌های اجتماعی، معیار مشابهت جاکارد (Jaccard similarity) است. این معیار مشابهت بر اساس تعداد مشترک گره‌هایی که دو گره در اختیار دارند، محاسبه می‌شود. با این معیار، دو گره که بیشترین تعداد گره‌های مشترک را دارند، بیشترین احتمال پیوند دارند.

معیار مشابهت جاکارد به راحتی قابل محاسبه است و توانایی پیش‌بینی پیوند را در شبکه‌های اجتماعی با دقت قابل قبولی فراهم می‌کند. این معیار همچنین مزیت کم‌هزینه بودن محاسبات را نیز دارد. با این حال، معیار مشابهت جاکارد تنها یکی از معیارهای مشابهت است و برای بهبود دقت پیش‌بینی پیوند، بهتر است از ترکیب چندین معیار با هم استفاده کرد.

ضریب جکارد [۳۷]، هر دو تعداد کل و تعداد مشترک همسایگان را برای محاسبه شباهت بین جفت گره‌ها استفاده می‌کند. ضریب جکارد بین دو گره u و v به صورت زیر محاسبه می‌شود:

$$(1)$$

$$Jc(u, v) = \frac{|\tau(u) \cap \tau(v)|}{|\tau(u) \cup \tau(v)|}$$

در اینجا، $\tau(u)$ به معنای مجموع گره‌هایی است که مجاور با گره u می‌باشند و $|\tau(u) \cap \tau(v)|$ تعداد همسایگان مشترک بین جفت گره‌های u و v می‌باشند.

۳.۳. شاخص سورنسن (Sorensen)

شاخص سورنسن یک معیار مشابهت بین دو مجموعه می‌باشد که برای محاسبه شباهت بین دو گره در یک شبکه استفاده می‌شود. این شاخص بر مبنای تعداد عناصری است که در هر دو مجموعه وجود دارند و همچنین تعداد عناصری که در حداقل یکی از دو مجموعه وجود دارد.

در تعیین معیار مشابهت جهت پیش‌بینی پیوند در شبکه‌های اجتماعی، روش تخصیص منابع (Resource Allocation) یکی از روش‌های پرکاربرد است. در این روش، معیار مشابهت برای دو گره، مجموع مقادیر معیار مشابهت بین هر یک از همسایگان آن‌ها محاسبه می‌شود. به عبارت دیگر، در این روش، منابع موجود در همسایگان هر گره برای تعیین شباهت آن با سایر گره‌ها استفاده می‌شود. برای دو گره i و j با همسایگان $N(i)$ و $N(j)$ به ترتیب، معیار تخصیص منابع به صورت زیر محاسبه می‌شود:

$$RA(i, j) = \frac{\sum_{k \in N(i) \cap N(j)} 1}{|N(k)|}$$

در اینجا، $|N(k)|$ تعداد همسایگان گره k است. این روش به دلیل سادگی و کارایی زیاد، در پیش‌بینی پیوند در شبکه‌های اجتماعی کاربرد فراوانی دارد. علاوه بر این، این روش نسبت به روش‌های دیگر، از جمله شاخص جاکارد و شاخص سورنسن، در صورت وجود اتصال بین دو گره، نتایج بهتری ارائه می‌دهد.

تخصیص یا تقسیم منبع، توسط ژو و همکاران [۴۰] پیشنهاد شده و به گره‌های درجه بالاتر، جریمه سنگین‌تری تحمیل می‌کند. این معیار شباهت، عملکرد بهتری برای شبکه‌ای که درجات میانگین زیاد دارد، دارا می‌باشد. جالب‌ترین حقیقت در مورد تقسیم منابع این است که نه تنها از همسایگان بلکه همچنین از همسایه‌های همسایگان استفاده می‌کند. شباهت بین جفت گره با استفاده از تقسیم منابع، به صورت زیر محاسبه می‌شود:

$$RA(u, v) = \sum_{z \in \tau(u) \cap \tau(v)} \frac{1}{|\tau(z)|} \quad (۴)$$

۴. معیار پیشنهادی

ما در این مقاله یک معیار شباهت جدید را پیشنهاد می‌دهیم، معیاری که شباهت X به Y و همچنین شباهت Y به X را مدنظر قرار می‌دهد. این معیار، کار محاسبه شباهت را به دو قسمت تقسیم می‌کند. اولی، با استفاده از معادله ۵، متوجه می‌شود که u چه میزان به v مشابه است. دومی، شباهت v به u را با استفاده از معادله ۶، محاسبه می‌کند. در نهایت، هر دو نتیجه معادلات ۵ و ۶ را با استفاده از معادله ۷ مورد استفاده قرار می‌دهد. در اینجا، به گره‌های دارای درجه بالاتر، جریمه تحمیل می‌شود تا پیوندهای بین گره‌های درجه بالاتر و گره‌های درجه پایین‌تر را پیدا کنند. در اکثر روش‌ها، پیوند بین گره‌های درجه بالاتر و درجه پایین‌تر نادیده گرفته می‌شود، اما در روش پیشنهادی ما، جریمه سنگینی بر گره‌های درجه بالاتر تحمیل می‌شود و شانس بیشتری به گره‌های درجه کمتر می‌دهد تا با گره‌های درجه بالاتر، پیوند پیدا کنند. علاوه بر آن، در روش‌های پیش‌بینی پیوند سنتی، جفت گره را به عنوان یک واحد در نظر می‌گرفتند و بر اساس اشتراک بین آن‌ها، تصمیم‌گیری می‌کردند. ما استدلال می‌کنیم که هر دو گره درون یک جفت، شباهت

به این صورت که اگر دو مجموعه A و B را داشته باشیم، سورنسن برای آن‌ها برابر است با:

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}$$

که در آن $|A \cap B|$ تعداد عناصری است که در هر دو مجموعه وجود دارند و $|A \cup B|$ تعداد عناصری است که در حداقل یکی از دو مجموعه وجود دارد. شاخص سورنسن مقداری بین ۰ تا ۱ دارد، که هر چه مقدار آن به یک نزدیک‌تر باشد، دو مجموعه به همان اندازه شبیه به یکدیگر هستند. این شاخص برای مقایسه مجموعه‌های با اندازه‌های مختلف مناسب است و به دلیل سادگی و سرعت محاسبه، در بسیاری از موارد به عنوان یک شاخص مشابهت مفید استفاده می‌شود.

معیارهای شباهت سورنسن [۳۸]، با معادله زیر تعریف می‌شوند. این معادله اندازه همسایگان مشترک را در نظر می‌گیرد و همچنین اشاره می‌کند که درجه کمتر گره‌ها، احتمال لینک بیشتر دارند.

$$SI(u, v) = \frac{|\tau(u) \cap \tau(v)|}{|\tau(u)| + |\tau(v)|} \quad (۲)$$

۴.۳. کسینوس سورنسن

شاخص کسینوس سورنسن (Cosine similarity index) یک معیار مشابهت بین دو بردار است که برای محاسبه شباهت بین دو گره در یک شبکه استفاده می‌شود. این شاخص برای بردارهای با ابعاد بالا و جستجوی سریع برای پیدا کردن بردارهای مشابه کاربرد دارد.

برای دو بردار X و Y با ابعاد n ، شاخص کسینوس سورنسن برابر با ضرب داخلی بین دو بردار تقسیم بر حاصل ضرب دو نرم آن‌ها است. به این صورت که ابتدا برای هر بردار، نرم آن با استفاده از جمع مربعات عناصر و جذر مجموع، محاسبه می‌شود. سپس برای دو بردار X و Y ، مقدار شاخص کسینوس سورنسن به صورت زیر محاسبه می‌شود:

$$\text{cosine_similarity}(X, Y) = \frac{(X \cdot Y)}{(\|X\| \|Y\|)}$$

در اینجا، $(X \cdot Y)$ نشان‌دهنده ضرب داخلی بین دو بردار است، و $\|X\|$ و $\|Y\|$ نشان‌دهنده نرم هر کدام از بردارهاست.

مقدار شاخص کسینوس سورنسن بین -۱ تا ۱ است، که مقدار ۱ نشان‌دهنده تمامی شباهت بین دو بردار و مقدار -۱ نشان‌دهنده تمامی عدم شباهت بین دو بردار است. اگر مقدار شاخص کسینوس سورنسن برابر با صفر باشد، این نشان‌دهنده عدم وجود هرگونه شباهت بین دو بردار است.

کسینوس سالتون [۳۹]، یک معیار کسینوس مشترک است که برای محاسبه شباهت بین جفت گره استفاده می‌شود. این معیار به صورت زیر تعریف می‌شود:

$$Sc(u, v) = \frac{|\tau(u) \cap \tau(v)|}{\sqrt{|\tau(u)| \cdot |\tau(v)|}} \quad (۳)$$

۵.۳. تخصیص منابع

خاص خود به همدیگر را دارند. می تواند این گونه باشد که یک شخص به شخص دیگر ۱۰۰٪ مشابه باشد، اما شخص دیگر، به آن اندازه شباهت به اولی نداشته باشد.

(۵)

$$sam(u_v) = \frac{|\tau(u) \cap \tau(v)|}{|\tau(u)|}$$

(۶)

$$sam(v_u) = \frac{|\tau(u) \cap \tau(v)|}{|\tau(v)|}$$

(۷)

$$sam(u, v) = (sam(u_v) + sam(v_u))/2$$

جهت مقایسه روش پیشنهادی خود با سایر روش های پیش بینی پیوند قبلی، درصد پیوندهای اجتماعی پیش بینی شده را با سایر روش ها مقایسه می کنیم. علاوه بر آن، معیار دقت (نسبت پیش بینی های صحیح به تعداد کل یال های ورودی)، برای ارزیابی معیارهای مشابهت را به صورت زیر استفاده می کنیم.

(۸)

$$Accuracy = 1 - \frac{E(G_V) + E(G_P) - 2E(G_V \cap G_P)}{Max(E(G_V), E(G_P))}$$

در معادله بالا، E نشان دهنده یال های گراف اجتماعی، G_V ، نشان دهنده گراف اجتماعی اصلی، G_P نشان دهنده گراف اجتماعی پیش بینی شده است، تابع Max، حداکثر تعداد یال ها از گراف اصلی و پیش بینی شده را تولید می کند.

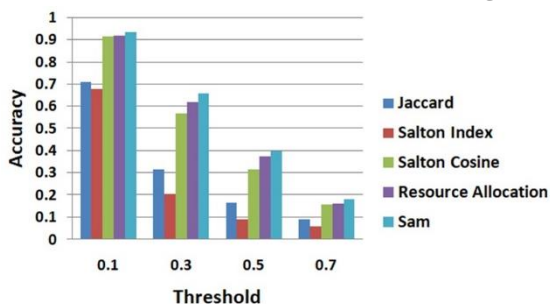
۵. شبیه سازی و مقایسه طرح پیشنهادی

ما در این مقاله جهت ارزیابی و مقایسه نتایج طرح پیشنهادی خود با سایر روش ها از پنج دیتاست (Astro، CondMat، GrQc، HepPh و HepTh) استفاده کرده ایم. در ابتدا، این دیتاست ها، در خود یک گراف اجتماعی از نویسندگان، با تعداد مختلف گره ها و یال ها به صورت آمار ارائه شده، در جدول ۱ نشان داده شده اند. سپس این یال ها را از هر گراف اجتماعی حذف کردیم و معیارهای شباهت بر روی هر جفت گره در هر گراف اجتماعی را بر اساس معادلات ۵، ۶ و ۷ محاسبه کردیم. بعد از اعمال معیارهای مشابهت، تعدادی آستانه بر روی شباهت بین هر جفت گره اعمال کردیم و ۱۰۰ گراف اجتماعی پیش بینی شده، ۲۰ گراف برای هر معیار شباهت و دیتاست بر روی چهار آستانه مختلف، ایجاد کردیم. در آخر، دقت هر معیار مشابهت را بررسی کردیم.

۱.۵. مقایسه با استفاده از دیتاست Astro

برای محاسبه شباهت، ۱۰۰۰ یال از دیتاست Astro برای پیش بینی استفاده شد. شکل ۱، نتایج دیتاست Astro را نشان می دهد. در این شکل آستانه بر روی محور X و درصد پیش بینی به شکل دقت بر روی

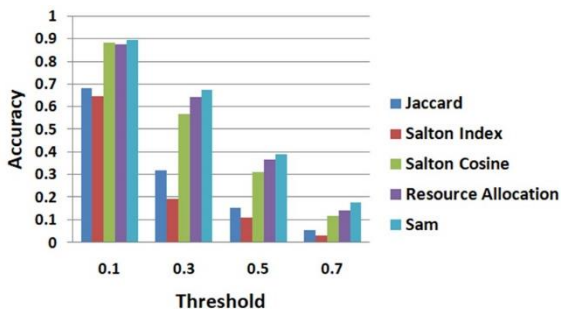
محور Y نشان داده شده. الگوی یکسانی در همه شکل ها دیده می شود. آستانه های حاصله نشان داد که در آستانه ۰/۱، معیار پیشنهادی به بالاترین نتایج با پیش بینی لینک ۹۳٪ رسید. به بیان دقیق تر، حداقل دقت حاصل شده توسط معیار پیشنهادی، ۱۸٪ است و بیشینه، ۹۳٪ است. علاوه بر آن، معیار پیشنهادی، Salton Cosine و «تقسیم منابع» بروی همه آستانه ها به نتایج خوبی رسیدند. از سوی دیگر، شاخص جاکارد و سالتون، نتوانستند به نتایج معقولی دست یابند. به طور کلی، حداقل دقت حاصل شده توسط جاکارد، شاخص سالتون، سالتون کسینوس (Salton Cosine) و «تقسیم منابع به ترتیب ۹٪، ۶٪، ۱۶٪ و ۱۶٪ می باشند.



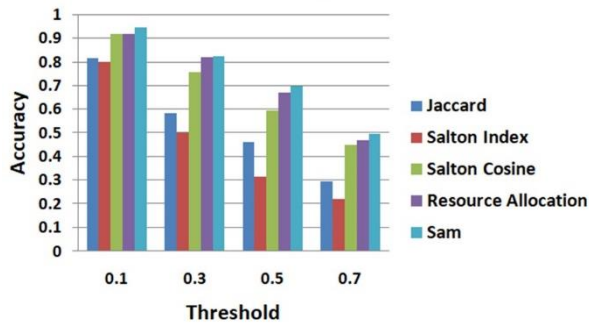
شکل ۱: نتایج دیتاست Astro برای پیش بینی ۱۰۰۰ یال

۲.۵. مقایسه ها با استفاده از دیتاست CondMat

جهت محاسبه شباهت، ۱۰۰۰ یال از ۹۳۴۹۷ یال برای پیش بینی از دیتاست CondMat، استفاده شد. شکل ۲، نتایج دیتاست CondMat را نشان می دهد. که در این شکل آستانه بر روی محور X و درصد پیش بینی به شکل دقت بر روی محور Y نشان داده شده است. در آستانه ۰/۱، معیار پیشنهادی به بالاترین نتایج پیش بینی لینک ۹۰٪ دست یافت. علاوه بر آن، حداقل دقت حاصل شده توسط معیار پیشنهادی، ۱۸٪ است. در آستانه های ۰/۱ و ۰/۳، معیار پیشنهادی، Salton Cosine و Resource Allocation، به نتایج خوبی رسیدند. از سوی دیگر، این بار هم Jaccard و Salton Index نتوانستند به نتایج مناسبی دست یابند. به طور کلی، حداقل دقت حاصل شده توسط Salton Index، Jaccard، Salton Cosine و Resource Allocation، به ترتیب ۶٪، ۳٪، ۱۲٪ و ۱۴٪ می باشند. برای آستانه های ۰/۵ و ۰/۷، با اینکه همه معیارهای شباهت قادر به تولید نتایج بهتری نبودند، اما توانستیم نتایج بهتری برای معیار پیشنهادی نسبت به دیگران، مشاهده کنیم.



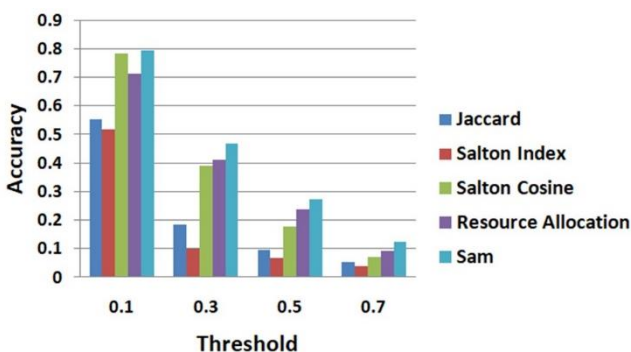
شکل ۲: نتایج دیتاست CondMat برای پیش بینی ۱۰۰۰ یال



شکل ۴: نتایج دیتاست HepPh برای پیش‌بینی ۱۰۰۰ یال

۵.۵. مقایسه با استفاده از دیتاست HepTh

برای پیش‌بینی لینک بر روی دیتاست HepTh، ۱۰۰۰ یال مختلف انتخاب شدند. شکل ۵ نتایج دیتاست HepTh را نشان می‌دهند که آستانه‌های حاصله، نشان دادند که در آستانه ۰/۱، روش پیشنهادی ما معیار پیشنهادی، به ۰/۷۹ پیش‌بینی لینک رسید که نسبت به باقی روش‌ها، بالاتر بود. به بیان دقیق‌تر، حداقل دقت حاصل شده توسط معیار پیشنهادی، ۰/۱۲ است و حداکثر ۰/۷۹. علاوه بر آن، معیار پیشنهادی، Salton Cosine و Resource Allocation بر روی همه آستانه‌ها، به نتایج خوبی رسیدند. از سوی دیگر، Jaccard و Salton Index به حداکثر ۰/۵۵ و ۰/۵۲ دست یافتند که قدری نسبت به دیتاست‌های قبلی، (یعنی Astro، CondMat، GrQc، HepPh)، کمتر رضایت‌بخش بودند. همین رفتار توسط Salton Cosine و Resource Allocation دنبال شد، زیرا حداکثر ۰/۷۸ و ۰/۷۱ لینک را پیش‌بینی کردند. به‌طور کلی، حداقل دقت حاصل توسط Jaccard، Salton Index، Salton Cosine و Resource Allocation به ترتیب ۰/۵، ۰/۴، ۰/۷ و ۰/۹ بودند که در مقایسه با دیتاست‌های قبلی (یعنی، Astro، CondMat، GrQc و HepPh)، بسیار کمتر بودند.



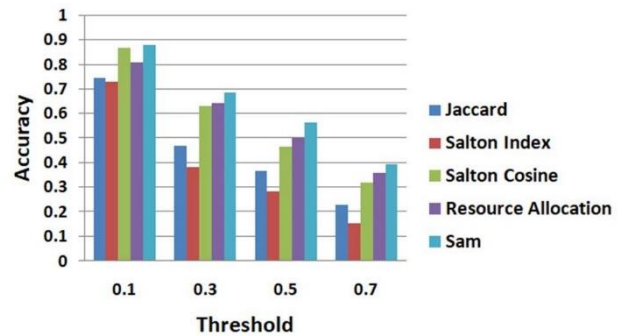
شکل ۵: نتایج دیتاست HepTh برای پیش‌بینی ۱۰۰۰ یال

۶. نتیجه‌گیری

در این مقاله، معیاری برای پیش‌بینی پیوندهای موجود در شبکه‌های اجتماعی را پیشنهاد دادیم و نتایج حاصل از روش پیشنهادی خود را بر روی پنج دیتاست مختلف، شبیه‌سازی کردیم. علاوه بر آن، این معیار پیشنهادی را با چهار روش پیش‌بینی پیوند Jaccard، Salton Index،

۳.۵. مقایسه با استفاده از دیتاست GrQc

برای پیش‌بینی لینک، ۱۰۰۰ یال مختلف از دیتاست GrQc انتخاب می‌شوند شکل ۳، نتایج دیتاست GrQc را نشان می‌دهند. آستانه‌های حاصله نشان دادند که در آستانه ۰/۱، معیار پیشنهادی به بالاترین نتایج با پیش‌بینی ۰/۸۸ پیوندها، دست یافت. به بیان دقیق‌تر، دقت حداقل حاصل شده به وسیله معیار پیشنهادی، ۰/۳۹ و حداکثر آن ۰/۸۸ است. علاوه بر آن، معیار پیشنهادی، Salton Cosine و Resource Allocation، بر روی همه آستانه‌ها به نتایج خوبی رسیدند. از سوی دیگر، Jaccard و Salton Index، حداکثر ۰/۷۴ و ۰/۷۳ را به دست آوردند که کمی نسبت به دو دیتاست قبلی (یعنی Astro و CondMat)، بهتر هستند. همین رفتار توسط Salton Cosine و Resource Allocation حاصل شد، زیرا حداکثر پیش‌بینی لینک ۰/۸۷ و ۰/۸۱ را حاصل کردند. به‌طور کلی، حداقل دقت توسط Jaccard، Salton Index، Salton Cosine و Resource Allocation به میزان ۰/۲۳، ۰/۱۵، ۰/۳۲ و ۰/۳۶ حاصل شدند که در مقایسه با دو دیتاست قبلی (یعنی، Astro و CondMat) بهتر هستند.



شکل ۳: نتایج دیتاست GrQc برای پیش‌بینی ۱۰۰۰ یال

۴.۵. مقایسه با استفاده از دیتاست HepPh

برای محاسبه شباهت، ۱۰۰۰ یال از ۱۱۸۵۲۱ یال، برای پیش‌بینی از دیتاست HepPh، استفاده می‌شوند. شکل ۴، نتایج دیتاست HepPh را نشان می‌دهند که در آن آستانه بر روی محور X و در صد پیش‌بینی به شکل دقت بر روی محور Y نشان داده شده است. در آستانه ۰/۱، معیار پیشنهادی، بالاترین نتایج را با پیش‌بینی ۰/۹۵ لینک‌ها حاصل کرد. از سوی دیگر، حداقل دقت توسط Sam ۰/۵۰ بود. در آستانه‌های ۰/۱ و ۰/۳، معیار پیشنهادی، Salton Cosine و Resource Allocation به نتایج خوبی رسیدند. از سوی دیگر، این بار نیز Jaccard و Salton Index نتوانستند نتایج مناسبی به دست آورند. به‌طور کلی، حداقل دقت توسط Jaccard، Salton Index، Salton Cosine و Resource Allocation به ترتیب ۰/۳۰، ۰/۲۲، ۰/۴۵ و ۰/۴۷ به دست آمدند. برای آستانه‌های ۰/۵ و ۰/۷، همه معیارهای شباهت، نتایج بهتری نسبت به دیتاست‌های قبلی (یعنی، Astro، CondMat، GrQc) به دست آوردند. موضوع جالب دیگری که می‌توانیم ببینیم این است که معیار پیشنهادی بهتر از بقیه فنون، کارایی داشته است.

Salton Cosine و Resource Allocation مقایسه کردیم. نتایج شبیه‌سازی بر روی پنج دیتاست مختلف (یعنی Astro، CondMat، GrQc، HepPh و HepTh)، نشان می‌دهد که روش پیشنهادی ما با دقت بیشتری می‌تواند پیوند موجود در یک شبکه اجتماعی را بین گره‌های

جدول ۲: نتایج مقایسه با معیار پیشنهادی

دیتاست	تعداد یال‌ها جهت پیش‌بینی	Jaccard		Salton Index		Salton Cosine		Resource Allocation		معیار پیشنهادی	
		Max	Min	Max	Min	Max	Min	Max	Min	Max	Min
Astro	۱۰۰۰	71%	9%	68%	6%	91%	16%	92%	16%	93%	18%
CondMat	۱۰۰۰	68%	6%	65%	3%	88%	12%	87%	14%	90%	18%
GrQc	۱۰۰۰	74%	23%	73%	15%	87%	32%	81%	36%	88%	39%
HepPh	۱۰۰۰	82%	30%	80%	22%	92%	45%	92%	47%	95%	50%
HepTh	۱۰۰۰	55%	5%	52%	4%	78%	7%	71%	9%	79%	12%

۷. پیشنهادهایی برای کارهای آتی

برای توسعه پژوهش‌های مرتبط با معیارهای مشابهت جهت شناسایی شبکه‌های اجتماعی موارد زیر پیشنهاد می‌گردد.

- بررسی تاثیر ساختار شبکه‌ها و ویژگی‌های گره‌ها بر پیش‌بینی پیوند در شبکه‌های اجتماعی
- بررسی و به‌کارگیری الگوریتم‌های یادگیری ماشین در پیش‌بینی پیوند در شبکه‌های اجتماعی
- بررسی تأثیر نویز و خطا در داده‌ها بر دقت پیش‌بینی پیوند در شبکه‌های اجتماعی
- بررسی تأثیر افزایش حجم داده بر دقت الگوریتم‌های پیش‌بینی پیوند در شبکه‌های اجتماعی
- بررسی تأثیر استفاده از الگوریتم‌های پیش‌بینی پیوند بر بهبود کارایی و عملکرد شبکه‌های اجتماعی

۸. مراجع

- [6] Alvi, Abdul & Islam, Arshad & Iqbal, Muhammad & Aleem, Muhammad. (2019). Centrality-Based Paper Citation Recommender System. EAI Endorsed Transactions on Industrial Networks and Intelligent Systems. 6. 159121. 10.4108/eai.13-6-2019.159121.
- [۷] Iraj Timuri; Mehdi Afzali "Increasing the accuracy of identifying overlapping communities using edge weighting". Intelligent multimedia processing and communication systems, ۱, ۱, ۲۰۱۹, ۲۰-۹
- [۸] Ajami, Mojtaba, & Asgari, Nasser. (۱۴۰۰). A comparative analysis of graphical structural measures to detect anomalies in online social networks. Intelligent multimedia processing and communication systems, ۲(۱), ۹-۱.
- [۹] Nderlou, Lida, & Hagazi Sharbian, Mohammad. (۱۴۰۱). Presenting a method for dynamic multi-layered social networks to discover influential groups based on the combination of evolutionary frog jump algorithm and C-means clustering. Intelligent multimedia processing and communication systems, ۳(۳), ۳۹-۲۹.
- [10] Z. Huang, X. Li, H. Chen, Link prediction approach to collaborative filtering, in: Proceedings of the 5th ACM/IEEE-CS joint conference on Digital libraries, JCDL'05, ACM, 2005, pp. 141-142.
- J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68-73.
- [۱۱] C. Lei, J. Ruan, A novel link prediction algorithm for reconstructing protein-protein interaction networks by topological similarity, Bioinformatics 29 (3) (2013) 355-364.
- K. Elissa, "Title of paper if known," unpublished.
- [۱۲] Samad, Abdul. Evaluation of Textual and Topological Similarity Measures for Citation Recommendation. Diss. CAPITAL UNIVERSITY, 2019.
- [۱۳] W. Peng et al., "Link Prediction in Social Networks: the State-of-the-Art," Sci China Inf Sci, vol. 58, no. 58, pp. 11101-38, 2015.
- [۱۴] B. Cao, N. N. Liu, and Q. Yang, "Transfer Learning for Collective Link Prediction in Multiple Heterogenous Domains," Int. Conf. Mach. Learn., pp. 180-186, 2010.
- [۱۵] G. Berlusconi, F. Calderoni, N. Parolini, M. Verani, and C. Piccardi, "Link prediction in criminal networks: A tool for criminal intelligence analysis," PLoS One, vol. 11, no. 4, p. e0154244, 2016.
- [1] Mallek, S., Boukhris, I., Elouedi, Z., & Lefèvre, E. (2019). Evidential link prediction in social networks based on structural and social information. Journal of computational science, 30, 98-107.
- [2] Pérez-Macías, N., Fernández-Fernández, J. L., & Rua Vieites, A. (2019). Entrepreneurial intentions: trust and network ties in online and face-to-face students. Education+ Training, 61(4), 461-479. Authorized licensed use limited to: Auckland University of Technology. Downloaded on June 05, 2020 at 21:29:50 UTC from IEEE Xplore. Restrictions apply.
- [3] Ahuja, R., Singhal, V., & Banga, A. (2019). Using Hierarchies in Online Social Networks to Determine Link Prediction. In Soft Computing and Signal Processing (pp. 67-76). Springer, Singapore.
- [4] Lai, Y. Y., Neville, J., & Goldwasser, D. (2019). TransConv: Relationship Embedding in Social Networks.
- [5] Yuan, W., He, K., Guan, D., Zhou, L., & Li, C. (2019). Graph kernel based link prediction for signed social networks. Information Fusion, 46, 1-10.

- similarity scores. *Data Mining and Knowledge Discovery*, 30(1), 147-180.
- [۳۰] Wang, D., Pedreschi, D., Song, C., Giannotti, F., & Barabasi, A. L. (2011, August). Human mobility, social ties, and link prediction. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 1100-1108). Acn.
- [۳۱] Samad, A., Islam, M. A., Iqbal, M. A., Aleem, M., & Arshed, J. U. (2017, December). Evaluation of features for social contact prediction. In *2017 13th International Conference on Emerging Technologies (ICET)* (pp. 1-6). IEEE.
- [۳۲] Junuthula, R. R., Xu, K. S., & Devabhaktuni, V. K. (2018, June). Leveraging friendship networks for dynamic link prediction in social interaction networks. In *Twelfth International AAAI Conference on Web and Social Media*.
- [۳۳] Zhou, K., Michalak, T. P., Wanek, M., Rahwan, T., & Vorobeychik, Y. (2019, May). Attacking Similarity-Based Link Prediction in Social Networks. In *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems* (pp. 305-313). International Foundation for Autonomous Agents and Multiagent Systems.
- [۳۴] Lim, M., Abdullah, A., Jhanjhi, N. Z., & Supramaniam, M. (2019). Hidden Link Prediction in Criminal Networks Using the Deep Reinforcement Learning Technique. *Computers*, 8(1), 8.
- [۳۵] Lim, M., Abdullah, A., & Jhanjhi, N. Z. (2019). Performance optimization of criminal network hidden link prediction model with deep reinforcement learning. *Jou*
- [۳۶] Leskovec, J., Kleinberg, J., & Faloutsos, C. (2007). Graph evolution: Densification and shrinking diameters. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 1(1), 2.
- [۳۷] Srilatha, P., & Manjula, R. (2016). Similarity index based link prediction algorithms in social networks: A survey. *Journal of Telecommunications and Information Technology*.
- [۳۸] Wang, P., Xu, B., Wu, Y., & Zhou, X. (2015). Link prediction in social networks: the state-of-the-art. *Science China Information Sciences*, 58(1), 1-38.
- [۳۹] Sarna, G., & Bhatia, M. P. S. (2017). Content based approach to find the credibility of user in social networks: an application of cyberbullying. *International Journal Of Machine Learning and Cybernetics*, 8(2), 677- 689.
- [۴۰] Zhou, T., Lü, L., & Zhang, Y. C. (2009). Predicting missing links via local information. *The European Physical Journal B*, 71(4), 623-630.
- [۱۶] Adar, E., Zhang, L., Adamic, L. A., & Lukose, R. M. (2004). Implicit+ Structure+ and+ the+ Dynamics+ of+ Blogspace.
- [۱۷] Samad, A., Qadir, M., Nawaz, I., Islam, M. A., & Aleem, M. (2020). A comprehensive survey of link prediction techniques for social network. *EAI Endorsed Transactions on Industrial Networks and Intelligent Systems*, 7(23), e3-e3.
- [۱۹] Tylenda, T., Angelova, R., & Bedathur, S. (2009, June). Towards timeaware link prediction in evolving social networks. In *Proceedings of the 3rd workshop on social network mining and analysis* (p. 9). ACM.
- [۱۸] Wang, C., Satuluri, V., & Parthasarathy, S. (2007, October). Local probabilistic models for link prediction. In *Seventh IEEE international conference on data mining (ICDM 2007)* (pp. 322-331). IEEE.
- [20] Song, H. H., Cho, T. W., Dave, V., Zhang, Y., & Qiu, L. (2009, November). Scalable proximity estimation and link prediction in online social networks. In *Proceedings of the 9th ACM SIGCOMM conference on Internet measurement* (pp. 322-335). ACM.
- [21] Munasinghe, L., & Ichise, R. (2011, August). Time aware index for link prediction in social networks. In *International Conference on Data Warehousing and Knowledge Discovery* (pp. 342-353). Springer, Berlin, Heidelberg.
- [22] da Silva Soares, P. R., & Prudêncio, R. B. C. (2012, June). Time series based link prediction. In *The 2012 international joint conference on neural networks (IJCNN)* (pp. 1-7). IEEE.
- [23] Zhang, J., & Philip, S. Y. (2014). Link prediction across heterogeneous social networks: A survey. *Social networks*.
- [24] Ibrahim, N. M. A., & Chen, L. (2015). Link prediction in dynamic social networks by integrating different types of information. *Applied Intelligence*, 42(4), 738-750.
- [25] Han, X., Wang, L., Farahbakhsh, R., Cuevas, Á., Cuevas, R., Crespi, N., & He, L. (2016). CSD: A multi-user similarity metric for community recommendation in online social networks. *Expert Systems with Applications*, 53, 14-26.
- [26] Murata, T., & Moriyasu, S. (2007, November). Link prediction of social networks based on weighted proximity measures. In *Proceedings of the IEEE/WIC/ACM international conference on web intelligence* (pp. 85- 88). IEEE Computer Society.
- [27] Newman, M. E. (2001). Clustering and preferential attachment in growing networks. *Physical review E*, 64(2), 025102.
- [28] Adamic, L. A., & Adar, E. (2003). Friends and neighbors on the web. *Social networks*, 25(3), 211-230.
- [29] Güneş, İ., Gündüz-Öğüdücü, Ş., & Çataltepe, Z. (2016). Link prediction using time series of neighborhood-based node

پی‌نوشت:

^۱ Lada A. Adamic