



فرماندهی کل قوا

ستاد کل نیروهای مسلح

دانشگاه و پژوهشگاه عالی دفاع ملی و تحقیقات راهبردی

پژوهشکده آماد، فناوری دفاعی و پدافند غیرعامل



تاریخ ارائه:

شماره:

«طرح همکاران تحقیقاتی و نخبگی جایگزین خدمت سربازی»

فارسی: پیش‌بینی پیوند موضوعی پژوهش‌های علمی با استفاده از تحلیل شبکه‌ی مقالات علمی	عنوان طرح پژوهشی
انگلیسی: Academic Researches Topical Link Prediction Using Academic Articles Network Analysis	

گروه: آماد و فناوری دفاعی
میز تخصصی: رصد و پایش علم، فناوری و نوآوری دفاعی و امنیتی
استاد راهنما: دکتر حنیف کازرونی
استاد راهنما دوم: دکتر مسعود رهگذر (دانشیار، دانشکده فنی دانشگاه تهران)
مجری طرح: رضا شکرچیان چالشتی
طبقه‌بندی طرح: عادی

«طرح پژوهشی کسر خدمت»

(۱) قسمت اول: اطلاعات مربوط به مجری

نام و نام خانوادگی: رضا شکرچیان چالشتی
 نام پدر: جنگیز
 سال تولد: ۱۳۷۰
 شماره ملی: ۴۶۱۰۲۱۶۴۰۱
 شماره تلفن همراه: ۰۹۱۳۱۸۵۴۹۵۰
 شماره شناسنامه و محل صدور: شهرکرد
 فارغ التحصیل مقطع: کارشناسی ارشد
 عنوان رشته: مهندسی نرم افزار
 دانشگاه: تهران
 آدرس محل کار: تهران. امیرآباد. کارگر شمالی. جلال آل احمد. قبل از پل گیشا. خیابان جنت. نبش کوچه اول پلاک ۲
 آدرس منزل: تهران. جمalzاده جنوبی. کوچه شعله ور. بن بست ۴ پلاک ۲
 (۳) سوابق تحصیلات دانشگاهی و تخصصی مجری (به ترتیب مدارج علمی):

ردیف	درجه تحصیلی	رشته تحصیلی	مؤسسه، دانشکده، دانشگاه	کشور	سال شروع	سال اخذ مدرک	معدل
۱	کارشناسی	مهندسی نرم افزار	دانشگاه تهران	ایران	۸۸	۹۲	۱۷,۱۵
۲	کارشناسی ارشد	مهندسی نرم افزار	دانشگاه تهران	ایران	۹۲	۹۵	۱۷,۴۰

(۴) سوابق شغلی و تجربی مجری طرح (به ترتیب از حال حاضر به ذیل):

ردیف	نام مؤسسه/سازمان	سمت و نوع کار	نام مسئول مستقیم بالاتر	از تاریخ	تا تاریخ	شهر	کشور
۱	پژوهشکده سامانه های هوشمند کاربردی دانشگاه تهران	توسعه دهنده	سجاد بهمنی	۹۶/۸	تاکنون	تهران	ایران
۲	شرکت تپسی	تحلیل گر داده	دکتر مهینی	۹۵/۱۱	۹۶/۸	تهران	ایران
	شرکت زعیم	توسعه دهنده	محمد عباسی	۹۴/۱۱	۹۵/۱۱	تهران	ایران

(۵) فعالیت های پژوهشی مجری:

مشخصات کامل تحقیقات و انتشارات مهم مجری پروژه (در صورت امکان نسخه ای را به پیوست ارائه فرمایید):

الف) مشخصات مقالات چاپ شده در نشریات

ردیف	عنوان مقاله	نام نشریه	نام سازمان متولی نشریه	تاریخ انتشار	درجه علمی نشریه	نام نویسنده همکار در تدوین مقاله
۱						

ب) مشخصات مقالات ارائه شده در همایش ها و نشست های علمی

ردیف	عنوان مقاله	عنوان	نام سازمان	نوع همایش	محل برگزاری	تاریخ	نام نویسنده
------	-------------	-------	------------	-----------	-------------	-------	-------------

ردیف	همایش یا نشست علمی	متولی برگزاری	بین‌المللی	ملی	سازمان ی (داخلی)	شهر	کشور	برگزاری	همکار در تدوین مقاله
۱	بهبود مدل‌سازی موضوعی اسناد با استفاده از دانش نهفته در مجموعه سندی	کنفرانس کامپیوتر	IPM	بین‌المللی		تهران	ایران		ناصر یزدانی
۲									

ج) مشخصات کتاب‌های چاپ شده

ردیف	عنوان کتاب	نام ناشر	سال انتشار چاپ اول	تیراژ	شماره آخرین تجدید چاپ	سال آخرین تجدید چاپ	نام نویسنده همکار
۱							
۲							

د) طرح‌های پژوهشی در دست اجرا یا خاتمه یافته

ردیف	عنوان پروژه / طرح پژوهشی	نام مؤسسه / سازمان کارفرما	نام همکاران	تاریخ شروع	تاریخ خاتمه	محل اجرای پروژه	ملاحظات

امضا: مجری

امضا: استاد راهنما

(۱) عنوان طرح :

پیش‌بینی پیوند موضوعی پژوهش‌های علمی با استفاده از تحلیل شبکه‌ی مقالات علمی

Academic Researches Topical Link Prediction Using Academic Articles Network Analysis

(۲) تعریف و بیان مسأله، ضرورت و اهمیت انتخاب و انجام پروژه:

تعریف:

هدف این تحقیق پیش‌بینی پیوند موضوعی پژوهش‌های علمی به کمک تحلیل شبکه‌های اجتماعی ساخته شده از مقالات علمی می‌باشد. برای پیش‌بینی پیوند موضوعی می‌توان از پیش‌بینی همکاری‌های علمی میان محققان و موضوعاتی که بر روی آن تمرکز دارند استفاده کرد. همچنین می‌توان موضوعات مقالات و ارتباط میان آن‌ها را مدل‌سازی کرد و از مدل‌سازی انجام شده برای پیش‌بینی روند ترکیب موضوعات و زمینه‌های علمی در آینده بهره برد. این پیش‌بینی‌ها را می‌توان با به کارگیری روش‌های پیش‌بینی لینک در شبکه‌های اجتماعی ساخته شده از مقالات و مدل‌سازی موضوعی اسناد انجام داد.

شبکه‌های اجتماعی، کاربردها و مدل‌سازی‌های آن امروزه در زمینه‌های مختلفی استفاده می‌شوند و مورد توجه بسیاری از محققان قرار گرفته است. منظور از شبکه‌های اجتماعی فقط مواردی همچون توئیتر، فیسبوک، پیام‌رسان‌ها و ... نیست و هر شبکه‌ای که عناصر آن به طریقی به یکدیگر مرتبط شده‌اند را شامل می‌شود. به طور مثال این شبکه‌ها می‌توانند از ارتباط میان کاربران، نویسندگان، سندها و مقالات علمی، کالاها، پروتئین‌ها و ... حاصل شوند. این شبکه‌ها می‌توانند برخط باشند و ارتباط میان انسان‌ها مانند رابطه‌ی دوستی را شکل دهند. و یا شبکه‌های حاصل از همکاری انسان‌ها باشند و به نوعی ارتباطات موجود در یک حرفه را شامل شوند. مانند همکاری در نگارش و چاپ مقالات علمی، شبکه‌ی ارتباطات پروتئین‌ها در سیستم‌های بیولوژی، شبکه‌های کامپیوتری همچون اینترنت از دیگر مثال‌ها می‌باشند. این سیستم‌ها را می‌توان در قالب گرافی نمایش داد که در آن نودهای گراف موجودیت‌ها و بازیگران شبکه می‌باشند و لینک‌ها نشان‌گر ارتباط میان نودها، همکاری‌ها و تاثیر بازیگران بر یکدیگر می‌باشند. اکثر شبکه‌ها ویژگی‌های ساختاری مشابهی دارند. مانند تنگی و چگالی کم ارتباطات، متوسط فاصله‌ی کم میان نودها، توزیع درجه‌ی مشابه نودها، وجود اجتماع‌ها در ساختار شبکه و این شبکه‌ها معمولاً پویا هستند و روابط و ارتباط‌ها در آن‌ها در حال تغییر است. همچنین اندازه و مقیاس بزرگ این شبکه‌ها از جمله چالش‌های بررسی و مطالعه‌ی آن‌ها می‌باشد.

پیش‌بینی لینک به عمل پیش‌بینی وجود و یا ایجاد لینکی در شبکه در زمان مشخص t با استفاده از اطلاعات و تاریخچه‌ی شبکه قبل از زمان t گفته می‌شود.

در یک دسته‌بندی روش‌های پیش‌بینی لینک به دو دسته تقسیم می‌شوند. روش‌های مبتنی بر شباهت و روش‌های مبتنی بر یادگیری. روش مبتنی بر شباهت به محاسبه‌ی شباهت بین دو جفت نودی می‌پردازد که به یکدیگر لینک ندارند. هر جفت نود (X, Y) که پتانسیل ارتباط دارند امتیازی می‌گیرند که امتیاز بالاتر به معنای بیشتر بودن احتمال شکل گرفتن ارتباط بین X و Y می‌باشد. لیست مرتبطی از لینک‌ها و امتیازها به دست می‌آید که لینک‌های با امتیاز بالاتر احتمال وقوع بیشتری در آینده دارند. روش‌های مبتنی بر یادگیری سعی دارند مساله پیش‌بینی لینک را با استفاده از روش‌های طبقه‌بندی حل کنند. روش‌های مرسوم یادگیری ماشین همچون طبقه‌بندها و مدل‌های احتمالی می‌توانند برای حل مساله به کار گرفته شوند. هر جفت نودی که لینکی بین آن‌ها نیست به عنوان یک

نمونه با مجموعه‌ای از ویژگی‌ها و برچسب طبقه می‌باشد. اگر که پتانسیل ارتباط بین دو نود وجود داشته باشد این جفت نود برچسب مثبت و در غیر این صورت برچسب منفی دریافت می‌کنند.

دو محقق در شبکه‌ی حاصل از ارتباطات مقالات از طریق‌های مختلفی می‌توانند با یکدیگر ارتباط داشته باشند. به طور مثال آن‌ها می‌توانند به یکدیگر لینک داشته باشند اگر مقاله‌ی مشترک داشته باشند و یا در یک ژورنال یا کنفرانس شرکت کرده باشند. راه ارتباطی دیگر می‌تواند بر اساس مقالاتی باشد که به آن‌ها ارجاع داده‌اند. می‌توان شبکه‌های مختلفی بر اساس تعاریف مختلف از ارتباط میان افراد ساخت که هر کدام در پیش‌بینی لینک و ارتباطات آینده می‌توانند اطلاعات مجزا و مفیدی در بر داشته باشند. شبکه‌هایی که نودها در آن‌ها یکسان می‌باشند ولی یال‌های متفاوتی دارند. این شبکه‌ها می‌توانند جداگانه بررسی شوند و امتیاز ایجاد لینک‌هایی که وجود ندارند محاسبه شوند. این امتیازها می‌توانند جداگانه استفاده شوند و یا از ترکیب آن‌ها برای پیش‌بینی ایجاد لینک استفاده شود. روش‌ها و الگوریتم‌های متفاوتی وجود دارند که لینک‌های آینده را با استفاده از ساختار شبکه و گراف ساخته شده از ارتباطات پیش‌بینی می‌کنند.

مدل‌سازی موضوعی اسناد مجموعه‌ای از الگوریتم‌هایی می‌باشد که کمک می‌کنند تا اطلاعات موضوعی در حجم عظیمی از اسناد کشف و تفسیر شوند. بسیاری از این الگوریتم‌ها روش‌های آماری‌ای هستند که با تحلیل کلمات متن‌ها تلاش دارند تا زمینه‌ها و موضوعاتی که در متن‌ها نهفته است را کشف کرده و به بررسی چگونگی ارتباط این موضوعات با هم و یا تغییرات آنها در طول زمان بپردازند. مجموعه‌های مورد بررسی معمولاً بدون ساختار هستند و این روش‌ها کمک می‌کنند تا اسناد به صورت موضوعی سازمان‌دهی شوند. این الگوریتم‌ها نیازی به تفسیر و برچسب اولیه بر روی اسناد ندارند و با تکیه بر محتوای اسناد، موضوعات آنها استخراج می‌شود. حجم اطلاعات به قدری زیاد است که برای سازمان‌دهی و موضوع‌بندی اسناد نمی‌توان به نیروی انسانی اتکا کرد. این روش‌ها مدل‌سازی موضوعات در زمینه‌های مختلفی استفاده می‌شود. از جمله‌ی این زمینه‌ها می‌توان به متن‌کاوی، تکنولوژی‌های جستجو، تکنولوژی‌های نرم‌افزاری، بینایی ماشین، بیوانفورماتیک و اقتصاد اشاره کرد.

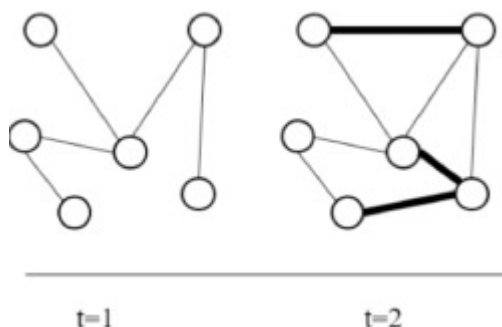
بسیاری از داده‌هایی که این روش‌ها تولید می‌شوند، برخلاف آنچه در پایگاه‌های داده‌ای وجود دارد، بدون ساختار هستند و معمولاً ارتباط معناداری با یکدیگر ندارند. کشف ساختار نهان و ارتباط معنایی میان مجموعه‌ها، اسناد و کلمات، می‌تواند به ساختاربخشی اطلاعات موجود کمک کند. از این ساختار ایجاد شده می‌توان برای گروه‌بندی موضوعی اسناد، راحت‌تر شدن بررسی اطلاعات و بالا بردن نتایج جستجو استفاده کرد. مدل‌سازی موضوعی از جمله روش‌هایی است که تلاش در استخراج ارتباط‌های معنایی میان کلمات دارد. روش‌های مختلفی برای نمایش‌سازی موضوعات وجود دارد. ساده‌ترین آن‌ها نشان دادن هر موضوع به صورت دسته‌ای از کلمات می‌باشد. تعیین کردن این‌که چه کلماتی در یک دسته ظاهر شوند بر عهده‌ی الگوریتم‌های مدل‌سازی موضوع می‌باشد. کلمات موضوع به هم مرتبط می‌باشند و اگر موضوع از انسجام کافی برخوردار باشد، کاربر می‌تواند بفهمد که این کلمات راجع به چه زمینه‌ای صحبت می‌کنند. برای درک بهتر موضوعات می‌توان نحوه‌ی نمایش را گسترش داد و از عبارت‌ها و موجودیت‌ها برای تفهیم و تعیین موضوعات کمک گرفت. در روش‌های مدل‌سازی موضوعی، موضوع‌ها بر روی مجموعه‌ای از اسناد به دست می‌آیند. همچنین از دیگر خروجی این روش‌ها توزیع موضوعات در هر سند می‌باشد. این‌گونه می‌توان متوجه شد که هر سند بیشتر راجع به چه موضوعاتی مطلب دارد و قدمی رو به جلو برای پیدا کردن سندهایی که در زمینه‌ی مورد علاقه هستند برداشته می‌شود. همچنین کلماتی که در یک موضوع هستند به هم مرتبط می‌باشند. می‌توان از این ارتباط در کاربردهایی همچون جستجوی اسناد استفاده کرد و به جای نگاهی کلمه‌ای به اسناد، نگاهی مفهومی و موضوعی به آن‌ها داشت.

به کمک مدل‌سازی موضوعی می‌توان استخراج کرد که در یک مجموعه از اسناد، راجع به چه موضوعاتی صحبت شده و هر سند در مجموعه‌ی اسناد به چه میزان مرتبط به هر موضوع می‌باشد. یکی از روش‌های مرسوم در نمایش موضوعات اسناد استفاده از لیست مرتب کلمات است. موضوع هر سند در غالب لیست مرتبی از کلمات با عددی نسبت داده شده به هر کلمه ارائه می‌شود. عدد احتمال تعلق کلمه به موضوع را نشان می‌دهد

بیان مساله:

هدف این تحقیق پیشنهاد و پیاده‌سازی روشی برای پیش‌بینی پیوند موضوعی در پژوهش‌های علمی با استفاده از اطلاعات و مجموعه داده‌ی ساخته شده از مقالات علمی می‌باشد. فرض این تحقیق بر آن است که مقالات منتشر شده حاوی اطلاعات با ارزشی از پژوهش‌ها و روند علمی روز دنیا می‌باشند و تحلیل پیوند موضوعی میان انبوه پژوهش‌های علمی شده می‌تواند دانش خوبی از اتفاقات علمی و موضوعاتی که کشورها بر روی آن تمرکز دارند و یا خواهند داشت، بدهد. از روش‌های پیش‌بینی لینک در شبکه‌های اجتماعی و مدل‌سازی موضوعی اسناد برای پیش‌بینی پیوند موضوعی پژوهش‌های علمی استفاده می‌شود.

شکل زیر مثال ساده‌ای از پیش‌بینی لینک می‌باشد. لینک‌های تیره لینک‌هایی هستند که در آینده ایجاد می‌شوند.



برای پیش‌بینی پیوند موضوعی دو نوع گراف از مجموعه اسناد مقالات بررسی می‌شود:

۱- **گرافی حاصل از ارتباط محققان:** در این گراف نودها محققان می‌باشند و قرار است ارتباط‌های آن‌ها در آینده پیش‌بینی شود. با بررسی موضوعاتی که هر محقق در آن فعالیت دارد می‌توان دانش مفیدی از فعالیت‌ها و موضوعات علمی فعلی و آتی به دست آورد. برای رسیدن به این هدف می‌توان گراف ارتباط محققان را ساخته و به کمک روش‌های پیش‌بینی لینک ارتباط محققان را پیش‌بینی کرد. همچنین مقالات هر محقق را گردآوری کرده و موضوعاتی که محقق بر روی آن‌ها فعالیت دارد را مدل‌سازی کرد. به کمک این دو روش می‌توان ارتباط موضوعی تحقیقات را پیش‌بینی کرد.

۲- **گراف حاصل از ارتباط موضوعات:** در این گراف مستقیماً مقالات و موضوعات آن‌ها مورد بررسی قرار می‌گیرند. اسناد و مقالات مدل‌سازی موضوعی می‌شوند. از شباهت موضوعی اسناد برای برقراری ارتباط و ساخت لینک استفاده می‌شود. در این گراف نودها موضوعات می‌باشند. پیش‌بینی ارتباط موضوعات در آینده کمک می‌کند دانش مفیدی از ارتباط موضوعات و روند ترکیب آن‌ها استخراج شود.

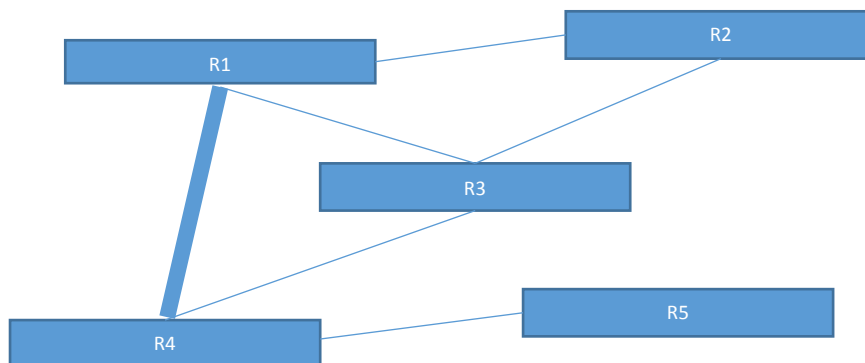
برای ساخت گراف همکاری‌های علمی از ارتباط‌های مختلفی همچون مقاله‌ی مشترک، مقالات ارجاع شده، موضوعات شبیه به هم و ... استفاده می‌شود. برای ساخت گراف حاصل از ارتباط موضوعات از مدل‌سازی موضوعی اسناد استفاده می‌شود.

در این تحقیق قرار است الگوریتم‌های مختلف پیش‌بینی لینک مطالعه و تحلیل شوند. زیرمجموعه‌ای از این روش‌ها انتخاب و پیاده‌سازی شوند. گراف‌های مختلفی از اسناد ساخته می‌شود که در هر گراف ارتباط میان نودها معنای متفاوتی دارد و حاوی اطلاعات مجزایی می‌باشد. مثلاً در یک گراف ارتباط میان دو نود به معنای مقاله‌ی مشترک داشتن دو محقق می‌باشد و در گراف دیگر ارتباط آن‌ها به معنای داشتن موضوع مشابه می‌باشد. نتایج عملکرد روش‌های مختلف پیش‌بینی لینک بر روی گراف‌های ساخته شده ارزیابی می‌شود. هر روش خروجی‌ای دارد و اطلاعاتی از هر گراف استخراج می‌کند. در ادامه روشی پیشنهاد می‌شود که خروجی‌ها و نتایج حاصل از روش‌های مختلف پیش‌بینی لینک را ترکیب کند و از ارتباطات مختلف میان نودها در پیش‌بینی لینک بهره گیرد. در نهایت برای هر گراف یک رتبه‌بندی پیشنهاد می‌شود که در آن لینک‌های با رتبه‌ی بالاتر با احتمال بیشتری در آینده تشکیل می‌شوند. پس از پیش‌بینی پیوندهای همکاری، باید موضوعاتی که هر شخص بر روی آن فعالیت می‌کند استخراج شود. در این قسمت مقالات محققان باید مدل‌سازی موضوعی شوند و با استفاده از پیوندهای پیش‌بینی شده ارتباط موضوعی پژوهش‌های علمی مدل می‌شود.

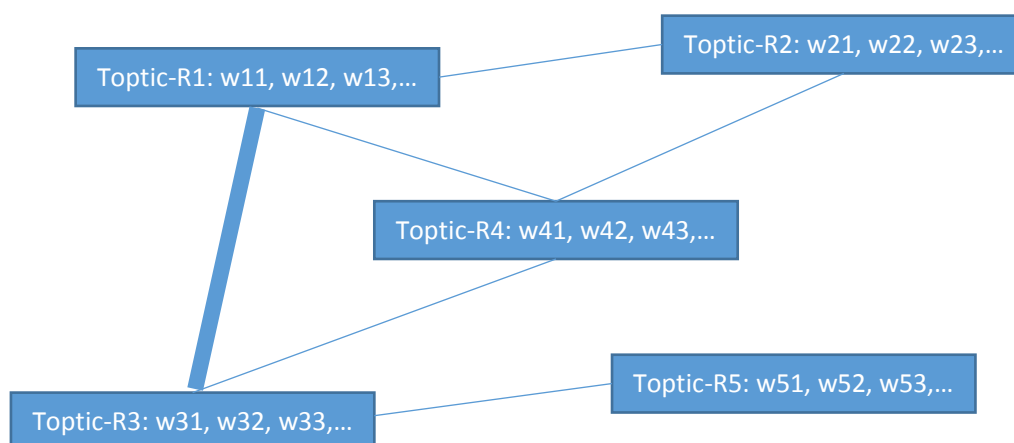
از پیش‌بینی پیوند موضوعی تحقیقات می‌توان در دروندیابی علمی استفاده کرد. برای به دست آوردن روند پژوهش‌های علمی پارامترهای مختلفی دخیل هستند. مانند بودجه‌ی صرف شده، فناوری‌های در دست تولید، مقالات منتشر شده و ... در این تحقیق از

مقالات منتشر شده به عنوان منبع اطلاعاتی استفاده می‌شود و تلاش می‌شود پیوند موضوعی پژوهش‌های علمی پیش‌بینی شود. از خروجی این تحقیق می‌توان در کنار دیگر پارامترها برای تحلیل دقیق‌تر دنیای دانش استفاده کرد.

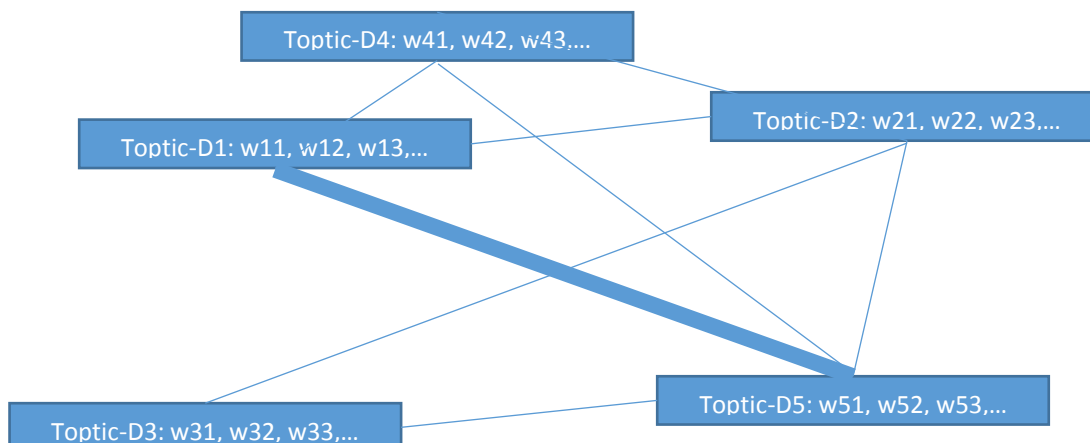
خروجی تحقیق:



یکی از گراف‌هایی که در این تحقیق ساخته می‌شود، گراف همکاری‌هاست و یکی از خروجی‌های این تحقیق پیش‌بینی شکل‌گیری همکاری‌ها بین محققان و رتبه‌بندی امتیاز و احتمال تشکیل همکاری می‌باشد. در این شکل نودها محققان می‌باشد. پیوند با ضخامت بیشتر پیوند پیش‌بینی شده می‌باشد.



برای هر محقق مقالاتی در آن‌ها همکاری داشته مدل‌سازی موضوعی می‌شود و موضوعاتی که در آن‌ها فعالیت داشته مدل می‌شوند. پس از پیش‌بینی ارتباط محققان و مدل‌سازی موضوعات می‌توان موضوعاتی که احتمالاً در آینده با یکدیگر ترکیب می‌شوند را کشف و رتبه‌بندی کرد. در این شکل هر نود موضوعات مدل شده برای یک محقق می‌باشد و لینک‌ها از ارتباط میان موضوعات ساخته می‌شوند.



یکی دیگر از گراف‌هایی که ساخته می‌شود گراف حاصل از موضوعات مقالات می‌باشد. در این روش مقالات مجموعه مدل‌سازی موضوعی می‌شوند و موضوعات هر مقاله مدل می‌شود. ارتباط میان موضوعات با توجه به میزان شباهت آن‌ها شکل می‌گیرد. با استفاده از روش پیش‌بینی لینک، پیوند موضوعی پژوهش‌های علمی کشف می‌شود. در این شکل نودها موضوعات هر مقاله می‌باشند و ارتباط آن‌ها بر اساس میزان شباهت موضوعات می‌باشد.

ضرورت و اهمیت انتخاب و انجام پروژه

پیش‌بینی مقالاتی که در آینده تاثیرگذاری بالایی دارند برای ذینفعانی همچون حکومت، دانشگاه‌ها، محققان و سرمایه‌گذاران حائز اهمیت است. توانایی پیش‌بینی آینده به تخصیص منابع بر روی زمینه‌های تحقیقاتی کمک می‌کند. در دنیای امروز تحقیق و توسعه فضایی رقابتی است که هر روز این رقابت میان کشورها در حال افزایش می‌باشد و هزینه‌های بسیاری صرف آن می‌شود. استفاده از راه‌های بهینه و خلاقانه کمک می‌کند از منابع موجود در این فضای رقابتی به درستی استفاده شود. ممکن است این امکان برای کشور وجود نداشته باشد که با صرف هزینه‌های بسیار برای تحقیق و توسعه در دنیا پیشرو باشد و اولین کشوری باشد که به یک علم و فناوری دست پیدا می‌کند. ولی با استفاده از تحلیل داده‌ها و پیش‌بینی روند علمی می‌توان امکان حرکت سریع و زود هنگام به سمت علوم و فناوری‌های مهم دنیا را فراهم کرد. تقویت قدرت تصمیم‌گیری و سیاست‌گذاری صحیح از مزایای این تحلیل‌ها می‌باشد. یکی از روش‌های پیش‌بینی روند علمی، پیش‌بینی همکاری‌های علمی و پیش‌بینی ترکیب موضوعات علمی با استفاده از داده‌های مقالات علمی می‌باشد. این پیش‌بینی می‌تواند با استفاده از روش‌های پیش‌بینی لینک بر روی شبکه‌ی حاصل از ارتباطات علمی افراد و ارتباطات موضوعات حاصل شود.

(۳) نوع تحقیق:

این تحقیق از نوع کاربردی است. مجموعه‌داده مقالات به عنوان ورودی این تحقیق می‌باشد. با استفاده از مجموعه داده و روش پیش‌بینی لینک و مدل‌سازی موضوعی، پیوند موضوعی پژوهش‌های علمی آینده پیش‌بینی می‌شوند.

(۴) اهداف تحقیق:

هدف اصلی این تحقیق پیشنهاد، پیاده‌سازی و ارزیابی روشی برای پیش‌بینی پیوند موضوعی پژوهش‌های علمی می‌باشد. از دو روش مدل‌سازی اسناد و پیش‌بینی لینک در شبکه‌ی اجتماعی حاصل از مقالات علمی برای پیش‌بینی موضوعی پژوهش‌های علمی استفاده می‌شود. برای رسیدن به این هدف باید از ارتباط‌های موجود میان مقالات علمی به درستی استفاده شود تا گراف‌ها ساخته شوند. ارتباط‌های ناموجود در گراف با روش‌های پیش‌بینی لینک استخراج می‌شوند. همچنین اسناد و مقالات باید مدل‌سازی موضوعی شوند و موضوعی که هر مقاله بر روی آن تمرکز دارد استخراج شود. با استفاده از مدل موضوعی ساخته شده از اسناد و ارتباط‌های پیش‌بینی شده در گراف‌های پیشنهاد شده، می‌توان پیوندهای موضوعی پژوهش‌های علمی را مدل‌سازی و پیش‌بینی کرد.

اهداف فرعی:

- ۱- تحقیق و مطالعه ساختارها، مدل‌ها و ویژگی‌های گراف‌های مختلف
- ۲- تحلیل و مطالعه روش‌های پیش‌بینی لینک در شبکه‌های اجتماعی، مقایسه مزایا و معایب روش‌ها برای انتخاب درست روش‌ها و توانایی در تحلیل نتایج تحقیق
- ۳- تحقیق و تحلیل روش‌های مدل‌سازی موضوعی اسناد
- ۴- ارائه ساختار کلی (framework) مرتبط با پیش‌بینی پیوند موضوعی پژوهش‌های علمی
- ۵- تشکیل و ساخت گراف و شبکه‌های مختلف همکاری‌ها با استفاده از مقالات علمی و ویژگی‌های آن‌ها (افراد گره‌ها می‌باشند و نحوه‌ی ارتباط افراد لینک‌ها را تشکیل می‌دهد)
- ۶- تشکیل شبکه‌های حاصل از ارتباط موضوعی مقالات. (در این شبکه گره‌ها موضوعات و لینک‌ها ارتباط موضوعات با یکدیگر می‌باشند.)
- ۷- انتخاب زیرمجموعه‌ای مناسب از روش‌های پیش‌بینی لینک. پیاده‌سازی آن‌ها و اعمال آن‌ها بر شبکه‌های ساخته شده
- ۸- بررسی روش‌های ترکیب رتبه‌بندی برای ترکیب نتایج روش‌های پیاده‌سازی شده. پیشنهاد و ساخت روشی برای ترکیب نتایج و رتبه‌بندی‌های حاصل از خروجی روش‌های مختلف. ارائه‌ی امتیازدهی نهایی برای مرتب‌سازی لینک‌های پیش‌بینی شده.

۹- تحلیل و ارزیابی نتایج و خروجی تحقیق.

۱۰- پیشنهاد و پیاده‌سازی راه‌کاری برای مقیاس‌پذیر کردن الگوریتم و اضافه کردن قابلیت استفاده از روش برای تعامل با کلان داده

(۵) سؤالات تحقیق:

سوال اصلی این مقاله این است که چگونه می‌توان روند علمی را با استفاده از پیش‌بینی لینک در شبکه‌ی اجتماعی حاصل از مقالات علمی کشف کرد؟

سؤالات فرعی:

- ۱- ساختارها، مدل‌ها و ویژگی‌های گراف‌های مختلف کدامند؟
- ۲- روش‌های پیش‌بینی لینک در شبکه‌های اجتماعی کدامند؟ مزایا و معایب هر روش چیست؟ کدام روش‌ها برای حل مساله‌ی پیش‌بینی لینک در شبکه‌ی مقالات علمی می‌توانند مناسب باشند.
- ۳- روش‌های مدل‌سازی موضوعی اسناد کدام است؟
- ۴- ساختار کلی پیش‌بینی موضوعی پیوند در پژوهش‌های علمی کدام است؟
- ۵- چگونه می‌توان گراف و شبکه‌های مختلف همکاری‌ها را ساخت؟
- ۶- چگونه می‌توان گراف و شبکه‌های اجتماعی حاصل از موضوعات مقالات را ساخت؟
- ۷- چه زیرمجموعه‌ای از روش‌های پیش‌بینی لینک برای حل مساله مناسب می‌باشند. چگونه می‌توان آن‌ها را پیاده‌سازی و بر شبکه‌های ساخته شده اعمال کرد؟
- ۸- چگونه می‌توان رتبه‌بندی و خروجی روش‌های پیاده‌سازی شده را ترکیب کرد؟ چگونه می‌توان روشی برای ترکیب رتبه‌بندی‌ها ساخت؟
- ۹- چگونه می‌توان نتایج و خروجی تحقیق را تحلیل و ارزیابی کرد؟
- ۱۰- چه راه‌کاری برای مقیاس‌پذیر کردن الگوریتم و آماده‌سازی آن برای تعامل با کلان داده پیشنهاد می‌شود؟

(۶) فرضیات تحقیق (در صورت وجود فرضیه‌سازی):

فرض این تحقیق وجود یک مجموعه داده مناسب و تمیز شده برای پیش‌بینی روند علمی می‌باشد.

(۷) قلمرو و محدوده تحقیق (مکانی، زمانی، موضوعی، سازمانی و...):

این تحقیق از لحاظ مکانی و زمانی محدودیت خاصی نداشته و از لحاظ موضوعی متناسب به نوع دادگان می‌باشد. از آنجایی که دادگان درباره مقالات و اطلاعات علمی افراد است، بنابراین محدوده موضوعی این تحقیق، موضوعات علمی و آکادمیک است.

(۸) روش تحقیق (توصیفی، پیمایشی، تاریخی، تجربی، کتابخانه‌ای و...):

در این تحقیق ابتدا مطالعه‌ای بر روی مقالات حوزه‌ی شبکه‌ی اجتماعی، گراف‌ها و پیش‌بینی لینک و مدل‌سازی موضوعی انجام خواهد شد. زیرمجموعه‌ای از روش‌های پیش‌بینی لینک انتخاب می‌شوند. مجموعه داده‌ی فراهم شده پیش‌پردازش می‌شود و گراف‌های مورد نظر با توجه به تعاریف مختلف ارتباط ساخته می‌شوند. دو نوع گراف در این تحقیق بررسی می‌شود. گراف مربوط به همکاری‌های علمی و گراف حاصل از موضوعات علمی و ارتباط آن‌ها با یکدیگر. روش‌های پیش‌بینی لینک انتخاب شده پیاده‌سازی و بر گراف‌ها اعمال می‌شوند. هر کدام از این روش‌ها امتیازی به لینک‌هایی که وجود ندارند اختصاص می‌دهند که این امتیاز پتانسیل ایجاد لینک در آینده را نشان می‌دهد. روش‌های رتبه‌بندی مطالعه و بررسی می‌شوند و روشی برای ترکیب اطلاعات رتبه‌بندی روش‌های مختلف به منظور تشکیل رتبه‌بندی واحد پیشنهاد می‌شود. هرچه امتیاز یک لینک بیشتر باشد احتمال به وجود آمدن آن در آینده بیشتر است. در نهایت خروجی روش با استفاده از معیارهای ارزیابی نتایج پیش‌بینی لینک در شبکه‌های اجتماعی ارزیابی می‌شود. به کمک مدل‌سازی موضوعی، موضوعات مجموعه و اسناد مدل‌سازی می‌شوند. با ترکیب نتایج پیش‌بینی لینک و مدل‌سازی موضوعی، پیوند موضوعی پژوهش‌های علمی مدل‌سازی و پیش‌بینی می‌شوند.

۹) معرفی جامعه آماری و نمونه آماری و ابزار گردآوری داده‌ها (پرسشنامه، مصاحبه، مشاهده و...) و روش نمونه‌برداری (تصادفی، ساده، خوشه‌ای، سلسله مراتبی و...):

مجموعه داده در این پژوهش از نوع داده مربوط پژوهش‌های علمی و دادگان مربوط به مقالات علمی می‌باشد. فرض بر این است که این داده‌ها قبلاً گردآوری شده‌اند.

۱۰) روش تجزیه و تحلیل آماری داده‌های تحقیق و روش اعتبارسنجی داده‌های گردآوری شده:

هدف این تحقیق مدل‌سازی و پیش‌بینی پیوند موضوعی در آینده با استفاده از پیش‌بینی لینک در شبکه و مدل‌سازی موضوعی می‌باشد. خروجی این روش یافتن امتیاز و رتبه‌بندی برای لینک‌هایی که در شبکه وجود ندارند و موضوعات آن‌ها می‌باشد. K لینک با امتیاز بالا به عنوان لینک‌های پیش‌بینی شده معرفی می‌شوند. مترهای مختلفی می‌تواند بر خروجی روش برای ارزیابی عملکرد آن اعمال شوند. از این میان می‌توان به مترهایی همچون Accuracy, Precision, Recall, F1, ROC اشاره کرد. برای ارزیابی نتایج این پژوهش از روش‌های ریاضی و مقایسه با نتایج روش‌های مشابه استفاده خواهد شد.

۱۱) پیشینه تحقیق (با ذکر مشخصات کامل منابع شناسایی شده):

پیش‌بینی لینک در شبکه‌ی اجتماعی، لینک‌های از قلم افتاده و لینک‌های احتمالی‌ای که در آینده رخ خواهند داد را بر اساس وضعیت فعلی شبکه پیش‌بینی می‌کند. این پیش‌بینی در کاوش و تحلیل شبکه بسیار حائز اهمیت است. یک شبکه‌ی اجتماعی شبکه‌ای حاصل از مجموعه‌ای از نقش‌ها، موجودیت‌ها، سندها می‌باشد که ارتباط و تعامل میان آن‌ها به صورت لینک در شبکه مشخص می‌شود. می‌توان این شبکه را در قالب گرافی نمایش داد. دو چالش اصلی در تحلیل شبکه‌های اجتماعی وجود دارد: ناقص بودن و پویایی شبکه. داده‌ها و اطلاعاتی که از شبکه به دست می‌آیند کامل نیستند چرا که گستردگی نقش‌ها و ارتباطات در شبکه‌های اجتماعی بسیار بالا می‌باشد و ساختن گرافی کامل مطابق با واقعیت بسیار هزینه‌بر می‌باشد. خاصیت دیگر شبکه‌های اجتماعی پویایی آن‌ها می‌باشد که باعث می‌شود نودها و ارتباط‌هایی در آینده به شبکه اضافه شوند و یا حذف گردند. بنابراین پیش‌بینی ارتباط‌های از قلم افتاده در شبکه‌ی فعلی و لینک‌هایی که در شبکه‌ی آتی اضافه و یا حذف می‌شوند بسیار با اهمیت می‌باشد. این اهمیت به خاطر پیش‌بینی سیر تکاملی شبکه و روند آتی ارتباطات در شبکه می‌باشد. این نیازمندی مسأله‌ای به نام پیش‌بینی لینک را مطرح می‌کند. پیش‌بینی لینک کاربردهای فراوانی دارد. یکی از کاربردهای استفاده در سیستم‌های توصیه‌گر بر اساس بازیابی اطلاعات می‌باشد. به طور مثال پیشنهاد همکاری‌های جدید بین اشخاص [۱، ۲]، یافتن خبرگان و همکاران در شبکه‌ی اجتماعی دانشگاهی [۳، ۴]، پیشنهاد همکاری‌های بین رشته‌ای [۵]، پیشنهاد دوستی در شبکه‌های اجتماعی [۶] و یا پیشنهاد کالا در خرید آنلاین از جمله کاربردهای این روش می‌باشد. کاربرد دیگر آن استنتاج شبکه‌ی کامل بر اساس مشاهدات ناقص از شبکه [۷، ۸]، فهم بهتر سیر تکاملی شبکه‌ها [۹، ۱۰، ۱۱، ۱۲]. همچنین در حوزه امنیت می‌تواند در شناسایی ارتباط‌های غیرعادی استفاده شود [۱۳].

در یک دسته بندی روش‌های پیش‌بینی لینک را می‌توان به دو روش بی‌ناظر و باناظر تقسیم‌بندی کرد.

روش‌های بی‌ناظر:

روش‌های بسیاری در دسته‌ی بی‌ناظر قرار می‌گیرند. در پیش‌بینی لینک، امتیاز لینک برای جفت نودهایی که به یکدیگر لینکی ندارند بر اساس ساختار شبکه محاسبه می‌شود. این امتیاز به نوعی نمایانگر میزان شباهت میان دو نود در شبکه می‌باشد و مشخص می‌کند امکان وجود لینک میان دو نود تا چه اندازه ممکن می‌باشد.

[۱۴] بر روی شبکه‌ی همکاری نویسندگان کار کرده است. در این مقاله نشان داده شده است که ویژگی‌های ساختاری می‌توانند بیانگر ارتباط میان جفت نودهایی باشد که لینکی بین آن‌ها نیست. در این مقاله آزمایشاتی بر روی ویژگی‌های مختلف ساختار شبکه انجام شده است که از جمله‌ی آن‌ها می‌توان به ویژگی‌های مبتنی بر شباهت نودها اشاره کرد. یکی از ویژگی‌های ساختاری می‌تواند تعداد همسایه‌ی مشترک باشد. تعداد همسایه‌ی مشترک تعداد نودهایی است که مستقیماً به دو نود متصل باشند. در [۱۵] این ویژگی در شبکه‌ی همکاری‌ها و در [۱۶] برای شبکه‌های با مقیاس بالا استفاده شده است.

از دیگر ویژگی‌های ساختاری، ویژگی‌های مبتنی بر مسیر می‌باشند. این ویژگی بر پایه‌ی مسیری می‌باشد که بین جفت نود بدون لینک وجود دارد. اندازه‌ی آن می‌تواند طول مسیر بین دو نود باشد. یک ایده این است که هر چه اندازه‌ی مسیر بین دو نود کمتر باشد، دو نود بیشتر به هم شبیه می‌باشند. اندازه‌ی مسیر بین دو نود می‌تواند به روش قدم‌زنی تصادفی نیز تعیین شود. دسته‌ی دیگر روش‌های ترکیبی می‌باشند. این دسته این ایده را مطرح می‌کند که دو نود می‌توانند مشابه باشند اگر ویژگی‌های ساختاری مشابهی داشته باشند. تک ویژگی‌های ساختاری می‌توانند ترکیب شوند تا ویژگی جدید را ایجاد کنند. در [۱۷] از مجموع همسایه‌ها به عنوان یک ویژگی جدید استفاده کرده است.

روش‌های باناظر:

در [۱۴] تلاش‌هایی برای ترکیب تاثیر تک ویژگی‌های ساختاری به منظور بهبود عملکرد پیش‌بین انجام شده است. بیشتر این تلاش‌ها در قالب روش‌ها یادگیری ماشین می‌باشند. در [۱۷] مساله‌ی پیش‌بینی لینک به مساله‌ی طبقه‌بندی دودویی تبدیل شده است. که نمونه‌ها جفت نودهای بدون لینک می‌باشند و ویژگی‌ها از مقادیر ویژگی‌های ساختاری به دست می‌آیند. [۱۸] یک روش پیش‌بینی لینک مبتنی بر یادگیری ماشین باناظر می‌باشد که از درخت تصمیم برای پیش‌بینی لینک استفاده می‌کند. در [۱۹] از روش فاکتوریزه کردن ماتریس برای پیش‌بینی لینک استفاده می‌کند. مدل ویژگی‌های نهان شبکه را از ساختار گراف یاد می‌گیرد. در روش‌های احتمالاتی معمولاً از مفاهیم بیزین استفاده می‌شود تا احتمال هم‌اتفاقی را بین دو نود جدا حساب کند [۲۰]. هم‌اتفاقی‌های با احتمال بالا امکان تشکیل لینک در آینده دارند. در جدول زیر چند مقاله‌ی موثر در روش پیشنهادی پیش‌بینی لینک معرفی شده‌اند.

مقاله	رویکرد
The link-prediction problem for social networks	استفاده از ویژگی‌های ساختاری شبکه همچون شباهت نودها برای پیش‌بینی لینک
Link prediction using supervised learning	تبدیل مساله پیش‌بینی لینک به مساله‌ی طبقه‌بندی دودویی. استفاده از ویژگی‌های ساختاری به عنوان ویژگی‌های مجموعه داده
A supervised machine learning link prediction approach for academic collaboration recommendation	استفاده از درخت تصمیم برای پیش‌بینی لینک
Local probabilistic models for link prediction.	مدل‌سازی احتمالی شبکه و پیش‌بینی لینک با استفاده از تخمین احتمال هم‌اتفاقی دو نود در شبکه

- [1] Mori J, Kajikawa Y, Kashima H, et al. Machine learning approach for finding business partners and building reciprocal relationships. *Expert Systems with Applications*, 2012, 39: 10402–10407
- [2] Wu S, Sun J, Tang J. Patent partner recommendation in enterprise social networks. In: *Proceedings of the 6th ACM International Conference on Web Search and Data Mining (WSDM'13)*, Rome, Italy, 2013. 43–52
- [3] Pavlov M, Ichise R. Finding experts by link prediction in co-authorship networks. In: *Proceedings of the 2nd International ISWC+ASWC Workshop on Finding Experts on the Web with Semantics (FEWS)*, Busan, Korea, 2007. 42–55
- [4] Wohlfarth T, Ichise R. Semantic and event-based approach for link prediction. In: *Proceedings of the 7th International Conference on Practical Aspects of Knowledge Management (PAKM'08)*, Yokohama, Japan, 2008. 50–61
- [5] Tang J, Wu S, Sun J M, et al. Cross-domain collaboration recommendation. In: *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'12)*, Beijing, China, 2012. 1285–1293
- [6] Aiello L M, Barrat A, Schifanella R, et al. Friendship prediction and homophily in social media. *ACM Transactions on the Web*, 2012, 6: 9
- [7] Marchette D J, Priebe C E. Predicting unobserved links in incompletely observed networks. *Computational Statistics & Data Analysis*, 2008, 52: 1373–1386
- [8] Kim M, Leskovec J. The network completion problem: inferring missing nodes and edges in networks. In: *Proceedings of the 11th SIAM International Conference on Data Mining (SDM'11)*, Mesa, Arizona, USA, 2011. 47–58
- [9] Barabási A L, Jeong H, Nédá Z, et al. Evolution of the social network of scientific collaborations. *Physica A*, 2002, 311: 590–614
- [10] Juszczyszyn K, Musial K, Budka M. Link prediction based on subgraph evolution in dynamic social networks. In: *Proceedings of the 2011 IEEE International Conference on Social Computing*, Boston, USA, 2011. 27–34
- [11] Bringmann B, Berlingerio M, Bonchi F, et al. Learning and predicting the evolution of social networks. *IEEE Intelligent Systems*, 2010, 25: 26–35
- [12] Raymond R, Kashima H. Fast and scalable algorithms for semi-supervised link prediction on static and dynamic graphs. In: *Proceedings of ECML/PKDD'10*, Barcelona, Spain, 2010. 131–147
- [13] Huang Z, Lin D K J. The time-series link prediction problem with applications in communication surveillance. *INFORMS Journal on Computing*, 2009, 21: 286–303
- [14] David Liben-Nowell and Jon M Kleinberg. The link-prediction problem for social networks. *JASIST*, 58(7):1019–1031, 2007.
- [15] M. E. J. Newman. Coauthorship networks and patterns of scientific collaboration. *Proceedings of the National Academy of Science of the United States (PNAS)*, 101:5200–5205, 2004a.
- [16] G. Kossinets. Effects of missing data in social networks. *Social Networks*, 28(3):247–268, 2006. ISSN 03788733. doi: 10.1016/j.socnet.2005.07.002.
- [17] Mohammad Al Hasan, Vineet Chaoji, Saeed Salem, and Mohammed Zaki. Link prediction using supervised learning. In *Workshop on link analysis, Counter-terrorism and security*, SIAM Data Mining Conference, Bethesda, MD, 2006.
- [18] Nesserine Benchettara, Rushed Kanawati, and Céline Rouveirol. A supervised machine learning link prediction approach for academic collaboration recommendation. In *Proceedings of the fourth ACM conference on Recommender systems - RecSys '10*, pages 253–256, New York, New York, USA, 2010b. ACM Press. ISBN 9781605589060. doi: 10.1145/1864708.1864760. URL <http://dblp.uni-trier.de/db/conf/recsys/recsys2010.html#BenchettaraKR10>.
- [19] Aditya Krishna Menon and Charles Eklun. Link prediction via matrix factorization. In *Dimitrios Gunopulos, Thomas Hofmann, Donato Malerba, and Michalis Vazirgiannis, editors,*

Machine Learning and Knowledge Discovery in Databases, volume 6912 of Lecture Notes in Computer Science, pages 437–452. Springer Berlin Heidelberg, 2011.

[20] Chao Wang, Venu Satuluri, and Srinivasan Parthasarathy. Local probabilistic models for link prediction. In Yong Shi and Christopher W Clifton, editors, Seventh IEEE International Conference on Data Mining (ICDM), pages 322–331. IEEE, October 2007.

۱۲) کلیدهای واژه‌ها (همراه با معادل انگلیسی آن)

شبکه‌های اجتماعی (Social Networks), پیش‌بینی لینک (Link Prediction), مجموعه داده علمی (Academic Dataset), مدل‌سازی موضوعی (Topic Modeling)

۱۳) معرفی مراحل انجام و گام‌های تحقیق و ارائه تاریخ پیشنهادی و جدول زمان‌بندی اجرای پروژه و تهیه نمودار کانت و ... (به شرح جدول زیر):

ردیف	گام‌های اجرای پروژه	درصد گام	پیش‌بینی ساعات	زمان اجرا (ماه)													
				۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰	۱۱	۱۲	۱۳	۱۴
۱	مطالعه مقالات و آشنایی با مفاهیم گراف، شبکه‌های اجتماعی			■	■												
۲	آشنایی با مفاهیم پیش‌بینی لینک در شبکه‌های اجتماعی				■	■											
۳	مطالعه، تحقیق و آشنایی با مدل‌سازی موضوعی					■	■										
۴	گردآوری مجموعه داده مناسب برای پژوهش					■											
۵	ساخت گراف‌ها و پیش‌پردازش و تمیز کردن مجموعه داده						■	■									
۶	پیاده‌سازی الگوریتم‌های مناسب برای پژوهش							■	■								
۷	بدست آوردن روش نهایی به منظور پیش‌بینی همکاری‌ها							■	■								
۸	انجام آزمایش‌های مرتبط								■	■							
۹	استخراج و تحلیل نتایج									■	■						
۱۰	ارزیابی نتایج بدست آمده با روش‌های مشابه										■	■					
۱۱	جمع‌بندی و نگارش پایان‌نامه												■	■	■		
جمع																	

۱۴) معرفی ساختار گزارش نهایی پروژه و تعیین سرفصل‌های گزارشی که در انتهای کار تحویل کارفرما خواهد گردید.

خروجی و گزارش کتبی متناسب با گام‌های نمودار گانت تنظیم خواهد شد. در انتهای کار گزارشی نهایی از نتایج تحقیق و عملکرد روش ارائه خواهد شد.

امضا: مجری

امضا: استاد راهنما

۱۵) تاییدیه پروپوزال طرح پژوهشی:

عنوان طرح: پیش‌بینی روند علمی با استفاده از روش پیش‌بینی لینک در شبکه‌ی مقالات علمی

مجری طرح: رضا شکرچیان چالشتی

اساتید راهنما:

نام و نام خانوادگی: دکتر حنیف کازرونی، دکتر مسعود رهگذر
امضا:

دبیر گروه:

نام و نام خانوادگی: فرهاد آفریده
امضا:

رئیس گروه:

نام و نام خانوادگی: دکتر حنیف کازرونی
امضا:

رئیس پژوهشکده:

نام و نام خانوادگی: سردار سرتیپ ۲ دکتر مجتبی سعادت
امضا: