علوم رایانش و فناوری اطلاعات
نشریه علمی انجمن کامپیوتر ایران
مجلد ۱۳، شماره ۱، ۱۳۹۴
صفحات ۱-۱۵
مقاله عادی

روش جدید پیش‌بینی لینک در شبکه‌های اجتماعی مبتنی بر مدل جستجوی گرانشی

اسماعیل بسطامی امین‌اله مه‌آبادی

دانشکده فنی و مهندسی، دانشگاه شاهد، تهران، ایران

چکیده

ما در این مقاله روش توزیعی مقیاس‌پذیر جدیدی برای پیش‌بینی دقیق لینک در شبکه‌های اجتماعی با استفاده از ویژگی‌های ساختاری آن شبکه‌ها و بدون هیچ نیازی به سابقه‌گیری ارائه می‌دهیم. این روش ضمن مدل‌سازی عامل‌گرای مساله و انتخاب انجمن‌ها، از الگوریتم جستجوی گرانشی برای تشخیص لینک‌های مناسب بین انجمن‌ها بهره می‌برد. نتایج ارزیابی آزمایش‌های تجربی نشان می‌دهد که عملکرد روش پیشنهادی در سناریوهای مجموعه داده‌ای مختلف شبکه، مقیاس‌پذیر و از دقت میانگین ۶۹ درصد و صحت میانگین ۶۸ درصد برخوردار است. ضمنا با داشتن ساختار توزیعی و با برخورداری از پاسخ زمانی مناسب، در صورت انتخاب بهینه عامل‌ها و تخصیص مناسب پردازنده به آن‌ها، با بهبود زمان پاسخ و افزایش قدرت مقیاس‌پذیری عامل‌ها مواجه خواهد شد.

کلمات کلیدی: مدل عامل‌گرا، شبکه‌های اجتماعی، پیش‌بینی لینک، الگوریتم جستجوی گرانشی، مدل‌سازی توزیعی.

۱- مقدمه

هر شبکه به‌عنوان مجموعه گروهی از عناصر و موجودیت‌ها تعریف شده است که دارای ویژگی‌های ارتباطی و رفتاری خاص خود است [۱]. تعداد عناصر، نوع و شیوه ارتباطات عناصر، تنوع و پیچیدگی‌های ارتباطی آن شبکه‌ها، باعث شده که آن‌ها را از شبکه‌های بسیار ساده تا بسیار پیچیده تقسیم‌بندی کنیم. هر شبکه در برابر تغییرات درونی اعضا و ارتباطات خود دارای ثبات است و یا در طول زمان با پویایی مداومی همراه خواهد بود [۲۶]. این تغییرات ممکن است به استحکام شبکه کمک کند و یا به تضعیف آن بیانجامد. استحکام یا فروپاشی یک شبکه از مسایل مهم مورد بررسی بعضی فراهم کننده شبکه‌ها است و در مقابل مهاجمان به شبکه درصدد اختلال یا فروپاشی آن هستند. این امر نشان‌دهنده اهمیت مطالعه و تحلیل ساختاری و پویایی شبکه‌ها است و می‌تواند نسبت به نوع شبکه، نیاز به بررسی و تحلیل‌های متفاوتی داشته باشد [۶۱].

شبکه‌های کامپیوتری و شبکه‌های اجتماعی دو موضوع مطرح مهم و اساسی جامعه موجود برای مطالعه و تحقیق می‌باشند که دارای ارتباط پیچیده و تنگاتنگ هستند [۱]. این دو شبکه ازنظر تعداد اعضا، نوع ارتباط، تنوع مبادله اطلاعات بین اعضا و درجه پیچیدگی آن‌ها، دارای مشکلات و معضلات خاص خود هستند. شبکه‌های کامپیوتری در حال تغییر و تحول عناصر، ارتباطات و الگوریتم‌های خود برای پشتیبانی فزاینده از شبکه‌های اجتماعی و پوشش و ارایه خدمات توسعه‌پذیر آن‌ها می‌باشند [۳۲]. گسترش روزافزون شبکه‌های اجتماعی باقابلیت ارایه خدمات گسترده دنیای امروزی، نیازمند وجود و گسترش شبکه‌های کامپیوتری است که بتواند اکوسیستم آن‌ها را در بستر اینترنت فراهم کند [۱]. شبکه‌های اجتماعی از موضوعات محبوب برای بررسی، مطالعه و تحقیق است [۶۲].

یک شبکه اجتماعی به‌عنوان گروهی از افراد[۱] یا موجودیت‌هایی[۲] که با یکدیگر در حال همکاری[۳] یا رقابت[۴] هستند، تعریف می‌شود [۱]. به کلیه همکاران یا رقیبان در اصطلاح بازیگران[۵] یک شبکه اجتماعی گفته می‌شود. این بازیگران می‌توانند موجودیت‌هایی به‌صورت اشخاص، سازمان‌ها، گروه‌ها، موجودات زنده، بیماری‌ها، صفحات اینترنتی و یا ترکیبی از آن‌ها باشند [۲۸]. در اصل هر موجودیتی که قابلیت اتصال به سایر موجودیت‌های دیگر را دارد می‌تواند به‌عنوان یک شبکه اجتماعی در نظر گرفته شود [۳۲]. نمونه‌های شبکه‌های اجتماعی عبارت‌اند از فیس‌بوک، توییتر، لینکداین و مانند آن این‌ها که براساس ابعاد، کاربرد و میزان بازیگران آن متفاوت هستند.

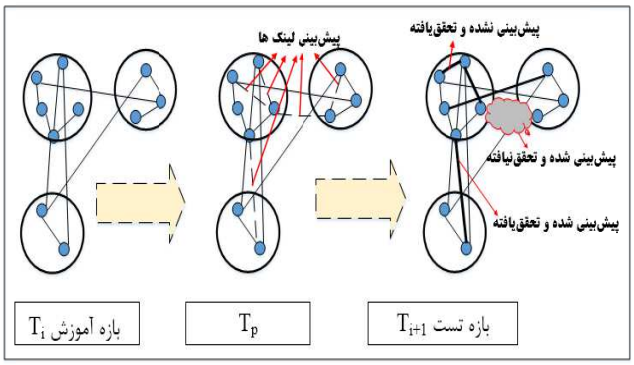
برای بررسی و مطالعه یک شبکه اجتماعی نیاز به مدل‌سازی ساختاری و رفتاری و ابزار مناسب آن شبکه است. در مدل‌سازی نوین عامل‌گرای شبکه، هر موجودیت شبکه را می‌توان به‌عنوان یک عامل در نظر گرفت [۴۶]. معماری مدل‌سازی چندعاملی می‌تواند رفتار و ساختار گراف‌گونه شبکه را به نحو مناسب مدل‌سازی و مدیریت کند. هر موجودیت در این شبکه را می‌توان به‌عنوان یک عامل در مدل‌سازی نوین شبکه در نظر گرفت [۴۸]. ساختار مدل‌سازی چندعاملی می‌تواند رفتار و ساختار گراف شبکه را به نحو مناسب مدل‌سازی و مدیریت کند.

اکنون مدل‌سازی چندعاملی به‌عنوان مدل برای بررسی و تحلیل ساختارهای مختلف موجودیت‌های هوشمند و غیرهوشمند طرح‌شده که برای مطالعه و مدل‌سازی ساختار و رفتار شبکه‌های اجتماعی بسیار مناسب است [۴۸]. این مدل نیازمند داشتن ابزارهای مدل‌سازی و پیاده‌سازی ماشینی برای مطالعه و مدل‌سازی ساختار و رفتار شبکه‌های اجتماعی است. ارایه ابزارهای جدید یک نیاز اساسی است و داشتن بستر مناسب و معماری متناسب با ساختارهای اجتماعی، آن‌ها را برای مدل‌سازی، کنترل و تحلیل، قدرتمندتر می‌سازد. ابزارهای قوی نظارت، مدیریت و بررسی ارایه شده است که شبکه‌های اجتماعی را در ساختارهای مجازی فراگیر کرده است [۶۳].

برای مدل‌سازی چندعاملی، تعاملات شبکه ممکن است با یک گراف ساده یا یک گراف چندگانه بیان شوند. هر یک از همکاران یا رقیبان با یک گره و همکاران یا رقابت‌های میان آن‌ها را با یال نشان می‌دهند. شبکه‌های اجتماعی می‌توانند توسط یک گراف پیچیده مرکب از گره‌ها و یال‌ها نمایش داده شوند که در آن افراد یا سازمان‌ها، همان گره‌ها و روابط میان آن‌ها همان یال‌های گراف را تشکیل می‌دهند. همچنین در شبکه اینترنت نیز صفحات وب یا سایت‌های وب به‌صورت گره‌ها و ارتباط میان آن‌ها، یال‌های گراف فرض می‌شوند. مدل‌سازی چندعاملی، معماری مناسب مطالعه و بررسی یک شبکه اجتماعی در ابعاد کوچک تا بسیار بزرگ است. گراف حاصل از شبکه دارای شکل یا ساختار خاص است که با استفاده از این ساختار می‌توان معیارهای مختلفی ازجمله مرکزیت، بینابینی و مانند آن برای شناخت شبکه به دست آورد. سپس با این دانش می‌توان مثلاً گره‌های بااهمیت و تأثیرگذار را شناسایی کرد. همچنین با توجه به روابط میان گره‌های مختلف، انجمن‌های موجود دربرگیرنده گره‌های مختلف را تعیین نمود که این‌گونه تحلیل را تحلیل ایستای شبکه گویند، زیرا این‌گونه تحلیل‌ها در طول زمان نیستند و صرفاً در یک‌زمان خاصی در شبکه صورت می‌گیرد [۲].

تحلیل شبکه‌های اجتماعی موضوعی است که امروزه بسیاری از پژوهشگران به آن می‌پردازند [۱]. تحلیل شبکه‌های اجتماعی [۶] یک تئوری علوم اجتماعی [۷] نیست بلکه یک رهیافت [۸] برای بررسی ساختارهای اجتماعی است. به این دلیل به تحلیل شبکه‌های اجتماعی، تحلیل ساختاری [۹] گفته می‌شود. مهم‌ترین تفاوت میان تحلیل شبکه‌های اجتماعی با روش‌های سنتی تحقیقات علوم اجتماعی این است که شخصیت یک بازیگر اجتماعی یا رابطه میان بازیگران مشکل اساسی آن است. درحالی‌که تحقیقات و مطالعات علوم اجتماعی اساساً به ویژگی‌های افراد می‌پردازد. در حقیقت یک تحلیل‌گر شبکه اجتماعی به دنبال نحوه دستیابی به ایجاد یک موجودیت و شیوه اتصال آن موجودیت‌ها در یک شبکه اجتماعی است [۱] [۳].

برخی از تحلیل‌گران شبکه اجتماعی معتقدند که موفقیت و شکست یک جامعه یا سازمان اغلب وابسته به الگوهای ساختاری است که آن‌ها در گراف شبکه اجتماعی به وجود می‌آورند. با تاریخچه‌ای بالغ بر ۷۰ سال، تحلیل شبکه‌های اجتماعی یک موضوع میان‌رشته‌ای بین رشته‌هایی مختلف جامعه‌شناسی [۱۰]، ریاضیات [۱۱] و علوم کامپیوتر [۱۲] مطرح است که در علوم مختلف مثل جامعه‌شناسی، اقتصاد [۱۳]، علوم ارتباطی [۱۴]، روانشناسی [۱۵]، فیزیک و کامپیوتر کاربرد دارد [۱]. شبکه‌های اجتماعی به‌طور عمده از دو دیدگاه ساختار و پویای قابل‌بررسی هستند. بررسی‌ها نشان می‌دهد که این شبکه‌ها در خصوصیات ساختاری به طرز جالب توجّهی اشتراک دارند [۲].



شکل ۱- فرآیند پیش‌بینی لینک

با این نگاه تغییرات این شبکه‌ها در طول زمان تا حدی قابل‌بررسی و پیش‌بینی است. به همین دلیل شبکه‌های اجتماعی را ازنظر ساختار و رفتار در طول زمان تحلیل و ارزیابی می‌کنند [۶۴]. در دهه‌ی اخیر، با بررسی ساختار شبکه‌ها و با استفاده از داده‌های واقعی و ابزارهای پردازشی قوی، بسیاری از خصوصیات مشابه شبکه‌ها استخراج شده است. مثلاً کوچک بودن میانگین کوتاه‌ترین فاصله موجودیت‌ها، بالا بودن ضریب خوشه‌بندی عناصر، و تشخیص وجود انجمن‌ها بین اعضای شبکه، بعضی از معیارهای طرح‌شده برای مطالعه خصوصیات مشترک تمامی شبکه‌ها در علوم مختلف هستند [۲۶]. تحلیل این شبکه‌ها از دو دیدگاه ساختاری (ایستا) و رفتاری (پویا) قابل‌مطالعه است [۶۵]. در تحلیل ایستا با برش تصویر در یک‌لحظه زمانی از شبکه، بررسی‌های مختلفی بر اساس ویژگی‌های ساختاری آن انجام می‌گیرد. ازآنجاکه پویایی یکی از ویژگی‌های مهم شبکه‌های اجتماعی است، تحلیل پویا نگاه دیگری است که در مطالعه شبکه‌ها به آن توجه می‌شود. نحوه انتشار یک پدیده در شبکه یکی از مهم‌ترین مصادیق پویایی آن‌ها است [۶۵]. اعضای هر شبکه در هر حوزه، رفتاری منحصربه‌فرد دارند که در کنار دیگر اعضا به ایجاد یک پدیده در شبکه منجر می‌شوند. انتشار ویروس‌ها در جوامع انسانی، خرابی پی‌درپی [۱۶] خطوط انتقال برق، انتشار شایعه و گسترش ویروس‌های رایانه‌ای در شبکه‌های مختلف اجتماعی و فناوری ازجمله مثال‌های انتشار یک پدیده در این شبکه‌ها هستند که هر یک از آن‌ها از رفتار جمعی گره‌های موجود در شبکه ناشی می‌شوند [۲].

موضوع پژوهش ما در این مقاله پیش‌بینی لینک در شبکه‌های اجتماعی با استفاده از ساختار ایستای شبکه است. پیش‌بینی برای بررسی وجود ارتباط میان دو موجودیت بر اساس ویژگی‌های موجودیت‌ها و دیگر ارتباط‌های موجود در گراف شبکه را "پیش‌بینی لینک" [۱۷] می‌گویند [۳]. به‌عبارت‌دیگر اگر در زمان t_i یک تصویر لحظه‌ای از مجموعه لینک‌ها داشته باشیم، هدف پیش‌بینی لینک‌ها در زمان t_{i+1} است [۴]. مساله پیش‌بینی در یک نگاه در شکل ۱ ارایه‌شده است. در این شکل، شبکه در زمان i شبکه‌ای است که موردمطالعه قرار می‌دهیم که آن را با T_i نمایش داده‌ایم. شبکه در زمان بعدی $(i+1)$ را برای تحقق پیش‌بینی‌ها خواهیم داشت که با T_{i+1} نمایش می‌دهیم. بر اساس ساختار شبکه در زمان T_i پیش‌بینی لینک برای زمان T_{i+1} را انجام می‌دهیم. زمان فرآیند پیش‌بینی را با T_p نمایش می‌دهیم.

منظور از بازه زمانی، فاصله بین دو برش زمانی است که پیش‌بینی انجام می‌شود و بسته به هر سیستم و هر شبکه متفاوت است. این بازه می‌تواند بر اساس ثانیه، دقیقه، ساعت، روز، هفته، ماه، سال و حتی چندین سال باشد. مثلاً در ساختار گرید برق بازه زمانی می‌تواند ساعتی باشد که می‌تواند با گزارش‌گیری و

ذخیره داده‌های موردنیاز برای تحلیل شبکه مرتبط باشد. یعنی حجم داده‌های موردنیاز بتواند ضمن داشتن مفهوم قابلیت بررسی، از جهت ذخیره و پشتیبانی داده‌های فوق، اقتصادی و امکان‌پذیر باشد.

معیارهای رایج موجود برای ارزیابی نتایج الگوریتم‌ها عبارت از صحت[۱۸]، دقت[۱۹]، فراخوانی[۲۰] و معیار f است. می‌توان به‌نوعی از این معیارها برای ارزیابی نتایج الگوریتم‌های پیش‌بینی لینک استفاده کرد. در پیش‌بینی لینک مساله موردتوجه ما دقت در تعیین لینک‌های پیش‌بینی‌شده و تحقق آن‌ها است. البته بازه زمانی پیش‌بینی لینک $[i , i + (n + 1)]$ است که می‌تواند مساله پیش‌بینی را برای دو یا چند بازه غیر متوالی نیز تعیین و آن را پیچیده کند [۲] [۵] [۶]. اگر داده‌ها شبکه بخواهد برای تحلیل در بازه زمانی $[i , i + (n + 1)]$ استفاده شود، مساله پیش‌بینی به‌عنوان یک چالش برای بررسی داده‌های بزرگ[۲۱] و شبکه‌های پیچیده[۲۲] تبدیل خواهد شد. ما در این مقاله سعی بر ارایه مدل مناسبی برای کاهش زمان پیش‌بینی لینک برای داده‌های بزرگ را داریم. معماری چندعاملی با ساختارهای موازی مناسب کاهش زمان جستجو فضای مساله است. همچنین مدل گرانشی نیز مناسب کاهش زمان جستجو برای مساله‌های بزرگ در کاهش زمان پیش‌بینی بسیار مؤثر است.

ما در این مقاله عملکرد الگوریتم‌های محلی پیش‌بینی لینک موجود را مطالعه و یک معماری ارایه می‌دهیم که از اطلاعات انجمنی گره‌های درونی برای پیش‌بینی‌های لازم استفاده کند. بیشتر اتکا و مانور معماری ما در مدل پیشنهادی مبتنی بر استفاده از اطلاعات مفید انجمن‌ها برای پیش‌بینی است. مقصود آن است که انجمن‌ها را به‌عنوان عامل‌های جستجوی الگوریتم گرانشی[۲۳]درنظر می‌گیریم. ضمناً سعی بر مدل‌سازی پیش‌بینی مبتنی بر الگوریتم مکاشفه‌ای جستجوی گرانشی برای بهبود عملکرد پیش‌بینی داریم. این کار جهت بهینه‌سازی عملکرد انجمن‌ها در بهبود دقت و سرعت پیش‌بینی لینک صورت گرفته است که بسط آن را برای مطالعه داده‌های بزرگ امکان‌پذیر می‌کند. نوآوری‌های قابل ارایه ما در این مقاله به‌طور خلاصه عبارت‌اند از:

- مدل‌سازی عامل‌گرای پیش‌بینی لینک مبتنی بر جستجوی گرانشی
- سرعت و دقت مناسب مدل پیش‌بینی لینک برای شبکه‌های حجیم
- قابلیت بسط، توسعه و توزیع مدل برای تحلیل داده‌های بسیار بزرگ

در ادامه مقاله و در بخش ۲ تحقیقات و "کارهای مرتبط" انجام‌شده با پیش‌بینی لینک را شرح می‌دهیم. در بخش ۳ ضمن تعریف مساله "پیش‌بینی لینک" و در بخش ۴ "الگوریتم‌های پیش‌بینی" را ارایه می‌کنیم. در ادامه و در بخش ۵ "الگوریتم جستجوی گرانشی" را شرح می‌دهیم. در بخش ۶ "معماری پیشنهادی" را بیان و "مدل‌سازی عامل‌گرای جستجوی گرانشی پیش‌بینی لینک" در شبکه‌های اجتماعی را ارایه می‌دهیم. در بخش ۷ به "نتایج آزمایش‌های تجربی" و تحلیل آن اشاره داریم. نهایتاً و در بخش ۸ به "نتیجه‌گیری" می‌پردازیم.

۲- کارهای مرتبط

بیشتر کارهای انجام‌شده سال‌های اخیر در مساله پیش‌بینی لینک از اطلاعات ساختاری گراف استفاده می‌کنند که در زیر به موضوعات مهم ارایه شده آن‌ها می‌پردازیم و همچنین در ادامه به تحقیقات مهم پیرامون الگوریتم جستجوی گرانشی و کاربردهای آن اشاره خواهیم کرد. می‌توان روش‌های موجود برای پیش بینی لینک بین گره‌های یک شبکه را به سه دسته عمده روش‌های مبتنی بر معیارهای شباهت[۲۴]، روش‌های مبتنی بر آمار و روش‌های مبتنی بر بیشترین احتمال تقسیم بندی کرد. در این روش‌ها علاوه بر بررسی ساختار گراف، ویژگی‌ها و خصوصیاتی که باعث می‌شود احتمال وجود لینک بیشتر شود را استخراج می‌کنند [۲۱].

الگوریتم‌های مختلف پیش‌بینی لینک معمولاً از شباهت بین دو گره استفاده می‌کنند [۴]. این الگوریتم‌ها فرض می‌کنند که هرچه قدر شباهت بین دو گره بیشتر باشد احتمال اینکه بین آن‌ها در آینده لینکی برقرار شود بیشتر است. موضوعی که در این مقاله به دنبال آن هستیم این است که چگونه از این معیار شباهت به نحو مناسبی برای هدف خود استفاده کنیم. روش‌های قدیمی بیان‌شده هرکدام از یک ویژگی خاص ساختار شبکه استفاده می‌کنند. درحالی‌که می‌توان از اطلاعات بیشتری نیز برای این منظور بهره برد. ازجمله می‌توان به اطلاعات گروه و انجمنی که هر گره در آن عضو است اشاره کرد. روش‌های این نوع پیش‌بینی لینک می‌تواند خیلی دقیق‌تر و بهتر از روش قبلی باشد. بااین‌وجود نسبت به روش قبلی کندتر هستند اما در شبکه‌هایی حجیم دارای میلیون‌ها عضو می‌توانند بهتر عمل کنند [۴۹]. بنابراین روشی که ما در این دنبال می‌کنیم روش‌های مبتنی بر معیار شباهت است.

ناول و همکاران [۴] اولین مدل پیش‌بینی لینک را که صراحتاً در شبکه‌های اجتماعی کاربرد داشت را ارایه کردند. روش پیش‌بینی بر اساس تشابه بین دو گره است که احتمال دارد در آینده با هم دوست شوند. آن‌ها گره‌ها را بر اساس میزان امتیاز شباهت بینشان رتبه‌بندی کردند. بعد از آنها الحسن و همکاران [۵۰] این رویکرد را با دو روش دیگر گسترش دادند. ابتدا نشان دادند که استفاده از داده‌های خارجی می‌تواند عملکرد پیش‌بینی لینک را بهبود دهد. سپس آن‌ها از معیارهای شباهت مختلفی برای حل مساله پیش‌بینی لینک استفاده کردند یعنی حل مساله پیش‌بینی لینک را به‌عنوان یک مساله دسته‌بندی باینری طرح کردند [۵۰]. بعداً مساله پیش‌بینی لینک در حوزه داده‌های رابطه‌ای و حوزه اینترنت مورد استفاده قرار گرفت [۶۶]. در مقاله شرکت و همکاران [۵۱] از الگوریتم بهینه‌سازی مورچگان برای پیش‌بینی لینک استفاده شده است.

الگوریتم جستجوی گرانشی یک الگوریتم بهینه‌سازی مکاشفه‌ای است که به‌تازگی در بین جوامع علمی رواج یافته است. این الگوریتم از قوانین طبیعت الهام گرفته است و در سال ۲۰۰۹ توسط راشدی و همکاران [۴۴] برای حل مسائل بهینه‌سازی به کار گرفته شد. این الگوریتم مبتنی بر جمعیت و از قوانین طبیعت الهام گرفته است. لذا از مجموعه‌ای از عوامل جستجو کننده تشکیل‌شده که با یکدیگر از طریق نیروی گرانش در ارتباط و کنش هستند [۴۵].

مدل عامل گرای جستجوی گرانشی دارای ساختار طبیعی و مناسب مدل‌های محیطی است و افراد مختلفی از این مدل استفاده کرده‌اند [۴۵]. یزدانی و همکاران [۵۲] از این مدل برای یافتن چندین جواب برای مساله‌هایی که دارای چندین مدل هستند استفاده کرده است. پی و همکاران [۵۳] از آن برای حل مساله زمان‌بندی در محیط‌های دو ولتاژی استفاده کرده‌اند. دراغی نژاد و همکاران [۵۴] از مدل مبتنی بر جستجوی گرانشی برای افزایش بازدهی شبکه و کاهش تداخل آن بهره برده‌اند که از یک اپراتور محلی با ترکیب با الگوریتم جستجوی گرانش استفاده کرده‌اند.

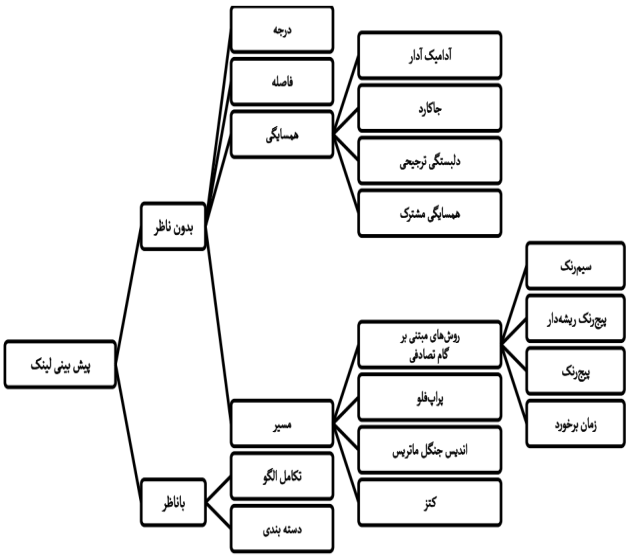
نیکنام و همکاران [۵۵] در مدلی سعی بر یافتن واکنش بهینه به کنترل ولتاژ و توان در سیستم‌های برق داشته‌اند. آن‌ها برای این کار با تنظیم پارامتر از طریق تولید تصادفی و خودتنظیم پارامترهای مختلف با ترکیب الگوریتم جستجوی گرانشی، سعی داشته‌اند جوابی بهینه بیابند. وانگ و لی [۵۶] مطالعه‌ای روی شرایط مرزی داشته‌اند که بهینه‌سازی نامحدود را نشان می‌دهد. حسن‌زاده و همکاران [۵۷] روش جدیدی برای کنترل و توزیع مساله بهینه‌سازی چند هدفی داشته‌اند که یک عملگر یکنواخت جهش[۲۵] و سیاست نخبه‌گرا را با الگوریتم جستجوی گرانش چندهدفه ترکیب کردند. راشدی و همکارانش [۵۸] روشی برای بهینه‌سازی فضای جستجوی باینری ارایه دادند که برای مدل‌سازی آن، سرعت را معادل تغییرات احتمالی بیت‌ها در زمانی که از حالت صفر به حالت یک تغییر می‌یابند قراردادند. شاو و همکاران [۵۹] بحث هزینه و مساله پخش بار در سیستم‌های برق را باهم ترکیب کردند. براساس دانش ما تاکنون مساله مدل‌سازی

مساله پیش‌بینی لینک مبتنی بر ساختار جستجوی گرانشی در کنار معماری چندعاملی برای حل مشکلات شبکه‌های با داده‌های بزرگ یا عناصر بسیار زیاد در قالب شبکه‌های اجتماعی ارایه نشده است.

۳- پیش‌بینی لینک

بیان شد که پیش‌بینی امری لازم برای مطالعه رفتار و ساختار بعدی یک شبکه خواهد بود. یعنی براساس توانایی ما در پیش‌بینی می‌توانیم وضعیت یک شبکه را درآینده برای تصمیم‌گیری مدنظر قرار دهیم. مثلاً پیش‌بینی وضعیت مصرف‌کنندگان برق در آینده می‌تواند بر وضعیت تولیدکنندگان برق و وضعیت تغذیه مصرف‌کنندگان بسیار مؤثر باشد. در یک پیش‌بینی درست میزان صحت پیش‌بینی، میزان دقت پیش‌بینی، و سرعت زمان از عوامل مؤثر بر مساله تصمیم‌سازی است [۲۱]. مساله پویایی در ساختار و رفتار شبکه نیازمندی به پیش‌بینی برای تصمیم‌گیری را نمایان می‌سازد. پیش‌بینی در یک شبکه می‌تواند برای بررسی ساختار آتی شبکه (شیوه ارتباط اعضا) یا رفتار آن (شیوه تعامل یا رفتار اعضا) صورت گیرد. پیش‌بینی ساختاری برای کاربردهای که نیاز به فرآیند یادگیری ندارد مؤثر است. پیش‌بینی رفتاری برای کاربردهای مفید است که حتماً نیازمند دانستن پیشینه ارتباطات مختلف هستیم.

ما در این مقاله به دنبال پیش‌بینی لینک برای کاربردهایی هستیم که نیاز به دانستن پیشینه وقوع ارتباطات ندارد تا بتوانیم در تصمیم‌سازی سرعت عمل مناسب داشته باشیم. پیچیدگی پیش‌بینی‌های رفتاری و نیاز به سابقه‌گیری، مشکل داده‌های بزرگ را به وجود می‌آورد که امر پیش‌بینی را برای تصمیم‌سازی نامناسب می‌سازد. لذا ما از مزیت پیش‌بینی‌های ساختار موجود شبکه بهره می‌بریم تا بتوانیم با رفع مشکل داده‌های بزرگ و تحلیل‌های طولانی مدت، در زمان کوتاهی به‌دقت مناسبی دست یابیم. پس ما به دلیل مشکلات زمانی و دقت کار، به سابقه اتصالات توجه نداریم و وجود ناظر را نیز مرتفع کرده‌ایم. بیشتر تمرکز خود را بر ویژگی‌های ساختاری موجود شبکه معطوف داشته‌ایم.



شکل ۲- روش‌های پیش‌بینی لینک

مساله پیش‌بینی با حجم داده‌های بزرگ بسیار پرهزینه و کند است لذا با ارایه معیارها و مدل‌های جدید بررسی آن را امری لازم و حیاتی می‌دانیم. در بسیاری از شبکه‌های بیولوژیکی، ازجمله شبکه‌های مواد غذایی، تعامل پروتئین- پروتئین و شبکه‌های متابولیک، پیش‌بینی این‌که آیا بین دو گره در شبکه یالی وجود دارد با انجام آزمایش‌ها مختلف امکان‌پذیر است که این کار معمولاً بسیار پرهزینه است. علاوه بر این دانش ما در این شبکه‌ها معمولاً محدود است. به‌عنوان مثال ۸۰ درصد از فعل و انفعالات مولکولی در سلول‌های مخمر [۱۰] و ۹۹ درصد از انسان [۱۱] [۱۲] هنوز شناخته‌شده نیست. به‌جای بررسی کورکورانه تمامی فعل و انفعالات امکان‌پذیر، می‌توان بر اساس فعل و انفعالات شناخته‌شده ارتباطات جدید را پیش‌بینی کرد [۱۳] [۱۴].

دوستی در شبکه‌های اجتماعی همان مفهوم عام رابطه بین دو نفر را دارد و تشخیص روابط مخفی یک مکانیزم تشخیص دوستی بالقوه برای پیشنهاد دوستی است [۱]. در این شبکه‌ها با کشف روابط مخفی، می‌توانیم از آن‌ها برای پیش‌بینی وقایع آتی یا پیشنهاد روابط لازم اعضاء بهره‌برداری کنیم. در شبکه های اجتماعی دوستی را این گونه تعریف می‌کنیم: "اگر دو گره با هم در ارتباط باشند می‌گوییم آن دو گره با هم دوست هستند". در چنین مواردی ارتباطات مخفی شاید در برگیرنده گره‌های اجتماعی موجودی باشد که هنوز آن اجتماع، اراتباطات ذکر شده را ایجاد نکرده است [۴]. معمولا ایجاد یا از دست رفتن ارتباطات در شبکه‌های اجتماعی تحت عنوان مساله پیش‌بینی لینک شناخته می‌شود [۳] [۴] [۱۶].

پیش‌بینی لینک در خارج از حوزه شبکه‌های اجتماعی نیز دارای کاربردهای بسیاری از قبیل تجارت الکترونیک، فعل و انفعالات ژنتیکی، کنترل مصرف و تولید برق، و ارتباط پروتئین با پروتئین است. از آنجا که این مساله در دامنه‌های مختلفی کاربرد دارد برای حل آن چندین روش پیشنهاد شده است. بسیاری از این روش‌ها از ویژگی‌های ساختاری گراف شبکه و در بعضی موارد هم از روش‌های غیرساختاری استفاده می‌کنند. در ساختار گرافی شبکه اجتماعی، پیش‌بینی وجود ارتباط میان دوموجودیت براساس خصوصیت‌های موجودیت‌ها و سایر لینک‌های مشاهده شده در گراف شبکه را پیش‌بینی لینک[۲۶] می‌گویند [۷]. به‌عبارت دیگر اگر در زمان t یک تصویر لحظه‌ای ازمجموعه لینک‌ها داشته باشیم، هدف پیش‌بینی لینک‌ها در زمان $t+1$ است [۸]. البته این مفهوم می‌تواند برای پیش‌بینی از زمان t_i تا زمان $t_{i+(n+1)}$ یعنی بازه‌های غیرمتوالی، مساله را بسیار پیچیده کند.

تعریف ریاضی پیش‌بینی لینک در مقاله ناول و همکاران [۴] بیان شده که پیش‌بینی لینک در یک گراف را به‌عنوان تعامل بین راس‌های گراف تعریف کرده‌اند. یعنی اگر شبکه اجتماعی به‌صورت $G=(V,E)$ مفروض باشد؛ به طوری که هر یال $e=(u,v)\in E$ نشان‌دهنده یک ارتباط[۲۷] میان u و v در زمان $t(e)$ باشد؛ ارتباط میان u و v در زمان‌های مختلف به‌صورت یال‌های موازی در درون گراف نشان‌داده می‌شوند. آنگاه $G[t,t'],t<t'$ زیرگرافی از G است که شامل تمامی یال‌های ایجاد شده در بازه زمانی t و t' است [۸].

اکنون تعریف پیش‌بینی لینک با انتخاب چهار بازه زمانی $t_0<t_0'<t_1<t_1'$ است. الگوریتم پیش‌بینی لینک تنها با دسترسی به گراف $G[t_0,t_0']$ باید به‌عنوان خروجی، یال‌هایی را پیش‌بینی کند که در گراف $G[t_0,t_0']$ وجود دارند ولی در گراف $G[t_1,t_1']$ وجود نداشته‌اند. بازه $[t_0,t_0']$ را بازه آموزش[۲۸] و بازه $[t_1,t_1']$ را بازه آزمون[۲۹] نام نهاده‌اند. همان‌گونه که در شکل ۲ نشان داده‌ایم، بازه T_i، بازه آموزش و بازه T_{i+1}، بازه آزمون است.

ما برای مطالعه و ارزیابی معماری و الگوریتم پیشنهادی خود، کلیه لینک‌های مورد پیش‌بینی در الگوریتم پیش‌بینی لینک را از نظر عملکرد مورد ارزیابی قرار دادیم. این معیارها برای ارزیابی عبارتند از:

- لینک‌های پیش‌بینی شده تحقق‌یافته (TP)
- لینک‌های پیش‌بینی شده تحقق‌نیافته (FP)
- لینک‌های پیش‌بینی نشده تحقق‌نیافته (TN)
- لینک‌های پیش‌بینی نشده تحقق‌یافته (FN)

روش‌های مختلف پیش‌بینی لینک به دلیل وجود ناظر مرکزی یا فقدان آن (به شرح شکل ۲) به دودسته کلی با ناظر[۳۰] و بدون ناظر[۳۱] تقسیم‌بندی می‌شوند [۹]. روش‌های بدون ناظر، روش‌هایی هستند که بدون بررسی پیش‌زمینه روند به وجود آمدن لینک ها درگذشته، تنها به کمک ویژگی های ساختاری موجود در گراف شبکه به پیش‌بینی لینک‌ها در آینده می‌پردازند. روش‌های با ناظر پس از یادگیری[۳۲] یک یا چند مرحله از فرآیند به وجود آمدن لینک‌ها درگذشته، به پیش‌بینی لینک‌های آینده می‌پردازند [۹].

الگوریتم‌های پیش‌بینی لینک بدون ناظر غالباً از ویژگی‌های ساختاری مانند تعداد همسایه‌های مشترک، طول کوتاه‌ترین مسیر میان دو گره، درجه[۳۳]دو گره استفاده می‌کنند [۹]. الگوریتم‌های پیش‌بینی لینک با ناظر گاهی اوقات با یادگیری پارامتر های یک مدل احتمالاتی و یا بررسی روند تکامل یک زیرساختار خاص در گراف شبکه به پیش‌بینی لینک می‌پردازند. درمجموع از اطلاعات هر روشی که به نوعی شباهت دو گره را نسبت به یکدیگر در گراف شبکه نشان دهد را می‌توان به‌نوعی برای پیش‌بینی لینک نیز استفاده نمود.

بسیاری از تحقیقات در تجزیه و تحلیل شبکه‌های اجتماعی به بهره‌برداری از اطلاعات کاربران و روابط اجتماعی مانند دوستی متمرکز شده‌اند. با این‌حال در بعضی از پژوهش‌ها، برای بهبود عملکرد پیش‌بینی لینک از اطلاعات دیگر ازجمله اطلاعات خوشه‌ها و انجمن‌ها استفاده می‌شود [۱۷]. در آزمایش‌های مختلف، فنگ و همکاران [۱۸] نشان دادند وقتی‌که ساختار خوشه‌ای شبکه‌ها رشد می‌کند، دقت و صحت اقدامات پیش‌بینی لینک براساس اطلاعات ساختاری بهتر می‌شود [۱۹] [۲۰]. این نتایج باعث شد که به‌نوعی از اطلاعات انجمن‌ها برای بهبود عملکرد پیش‌بینی لینک استفاده شود.

ما از اطلاعات انجمن‌ها در کنار الگوریتم جستجوی گرانشی الهام گرفته از قوانین طبیعی فیزیکی برای بهبود عملکرد پیش‌بینی لینک استفاده کرده‌ایم. این الگوریتم از مجموعه‌ای از عوامل جستجوگر تشکیل‌شده است که با یکدیگر از طریق نیروی گرانش در ارتباط هستند [۴۴].

۴- الگوریتم‌های پیش‌بینی

همانطورکه در پیشتر گفته شد می‌توان روش‌های موجود برای پیش بینی لینک بین گره‌های یک شبکه را به سه دسته عمده تقسیم‌بندی کرد. دسته اول روش‌های مبتنی بر معیارهای شباهت است. این روش‌ها ساده ترین روش پیش‌بینی لینک هستند که در آن به هر جفت از گره‌های x و y یک امتیاز داده می‌شود که این امتیاز بر حسب میزان شباهت دو گره به هم است. تمام لینک‌هایی که مشاهده نشده اند و قرار است که پیش بینی شوند بر حسب میزان شباهت دو گره به هم رتبه بندی می‌شوند و لینک هایی که امتیازشان بیشتر است پیش بینی می‌شوند. این روش بیشتر بر اساس ساختار شبکه، لینک‌ها را پیش بینی می‌کند. مثل روش‌های مبتنی بر فاصله که در آن گره‌هایی که در کمترین فاصله از هم قرار دارند شانس بیشتری برای پیش بینی لینک بین آنها در آینده دارند. دسته دوم روش‌های مبتنی بر آمار هستند.. در این روش‌ها از مدل‌های آماری و توزیع‌های آماری استفاده می‌شود. دسته سوم روش‌های مبتنی بر بیشترین احتمال هستند. در این روش‌ها علاوه بر بررسی ساختار گراف، ویژگی‌ها و خصوصیاتی که باعث می‌شود احتمال وجود لینک بیشتر شود را استخراج می‌کنند [۲۱]. روش‌هایی که ما سعی در بیان آن‌ها داریم مبتنی بر معیار شباهت هستند.

روش‌های بدون ناظر پیش‌بینی لینک عمدتاً تنها از یک تصویر لحظه‌ای[۳۴] گراف شبکه استفاده کرده و بدون در نظر گرفتن محتوای شبکه موردبررسی، تنها به کمک ویژگی‌های ساختاری گراف شبکه سعی در پیش‌بینی لینک‌ها در تصویر لحظه بعدی از شبکه می کند. به این روش‌ها، روش‌های مبتنی بر معیارهای شباهت نیز گفته می‌شود. این دسته از روش‌ها عمدتاً از الگوریتم‌های ساده‌ای استفاده نموده که در بدترین حالت پیچیدگی زمانی از مرتبه $O(n^3)$ دارند. این روش ها معمولاً از سرعت مناسبی برخوردار بوده ولی دقت خروجی این دسته از روش‌ها نسبت به روش‌های پیش‌بینی لینک با ناظر کمتر است. در این فصل به بررسی تعدادی از الگوریتم‌ها و روش‌های پیش‌بینی لینک بدون ناظر در شبکه‌های اجتماعی پرداخته می‌شود. در هر زیر بخش یک روش یا یک دسته روش موردبررسی قرارگرفته است.

۴-۱- روش‌های مبتنی بر فاصله

یکی از ساده‌ترین روش‌ها برای پیش‌بینی لینک استفاده از معیار فاصله[۳۵] است [۲۲] در این روش زوج گره‌ای که بین آن‌ها لینکی وجود ندارد و در فاصله کمی از یکدیگر قرار دارند، بیشترین احتمال برای به وجود آمدن ارتباط میان آن دو نسبت به سایرین وجود دارد. ازآنجایی‌که الگوریتم‌های پیش‌بینی لینک پیش‌بینی‌های خود را برحسب امتیاز لینک‌ها به‌صورت نزولی ارایه می‌دهند، در این روش امتیاز لینک‌ها، منفی کوتاه‌ترین فاصله میان دو گره در نظر گرفته می‌شود. در عمل خروجی این الگوریتم زوج گره هایی است که در فاصله دو از یکدیگر قرار دارند.

۴-۲- روش‌های مبتنی بر درجه

در این روش تنها از درجه گره‌ها برای رتبه‌بندی لینک‌ها استفاده می‌شود. در این معیار احتمال اینکه یک یال جدید دارای نقطه پایانی x (یا y) باشد، متناسب با درجه‌ی گره‌ی x است. در [۳۲] [۳۳] احتمال آنکه دو نویسنده با یکدیگر همکاری کنند، متناسب با ضرب درجه‌ی آن دو نویسنده در گراف است. به این روش دل‌بستگی ترجیحی[۳۶] گفته می‌شود. اگر $|\Gamma(x)|$ درجه‌ی گره‌ی x باشد، رابطه (۱) نحوه محاسبه امتیاز لینک‌ها بر اساس معیار دل‌بستگی ترجیحی را نشان می‌دهد.

$$Score(x,y) = |\Gamma(x)|.|\Gamma(y)| \quad (۱)$$

۴-۳- روش‌های مبتنی بر همسایگی

در این بخش روش‌هایی که برحسب همسایه‌های یک گره میزان شباهت آن را با سایر گره‌ها را محاسبه می‌نماید، مطرح می‌شود. اساس تمامی این روش‌ها بر پایه ویژگی تراگذری[۳۷] در شبکه‌های اجتماعی است. طبق ویژگی تراگذری اگر گره x با گره y در ارتباط باشد و گره‌ی y نیز با گره‌ی z، آنگاه احتمال زیادی وجود دارد که گره‌ی x با گره‌ی z نیز در ارتباط باشد [۳۳].

۴-۳-۱- روش همسایگی مشترک

در این روش امتیاز یک لینک برابر با تعداد همسایه‌های مشترک[۳۸] میان نقاط انتهایی آن خواهد بود [۳۳]. رابطه (۲) نحوه محاسبه‌ی تعداد همسایه‌های مشترک میان دو گره را نشان می‌دهد.

$$Score(x,) = |\Gamma(x) \cap \Gamma(y)| \quad (۲)$$

رابطه (۲) کلی است و می‌توان در حالت‌های مختلفی آن را محاسبه نمود. می‌توان همسایه‌های مشترک را تنها همسایه‌های مرتبه اول میان دو گره در نظر گرفت و یا همسایه های مرتبه دوم یعنی همسایه های همسایه گره موردنظر و یا مراتب بالاتر نیز محاسبه کرد.

۴-۳-۲- ضریب جاکارد

یکی از معیارهای سنجش شباهت در استخراج داده‌ها، ضریب جاکارد[۳۹] است [۳۴] اگر ضریب جاکارد را به‌عنوان معیاری برای رتبه‌بندی یال‌ها استفاده کنیم، در کل به نوعی حالت نرمال شده روش همسایه های مشترک است. در این روش تعداد همسایه‌های مشترک میان دو گره بر اجتماع همسایه‌های آن دو تقسیم می‌شود. رابطه (۳) نحوه‌ی محاسبه‌ی ضریب جاکارد را نشان می‌دهد. ضریب جاکارد می‌گوید که دو گره شبیه‌تر به یکدیگر هستند اگر تعداد همسایه‌های مشترک زیاد و تعداد همسایه‌های غیرمشترک کمی داشته باشند.

$$Score(x,y)=\frac{|\Gamma(x)\cap\Gamma(y)|}{|\Gamma(x)\cup\Gamma(y)|} \qquad (۳)$$

۴-۳-۳- آدامیک آدار

آدامیک و آدار[۴۰] در زمینه وب برای تشخیص شباهت دو وب‌سایت به یکدیگر معیاری را به نام آدامیک آدار معرفی کردند [۳۶]. رابطه (۴) نحوه محاسبه آدامیک آدار را نشان می‌دهد. طبق رابطه پس از به دست آوردن ویژگی‌های مشترک میان دو صفحه در وب، به ویژگی‌هایی که نادر بوده و تنها میان دو صفحه موردنظر به وجود آمده است، وزن بالاتری داده می‌شود.

$$Score(x,y)=\sum_{\substack{z:feature\\ shared\ by\ x,y}}\frac{1}{log(frequency(z))} \qquad (۴)$$

۴-۳-۴- اختصاص منابع

روش اختصاص منابع[۴۱] براساس ایده اختصاص منابع در شبکه‌های پیچیده ارایه شده است [۴۰]. دو گره‌ی x و y که به‌طور مستقیم میان آن‌ها لینکی وجود ندارد، را در نظر بگیرید. گره‌ی x می‌تواند یک سری منبع[۴۲] را به کمک همسایه‌هایش به‌عنوان واسطه، به سمت گره‌ی y هدایت نماید. در حالت ساده هر گره‌ی واسطه تنها توانایی انتقال یک واحد از منبع را دارا است و به این صورت عمل می کند که تمامی منبع دریافتی را به سمت تمامی همسایه‌هایش رها می‌سازد. اکنون میزان شباهت گره‌ی y به گره‌ی x برابر با میزان منابعی است که این گره از گره‌ی x دریافت می‌کند. نحوه محاسبه شباهت بدین روش در رابطه (۵) نشان داده‌شده است.

$$Score(x,y)=\sum_{z\in\Gamma(x)\cap\Gamma(y)}\frac{1}{deg(z)} \qquad (۵)$$

۴-۴- روش‌های مبتنی بر مسیر

در این بخش دسته روش‌هایی که بر اساس معیار مسیر امتیاز لینک‌ها را مشخص می‌نمایند شرح داده می‌شود. این روش‌ها به‌نوعی ایده‌ی کوتاه‌ترین مسیر میان دو گره را به سمت محاسبه‌ی تمامی مسیرهای میان دو گره ارتقاء داده‌اند. روش‌های مبتنی بر گام تصادفی[۴۳] نیز جزو این دسته از روش‌ها محسوب می‌شوند. در زیر به روش‌های مبتنی بر مسیر اشاره خواهیم کرد.

۴-۴-۱- کتز

معیار کتز[۴۴] [۲۱] یک فرمول برای استفاده از تعداد و طول مسیرهای میان دو گره برای پیش‌بینی لینک است. مسیرها نیازی نیست که کوتاه‌ترین مسیر باشند. برای کاهش اثر مسیرهای طولانی در تعیین امتیاز یک لینک، از یک ضریب به نام β در این فرمول استفاده‌شده است. این کار باعث می‌شود که پیش‌بینی لینک میان گره‌ها در نواحی پرجمعیت اطراف آن‌ها نسبت به سایر نقاط بیشتر باشد. رابطه (۶) نحوه محاسبه‌ی امتیاز یک لینک بر اساس معیار کتز را نشان می‌دهد.

$$Score(x,y)=\sum_{L=1}^{\infty}\beta^{L}.\ |paths_{x,y}^{L}| \qquad (۶)$$

در رابطه‌ی (۶) مقدار $paths_{x,y}^{L}$ مجموعه‌ی تمامی مسیرهای به طول L از گره‌ی x به گره‌ی y است. همان‌طور که قبلاً نیز اشاره شد، $\beta > 0$ یک ضریب کنترلی است. اگر مقدار این ضریب خیلی کم باشد، معیار کتز شبیه همسایه‌های مشترک می‌شود. در مقاله [۲۰] مقدار β بین ۰.۰۰۰۵ تا ۰.۰۲۵ در نظر گرفته شده است.

۴-۴-۲- پراپ‌فلو

پراپ‌فلو[۴۵] [۴۳] یک روش مبتنی بر مسیر برای پیش‌بینی لینک است، به‌طوری‌که نسبت به نویزهای ساختاری که در فاصله دوری از گره‌ی مبدأ قرار دارند، حساس نیست. این روش به این صورت عمل می‌کند که ابتدا از یک گره یک واحد جریان به سمت سایر گره‌ها فرستاده می‌شود. یال‌های گراف وزن‌دار فرض می‌شوند. وزن هر یال نشان‌دهنده‌ی ظرفیت آن یال برای عبور جریان است. درصورتی‌که گراف وزن‌دار نباشد، می‌توان وزن همه‌ی یال‌های را برابر با یک در نظر گرفت. اکنون برحسب وزن یال‌ها، احتمال آنکه جریان موردنظر از یک یال عبور کند محاسبه می‌شود. هرچقدر وزن آن یال بیشتر باشد، احتمال بیشتری دارد که جریان از آن عبور کند. اکنون پس از محاسبه‌ی تمامی احتمال‌ها، گره‌ای که دارای بیشترین احتمال باشد، گره‌ای است که بیشترین شانس برای ایجاد لینک با آن وجود دارد.

۴-۴-۳- اندیس جنگل ماتریس

طبق اندیس جنگل ماتریس[۴۶] [۴۲]، شباهت میان دو گره‌ی x و y برابر با نسبت تعداد جنگل‌های ریشه‌دار پوشایی[۴۷] است که گره‌های x و y به آن‌ها با ریشه‌ی x تعلق دارند، به‌تمامی جنگل‌های ریشه‌دار پوشا در شبکه، است. برای بیان تعریف جنگل های ریشه دار پوشا نیاز است که تعاریف زیر مطرح شود. گراف بدون دور را جنگل[۴۸] می‌گویند. درخت، یک جنگل متصل است. هر جنگل از تعدادی درخت تشکیل‌شده است. درخت ریشه‌دار ، درختی است که یک گره به نام ریشه دارد. سایر گره‌ها فرزند ریشه هستند. جنگل پوشا[۴۹] شامل یک سری درخت ریشه‌دار است. زیرگراف ماتریس G گرافی است که تمامی گره‌ها و یال‌های آن متعلق به گراف G است. زیرگراف پوشا[۵۰] گراف G گرافی است که مجموعه گره‌های آن

متعلق به G است. برای محاسبه‌ی اندیس جنگل ماتریس، رابطه‌ی (۷) بیان‌شده است.

$$Score(x,y)=(I+\alpha L)^{-1}, \alpha>0 \quad (۷)$$

در رابطه (۷)، I ماتریس واحد است α یک ضریب کنترلی است L ماتریس کیرشهف[۵۱] است.

۵- الگوریتم جستجوی گرانشی

در حل مسائل با فضای جستجوی زیاد، راه حل های کلاسیک به دلیل اینکه فضای جستجو به صورت نمایی افزایش می‌یابد نمی‌توانند جواب بهینه بدهند. در طول دهه گذشته استفاده از الگوریتم هایی که از طبیعت الهام می‌گیرند روبه رشد است. این مسأله توسط بسیاری از پژوهشگران نشان داده شده است که این راه حل‌ها برای حل مسائلی که محاسبات پیچیده دارند از قبیل بهینه کردن تابع هدف، تشخیص الگو، کنترل اهداف، پردازش تصویرمدل سازی و مانند این‌ها. روش‌های مختلف اکتشافی[۵۲] بسیاری مانند الگوریتم‌های ژنتیک، شبیه‌سازی گداخت[۵۳] (SA)، بهینه‌سازی ازدحام ذرات[۵۴] توسط پژوهشگران تاکنون ارایه شده‌اند. این الگوریتم‌ها مسائل بهینه‌سازی مختلفی را حل می‌کنند. اگرچه تاکنون هیچ الگوریتم خاصی بهترین راه حل را در تمام حوزه‌ها نداشته است [۴۴].

الگوریتم‌های مکاشفه ای خاصیت تصادفی[۵۵] دارند. در بعضی از الگوریتم های با خاصیت تصادفی مثل SA، عملیات جستجو از یک نقطه شروع می‌شود و سپس به صورت ترتیبی ادامه می‌یابد، اگرچه بعضی از الگوریتم‌ها عمل جستجو را به صورت موازی عمل می‌کنند و از چندین عامل و از چندین نقطه برای این کارشروع می‌کنند. در این الگوریتم‌ها ارتباط محلی بین عامل‌ها می‌تواند منجر به یک نتیجه سراسری شود که به سیستم اجازه می‌دهد مساله را بدون کنترل مرکزی حل کند. در این حالت هرکدام از اعضا عملیات مختلفی را انجام می‌دهند از جمله جستجوی تصادفی، بازخورد و تعاملات مختلف. دوجنبه مهم در الگوریتم‌های اکتشافی مبتنی بر جمعیت اکتشاف[۵۶] و بهره‌برداری[۵۷] است که باید الگوریتم ها از آنها به نحو خوبی استفاده کنند [۴۴].

الگوریتم جستجوی گرانشی (GSA[۵۸]) یک الگوریتم بهینه‌سازی مکاشفه‌ای مبتنی بر جمعیت است که از عامل های جستجوگر تشکیل‌شده و با یکدیگر کنش و واکنش دارند [۴۵]. که در آن عوامل به‌عنوان اشیا در نظر گرفته می‌شوند و عملکرد آن‌ها با معیار میزان جرم‌شان سنجیده می‌شوند. نیروی جاذبه، جابجایی سراسری را باعث می‌شود و همه اجرام به سمت جرم‌های سنگین‌تر جذب می‌شوند. GSA می‌تواند به عنوان یک سیستم ایزوله شامل یک سری جرم در نظر گرفت. این سیستم مثل یک جهان مصنوعی کوچک است که اجرام از قوانین نیوتون شامل قانون جاذبه و حرکت پیروی می‌کنند. به طور دقیق تر اجرام از قوانین زیر پیروی می‌کنند.

- **قانون جاذبه**: هر ذره سعی دارد ذرات دیگر را به یک نیرویی جذب کند. این نیرو طبق رابطه (۸) محاسبه می‌شود.

$$F = G\ (M_1 M_2 / R_2) \quad (۸)$$

- **قانون حرکت**: سرعت فعلی هر جسم متأثر است از سرعت قبلی خود و شتابی که جسم در حالت کنونی دارد. این سرعت مطابق رابطه (۹) است.

$$a = F/M \quad (۹)$$

1. **Generate** the initial population using test dataset
2. **Evaluate** the fitness value of the population
3. **Calculate** M, F and a of the population
4. **Update** the velocity and position of the population
5. **IF** termination criteria **is met** pass the solution to the heuristic search **Otherwise** go to step 2

الگوریتم ۱- الگوریتم جستجوی گرانشی

براساس رابطه (۸)، F مقدار نیروی گرانشی که دو جسم به هم وارد می‌کنند، G ثابت گرانشی، M1 و M2 جسم‌های اول و دوم را نمایش می‌دهند. R فاصله بین دو جسم را نشان می‌دهد.

رابطه (۸) قانون نیوتون را نمایش می‌دهد که طبق این قانون، مقدار نیروی بین دو جسم با ضرب جرم دو جسم رابطه مستقیم و با عکس فاصله بین دو جسم رابطه معکوس دارد. رابطه (۹) نشان می‌دهد وقتی‌که به یک جسم نیرویی وارد می‌شود در همان جهت به جسم شتاب وارد می‌شود.

در الگوریتم جستجوی گرانشی، عامل‌ها دارای چهار پارامتر مکان[۵۹]، جرم لختی[۶۰]، جرم گرانشی فعال[۶۱] و جرم گرانشی غیرفعال[۶۲] هستند [۴۵]. مکان هر جسم نشان‌دهنده یک راه‌حل برای مساله است. گرانش و جرم لختی با یک تابع تناسب[۶۳]تعیین می‌شوند. الگوریتم با جرم لختی و گرانش جهت‌دهی می‌شود درحالی‌که هر جسم خود یک راه‌حل است. اجسام سنگین‌تر بقیه اجسام را به سمت خود جذب می‌کنند. بنابر این در فضای مساله اجسام سنگین‌تر یک راه‌حل بهینه برای مساله هستند.

همانطور که از قوانین نیوتون می‌دانیم جرم لختی جسم در برابر حرکت جسم مقاومت می‌کند یعنی اینکه جسمی که جرمش بیش‌تر باشد حرکتش کندتر است. بنابراین عامل‌های با جرم بیشتر به آرامی جابجا می‌شوند از این رو فضای محلی بیشتری را جستجو می‌کنند. ثابت گرانش، دقت الگوریتم را تنظیم می‌کند این مورد شبیه به دما در الگوریتم SA[۶۴] است. این الگوریتم یک الگوریتم با حافظه کم است با این حال می‌تواند شبیه به یک الگوریتم با حافظه بالا کارآمد باشد.

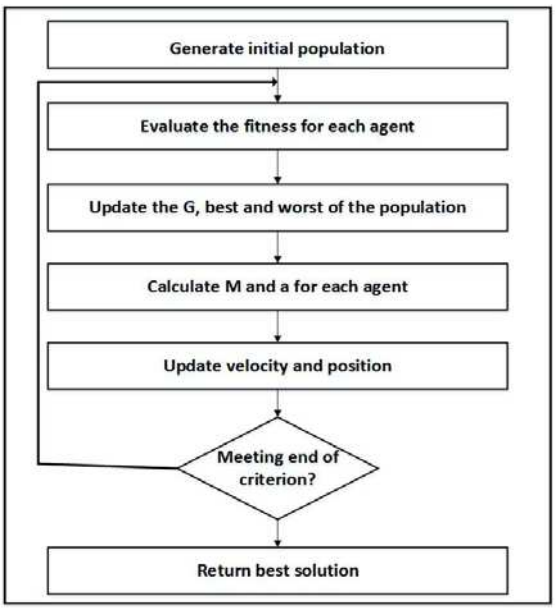
ما در اینجا فرض می‌کنیم که ثابت گرانشی برای همه یکسان است. جرم اینرسی بیشتر یک حرکت کُند از عامل‌ها را در فضای جستجوی خود فراهم می‌کند که همین امر باعث می‌شود عمل جستجو دقیق‌تر باشد. در مقابل جرم گرانشی بیشتر باعث جاذبه بیشتر عامل‌ها می‌شود و همین امر باعث می‌شود که عمل همگرایی زودتر صورت گیرد. مراحل الگوریتم جستجوی گرانشی به شرح الگوریتم ۱ عبارت‌اند از: (مرحله ۱) تعریف اولیه عامل، (مرحله ۲) محاسبه بهترین میزانبرازش، (مرحله ۳) محاسبه ثابت گرانش، (مرحله ۴) محاسبه جرم عامل‌ها، (مرحله ۵) محاسبه شتاب عامل، (مرحله ۶) سرعت و مکان عامل‌ها، (مرحله ۷) تکرار مراحل ۲ تا ۶ است. در ادامه به شرح مراحل می‌پردازیم.

مرحله ۱: تعریف اولیه عامل

مکان N عدد از عامل‌ها به‌صورت تصادفی تعیین می‌شوند. برای تعیین مکان عامل‌ها از رابطه (۱۰) استفاده می‌کنیم.

$$X_i = (x_i^1, \dots x_i^d, \dots, x_i^n)\ \ for i = 1, 2, \dots, N \quad (۱۰)$$

که در این رابطه x_i^d مکان iامین عامل در dامین بعد را بیان می‌کند در حالی‌که n تعداد بعد را نشان می‌دهد.



شکل۳ - الگوریتم جستجوی گرانشی

مرحله ۲: محاسبه بهترین میزان برازش

در این مرحله بسته به نوع مساله کمینه یا بیشینه بودن مساله، محاسبه میزان برازش[۶۵] با محاسبه بهترین حالت یا بدترین حالت تابع حالت انجام می‌شود.

اگر مساله کمینه باشد:

$$best(t) = \min fit_j(t) \quad j = \{1,\ldots,N\}$$

$$worst(t) = max\, fit_j(t) \quad j = \{1,\ldots,N\}$$

اگر مساله بیشینه باشد:

$$best(t) = max\, fit_j(t) \quad j = \{1,\ldots,N\}$$

$$worst(t) = \min fit_j(t) \quad j = \{1,\ldots,N\}$$

$fit_j(t)$ میزان برازش j امین عامل را نشان می‌دهد. best(t) و worst(t) بهترین و بدترین میزان برازش را نشان می‌دهد.

مرحله ۳: محاسبه ثابت گرانش

ثابت گرانشی به‌صورت رابطه (۱۱) محاسبه می‌شود:

$$G(t) = G_0 e^{(-\alpha t/T)} \qquad (۱۱)$$

که در اینجا G_0 و α ثابت است که در ابتدای برنامه تعیین می‌شوند. T تعداد گام‌های تکرار است.

مرحله ۴: محاسبه جرم عامل‌ها

$$M_{ai} = M_{pi} = M_{ii} = M_i, i = 1, 2, \ldots, N \qquad (۱۲)$$

$$m_i = \frac{fit_i(t) + worst(t)}{best(t) - worst(t)}$$

$$M_i = \frac{m_{i(t)}}{\sum_{j=1}^{N} m_j(t)}$$

که M_{ai} و M_{pi} به ترتیب جرم فعال و غیرفعال هستند درحالی‌که M_{ii} جرم لختی عامل i است.

مرحله ۵: محاسبه شتاب عامل

شتاب i امین عامل در گام t به‌صورت رابطه (۱۳) محاسبه می‌شود.

$$a_i^d(t) = \frac{F_i^d(t)}{M_{ii}(t)} \qquad (۱۳)$$

$F_i^d(t)$ کل نیروی وارده به عامل i ام است که به‌صورت رابطه (۱۴) محاسبه می‌شود.

$$F_i^d = \sum_{j\ \epsilon Kbest, j!=i} rand_j F_{ij}^d(t) \qquad (۱۴)$$

Kbest مجموعه k عامل اول با بهترین مقدار تناسب و بزرگ‌ترین جرم است. Kbest بازمان به‌صورت خطی کاهش می‌یابد. $F_{ij}^d(t)$ به‌صورت رابطه (۱۵) محاسبه می‌شود.

$$F_{ij}^d(t) = G(t).\left(M_{pi}(t)\ x\ M_{aj}(t)/R_{ij}(t) + \varepsilon\right).\left(x_j^d(t) - x_i^d(t)\right) \qquad (۱۵)$$

$F_{ij}^d(t)$ نیروی وارده از عامل j به i در بعد d و در گام t است. $R_{ij}(t)$ فاصله اقلیدسی بین دو عامل i و j است. G(t) ثابت گرانشی است که ثابت است.

مرحله ۶: سرعت و مکان عامل‌ها

سرعت و مکان هر عامل در گام (t + 1) بر اساس رابطه (۱۶) محاسبه می‌شود.

$$v_i^d(t+1) = rand_i x\ v_i^d(t) + a_i^d(t)$$
$$x_i^d(t+1) = x_i^d(t) + v_i^d(t+1) \qquad (۱۶)$$

v_i^d سرعت عامل iام در dامین بعد را نشان می‌دهد. a_i^d شتاب iامین عامل در dامین بعد را نشان می‌دهد. x_i^d مکان iامین عامل در dامین بعد را نشان می‌دهد.

مرحله ۷: تکرار مراحل ۲ تا ۶

مراحل ۲ تا ۶ تا زمانی که به بهترین جواب نرسیده‌ایم تکرار می‌شود. بهترین مقدار تابع تناسب در نهایت به‌عنوان تناسب نهایی در نظر گرفته می‌شود. نمودار و روند کلی کار در شکل۳ نشان داده شده است [۴۵].

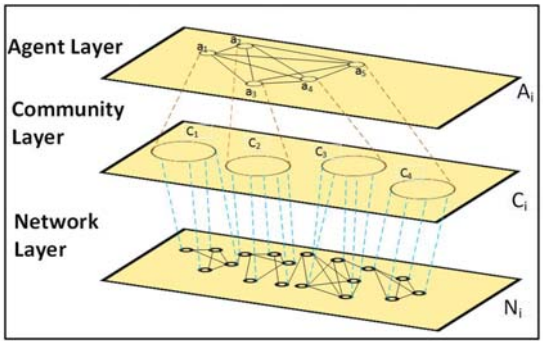
۶- روش پیشنهادی پیش‌بینی لینک

در این بخش ابتدا به معرفی روش پیشنهادی در قالب دو بخش معماری و مدل جدید پیش‌بینی می‌پردازیم. روش، ابعاد آن، مدل جدید گرانشی پیش‌بینی لینک، الگوریتم‌ها و معیارهای لازم مساله را ارایه خواهیم داد.

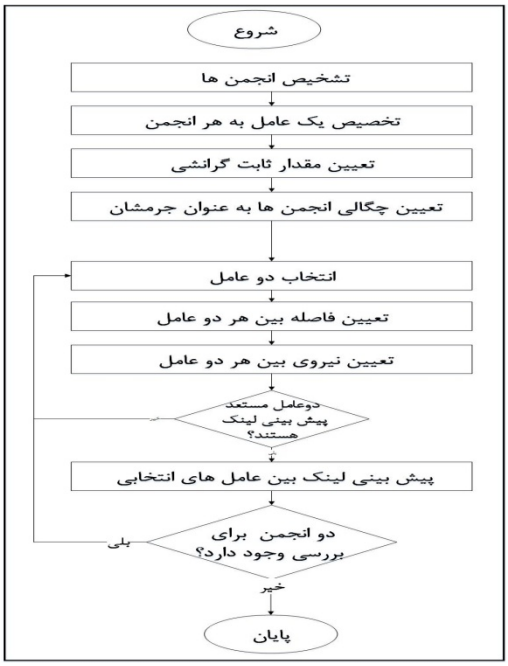
۶-۱- معماری روش پیش‌بینی لینک

قبل از تعریف سیستم مبتنی بر چند عامل، ابتدا به معرفی عامل معرفی می‌پردازیم. عامل[۶۶] یک سیستم کامپیوتری است که قادر به انجام اعمال مستقل و خودکار از طرف کاربر یا صاحب خود است. نکته اصلی در مورد عامل‌ها،

خودمختاری آن‌ها است. به این معنا که عامل‌ها قادرند به‌طور مستقل عمل کنند و روی حالات داخلی خود کنترل داشته باشند. به این ترتیب تعریف دیگری برای عامل می‌توان ارایه داد یعنی عامل یک برنامه کامپیوتری است که ضمن دارا بودن قدرت تصمیم‌گیری مستقل و درحال مشورت با دیگر عامل‌ها خواهد بود [۴۷]. بر این اساس معماری پیشنهادی ما از سه لایه شبکه اجتماعی، انجمن‌ها اجتماعی و عامل‌ها تصمیم‌گیرنده تشکیل شده است (شکل ۵). در پایین‌ترین لایه، لایه شبکه اجتماعی قرار دارد (لایه شبکه در شکل ۴). لایه میانی همان لایه انجمن‌ها است (لایه انجمن در شکل ۴). در بالاترین لایه، لایه عامل‌ها یا سامانه چندعاملی ما قرار دارد (لایه عامل در شکل ۴). یک سامانه چندعاملی[۶۷] سامانه‌ای است که از چندین عامل هوشمند تعاملی تشکیل‌شده باشد (لایه عامل در شکل ۴). از سامانه‌های چندعاملی می‌توان برای حل مساله‌هایی استفاده کرد که حل آن برای یک عامل منفرد یا یک سامانه‌ی یکپارچه مشکل یا غیرممکن است.



شکل ۴- معماری پیشنهادی سه لایه‌ای عامل‌گرای پیش‌بینی لینک



شکل ۵- مدل پیشنهادی پیش‌بینی گرانشی لینک

هوشمندی می‌تواند شامل رویکردهای جستجو، یافتن و پردازش روش‌مند، کاربردی، رویه‌ای، یا الگوریتمی باشد. موضوع‌هایی که تحقیق سامانه‌های چندعاملی در آن‌ها می‌تواند رویکرد مناسبی ارایه کند شامل مبادلات برخط، پاسخ‌دهی به فجایع و مدل‌سازی ساختارهای اجتماعی است. سامانه‌های چندعاملی روش‌های نوینی برای حل مسائل محاسباتی و پیاده‌سازی پروژه‌های نرم‌افزاری رایانه‌ای ارایه می‌دهند. ازآنجاکه در این‌گونه سامانه‌ها مجموعه‌ای متشکّل از چندین عامل سیستم را به وجود می‌آورد، دست‌یابی و نیل به اهدافی امکان‌پذیر می‌گردد که به‌وسیله سامانه‌های تک عاملی میسّر نیست [۴۶].

این امر مساله مقیاس‌پذیری مدل پیشنهادی را از طریق توسعه و بسط عامل‌ها امکان‌پذیر می‌سازد. مقیاس‌پذیری در حل این مساله دارای ابعاد مختلفی از قبیل عامل‌ها، امنیت[۶۸]، نقل وانتقال[۶۹]، سیاربودن[۷۰]، ارتباطات[۷۱]، مدیریت منابع[۷۲]، کشف منابع[۷۳]، کنترل[۷۴] و مدیریت داده‌ها[۷۵] است [۶۸، ۶۹]. با توجه به مساله مدل‌سازی عامل‌گرا، ما در این مقاله به مقیاس‌پذیری عامل‌ها از نظر توزیع زمان جستجو، امکان نقل وانتقال، و سیاربودن در حل مساله پیش‌بینی لینک پرداخته‌ایم. البته راه‌حل پیشنهادی همچنین دارای قابلیت مقیاس‌پذیری حل مساله از نظر کنترل و مدیریت داده‌ها است.

۶-۲- مدل پیشنهادی پیش‌بینی لینک

در این بخش به مساله نگاشت مدل گرانشی برای پیش‌بینی لینک می‌پردازیم. در روش پیشنهادی (طبق شکل ۵) ابتدا انجمن‌ها را از گراف استخراج می‌کنیم. تشخیص انجمن‌ها و استخراج آن‌ها طبق پیاده‌سازی الگوریتم و روش لوین[۷۶] صورت گرفته‌است [۷۰]. برای مدیریت و تصمیم‌سازی‌ها، پس از استخراج انجمن‌ها به هر انجمن یک عامل تخصیص می‌دهیم. یعنی به تعداد انجمن‌ها عامل ایجاد می‌کنیم و هر عامل را به یک انجمن انتساب می‌دهیم. در این مدل پس از اختصاص عامل‌ها از الگوریتم جستجوی گرانشی استفاده می‌کنیم. یعنی ابتدا مقدار ثابت گرانشی را تعیین می‌کنیم که ما آن را برابر مقدار ۹.۸ انتخاب کرده‌ایم. سپس چگالی انجمن‌ها را طبق رابطه (۱۷) محاسبه و به عنوان مقدار جرم آن‌ها قرار می‌دهیم. حال دو انجمن (یا دو عامل) را انتخاب می‌کنیم تا میزان فاصله آن‌ها (تناسب معکوس با تعداد یال‌های بین آن‌ها) را مشخص کنیم.

در ادامه نیروی بین دو عامل را با استفاده از رابطه (۸) اندازه می‌گیریم. سپس بر اساس مقدار یک آستانه (مقداری ثابت متناسب با نیروی بیشینه بین تمام عامل‌ها) میزان انتخاب دو عامل را برای پیش‌بینی تعیین می‌کنیم. اگر این نیرو از این آستانه کوچک‌تر بود دو عامل برای پیش‌بینی برگزیده نمی‌شوند. بعد از گزینش عامل‌ها، پیش‌بینی لینک بین دو عامل گزینش شده انجام می‌شود. برای تمامی زوج عامل‌ها این روال تکرار می‌شود و در نهایت فرآیند پیش‌بینی لینک کامل می‌گردد. ما در صورت انتخاب راهبرد مناسب در تخصیص پردازنده‌ها به عامل‌ها، به افزایش سرعت پاسخگویی (کاهش زمان پیش‌بینی) و مقیاس‌پذیری مساله پیش‌بینی (مقاومت در برابر افزایش عناصر و گره‌ها)، کمک می‌کنیم.

همان‌گونه که در بخش ۵ بیان شد، الگوریتم جستجوی گرانشی دارای مراحل مختلفی است که ما برای نگاشت این مدل گرانشی در مساله پیش‌بینی لینک به شرح زیر عمل می‌کنیم و معماری مناسبی ارایه می‌دهیم. در این معماری هر عامل را به‌عنوان یک جسم در نظر می‌گیریم که دارای جرم است. برای محاسبه جرم هر عامل ابتدا باید در GSA میزان برازش را به دست بیاوریم. در مدل پیشنهادی چگالی گراف را به عنوان تابع برازش در نظر می‌گیریم. چگالی گراف از رابطه (۱۷) محاسبه می‌شود که در آن |E| تعداد یال‌های یک انجمن و |V| تعداد گره‌های موجود آن انجمن است.

$$Density = \frac{2|E|}{|V|(|V|-1)} \quad (۱۷)$$

پس از محاسبه چگالی گراف، جرم هر عامل را مطابق رابطه (۱۲) محاسبه می‌کنیم و سپس نیروی بین اجرام را به دست می‌آوریم. طبق رابطه (۱۵) برای محاسبه نیروی بین دو جسم باید فاصله بین دو آن جسم را بیابیم. فاصله بین دو

جسم را به این صورت تعیین می‌کنیم که هر چه تعداد یال بین دو انجمن بیشتر باشد فاصله بین آن‌ها کمتر در نظر گرفته می‌شود. به این صورت فاصله بین دو عامل را به‌عنوان معکوس تعداد یال بین دو عامل در نظر می‌گیریم. لذا نیروی بین دو عامل را به دست می‌آوریم. به‌این‌ترتیب می‌توان عامل‌ها را به‌تناسب نیروی بین گرانشی بین آن‌ها برای محاسبات پیش‌بینی لینک در نظر گرفت. دوعاملی که نیروی بین آن‌ها کم‌تر از یک حد آستانه معین است، از محاسبات پیش‌بینی لینک حذف می‌شود.

۷- نتایج آزمایش‌های تجربی

ما نتایج آزمایش‌های تجربی خود را در سه مرحله و به این شرح ارایه می‌دهیم. ابتدا تنظیمات آزمایشات، ابزار، و مجموعه‌های داده‌ای مورد بررسی و ساختار آن‌ها را معرفی می‌کنیم. سپس به معیارهای ارزیابی آزمایش‌ها و الگوریتم می‌پردازیم. نهایتا موارد مورد بررسی در آزمایش‌های تجربی و تحلیل نتایج آن‌ها را بیان می‌کنیم.

جدول ۱- مشخصات مجموعه‌های داده‌ای آزمایش‌ها

مجموعه داده	شبکه در زمان t		شبکه در زمان t+1	
	تعدادگره	تعدادیال	تعدادگره	تعداد یال
CiteSeer	۳۸۴۴۱۳	۱۷۵۱۴۶۳	۳۸۴۴۱۳	۱۹۲۶۳۷۹
hep-ph	۳۴۵۴۶	۴۲۱۵۷۸	۳۴۵۴۶	۴۸۶۹۸۲
Wiki	۱۶۷۵۲۵	۱۱۶۴۵۷۶	۱۶۷۵۲۵	۱۲۵۹۶۵۹
emails	۱۶۷	۵۷۸۴	۱۶۷	۵۹۵۶
movielens	۲۶۲۵	۱۰۰۰۰۰	۲۶۲۵	۱۲۳۱۴۵
US airports	۱۵۷۴	۲۸۲۳۶	۱۵۷۴	۳۱۵۸۴

۷-۱- تنظیمات آزمایش‌ها

ما مدل‌سازی و معماری پیشنهادی مساله و الگوریتم جستجوی گرانشی مورد استفاده را در ابزار NetBeans IDE و با زبان Java پیاده‌سازی و بوسیله مجموعه‌های داده‌ای مختلف hep-ph، CiteSeer، Wikibooks، Manufacturing emails، maayan-faa و movielens-100k_rating مورد ارزیابی قرار می‌دهیم. سیستم مورد آزمایش یک ماشین core i7، 2.79GHz و با حافظه اصلی 4G RAM با سیستم عامل Win7 است. در جدول۱ مشخصات و در زیر شرح بیشتر مجموعه‌های داده‌ای آورده شده است. هر مجموعه داده‌ای ارائه‌کننده یک کاربرد است.

از آنجا که روش‌های پیش‌بینی لینک وابسته به ساختار و نوع شبکه هستند لذا نمی‌توان یک روش غالب پیش‌بینی لینک را برای تمامی شبکه‌ها ارایه داد. بنابراین انتخاب مناسب مجموعه‌های داده‌ای مختلف می‌تواند به رفع این نقیصه کمک کند. لذا ما نخست انواع مختلف مجموعه‌های داده‌ای با ویژگی‌های ساختاری متفاوت را برای آزمون الگوریتم‌ها انتخاب می‌کنیم و سپس در ادامه روش‌های مختلف ارزیابی را برای مقایسه الگوریتم‌های پیش‌بینی لینک ارایه خواهیم داد. در ادامه مجموعه‌های داده‌ای مورد استفاده معرفی خواهد شد.

مجموعه داده‌ای hep-ph مربوط به شبکه نشریات مرتبط با موضوعات فیزیک است که در آن هر گره به عنوان یک نشریه فرض شده و لینک بین آن‌ها به عنوان ارتباط بین نشریات مختلف در نظر گرفته شده است. این مجموعه داده ای ۳۴۵۴۶ گره دارد. مجموعه داده‌ای CiteSeer مبتنی بر شبکه استنادها و ارجاع به منابع است و از کتابخانه دیجیتالی CiteSeer این داده‌ها استخراج شده‌اند. گره‌ها در این مجموعه داده‌ای نشریات هستند و یال‌ها استنادها یا ارجاع‌هایی است که به این نشریات شده است. نشریات خود دارای مجموعه داده‌ای هستند و به این ترتیب در این مجموعه داده‌ای حلقه داریم. این مجموعه داده‌ای شامل ۳۸۴۴۱۳ گره است.

مجموعه داده‌ای Wikibooks شبکه دوبخشی، بخش ویرایش‌گران و بخش از صفحات ویکی‌پدیای زبان انگلیسی است. این شبکه شامل کاربران و صفحات ویکی‌پدیا است. در این شبکه هر یال به عنوان یک فعالیت ویرایشی در نظر گرفته می‌شود و گره‌ها نیز کاربران و صفحات ویکی‌پدیا فرض می‌شوند. در این شبکه ۳۲۵۸۳ کاربر و ۱۳۴۹۴۲ مقاله و صفحه وجود دارد. مجموعه داده‌ای Manufacturing emails مربوط به ایمیل‌های رد و بدل‌شده بین کارمندان یک کارخانه است. در این شبکه گره‌ها به‌عنوان کارمندان در نظر گرفته می‌شوند و یال بین دو گره همان ایمیل مورد مبادله است. یعنی سمت چپ هر یال فرستنده ایمیل، سمت راست آن یال گیرنده ایمیل، و یال نیز ایمیل مورد مبادله در نظر گرفته شده است. این شبکه شامل ۱۶۷ گره است.

شبکه movielens یک شبکه دوبخشی است که در آن گره‌های گراف همان کاربران و فیلم‌ها هستند. در این شبکه امتیازی که کاربران به فیلم‌ها می‌دهند به عنوان یال بین گره‌ها در شبکه تعریف شده است. این شبکه شامل ۹۴۳ کاربر و ۱۶۸۲ فیلم است که در مجموع ۲۶۲۵ گره در این شبکه وجود دارد. شبکه US airports شبکه پروازی بین فرودگاه‌های مختلف ایالات متحده آمریکا در سال ۲۰۱۰ است. در این شبکه هر گره به عنوان یک فرودگاه در نظر گرفته شده و یال بین گره‌ها به عنوان ارتباطات بین فرودگاه‌های مختلف فرض شده است که این شبکه دارای ۱۵۷۴ گره است.

۷-۲- معیارهای ارزیابی

همان‌گون که بیان شد لینک‌های مورد پیش‌بینی در یک الگوریتم پیش‌بینی لینک را می‌توان به چهار دسته TP، FP، TN و FN تقسیم‌بندی نمود. معمولا برای ارزیابی هر الگویتم از چند معیار مختلف استفاده می‌شود که این معیارها عبارت از صحت[۷۷]، دقت[۷۸]، فراخوانی[۷۹] و معیار f هستند. از این معیارها برای ارزیابی نتایج الگوریتم‌های پیش‌بینی لینک نیز می‌توان استفاده کرد. ما از صحت و دقت استفاده و آن‌ها را به صورت روابط (۱۸) و (۱۹) محاسبه می‌کنیم. منظور از صحت، میزان درجه نزدیکی یک اندازه به مقدار واقعی خودش است. این معیار حاکی از نزدیک بودن نتیجه یک آزمایش به ارزش واقعی مقدار خودش می‌باشد. منظور از دقت درجه‌ای است که تکرار آزمایش‌ها، نتایج یکسانی را نشان می‌دهند. یعنی نشان‌دهنده میزان احتمال یا ارتباط لینک مورد پیش‌بینی با لینک واقعی است [۶۷].

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \quad (۱۸)$$

$$Precision = \frac{TP}{(TP + FP)} \quad (۱۹)$$

۷-۳- تحلیل آزمایش‌ها

برای پیش‌بینی لینک نیاز داریم که مجموعه‌های داده‌ای ما در زمان رخ دادن، کل یال‌های آن معلوم باشد ولی به دلیل عدم دسترسی به چنین منابعی، مجموعه‌های داده‌ای را بر حسب بازه زمانی با واحدهای مختلف ایجاد کرده‌ایم. مثلا از آنجا که سال انتشار نظریات فیزیک در دسترس بوده است، برای شبکه hep-ph بازه‌های زمانی را بر حسب سال در نظر گرفته‌ایم. پس در ابتدا ما باید مجموعه داده‌ای را بر

حسب بازه‌های زمانی مناسب واحدهای آن داده‌ها دسته‌بندی کنیم. هرچند که در برخی از مجموعه‌های داده‌ای به علت عدم وجود اطلاعات از زمان رخدادن لینک‌ها، دو مجموعه آموزشی و تست را به کمک یک فرآیند تصادفی و با حذف یک سری از لینک‌ها در مجموعه آموزشی و اضافه نمودن آن‌ها به مجموعه‌ی آزمون، آن‌ها را برای پیش‌بینی لینک آماده نموده‌ایم. در این بخش روش سنتی (آدامیک‌آدار) و روش پیشنهادی پیش‌بینی لینک مورد مقایسه و بررسی قرار می‌گیرند. میزان دقت و صحت روش پیشنهادی ما با الگوریتم آدامیک‌آدار که یکی از روش‌های همسایگی است به صورت جداول ۲ و ۳ سنجیده شده است. در این جداول نتایج آزمون دو روش را روی مجموعه داده‌ای بخش ۷-۱ ارایه شده است.

برای ارزیابی مدل چندعاملی و اثر آن در عملیات پیش‌بینی از ساختار تک عاملی و چند عاملی برای یک مجموعه داده‌ای یکسان بهره برده‌ایم که نتایج بهتر با بیش از یک عامل حاصل می گردد. میزان دقت در روش سنتی ۴۳ درصد است و در روش پیشنهادی ۶۹ درصد است (شکل ۶). به این ترتیب بهبودی که به طور میانگین روی مجموعه داده‌ای مختلف جدول ۱ به میزان ۲۶ درصد است. میزان صحت در روش سنتی ۳۵ درصد است و در روش پیشنهادی ۶۸ درصد است (شکل ۷). به این ترتیب بهبودی که روی دقت الگوریتم‌ها صورت گرفته به‌طور میانگین ۳۳ درصد است.

با توجه به تناسب ساختاری معماری، الگوریتم و مقیاس‌پذیری شبکه مورد بررسی به نکاتی در باره معماری، الگوریتم و حجم شبکه مورد مطالعه اشاره می‌کنیم که برای بسط و توسعه مدل پیشنهادی جهت پوشش داده‌های بزرگ و هوشمندی عامل‌ها قابل توجه است. یکی از معیارهای مهم برای یک الگوریتم پیش‌بینی لینک، داشتن زمان اجرای معقول است هر چند که با افزایش ابعاد شبکه‌های اجتماعی، زمان مورد نیاز برای پیش‌بینی‌ها بیشتر می‌شود. الگوریتمی مقیاس‌پذیر[۸۰] است که با افزایش نمایی اندازه شبکه، زمان اجرای آن به صورت نمایی افزایش پیدا نکند. پژوهشگران مختلفی روی مقیاس پذیری کارهای مختلفی را انجام داده‌اند. برای مقیاس‌پذیری الگوریتم‌ها می‌توان از روش‌های موازی‌سازی مانند نگاشت‌کاهش[۸۱] به همراه روش‌های تجزیه گراف‌ها استفاده نمود. پیاده‌سازی این روش‌ها یعنی مقیاس‌پذیری در قالب الگوریتم‌های کنونی کاری بسیار دشوار است. در مقابل، روش پیشنهادی ما به سهولت قابل پیاده‌سازی و ضمنا از نظر توسعه و توزیع عامل‌ها مقیاس‌پذیر است. از زمان پاسخ خوبی نسبت به روش‌های مرسوم برخوردار است که توسعه ابعاد "مساله پیش‌بینی" خلل اساسی به آن وارد نمی‌کند.

۸- نتیجه‌گیری

ما در این مقاله روش جدید پیش‌بینی لینک از نوع ساختاری بدون ناظر ارایه دادیم که اصول آن بر مدل عامل‌گرا و الگوریتم جستجوی گرانشی قرار دارد. مساله انتخاب انجمن‌ها و انتساب عامل‌ها به آن‌ها از محورهای معماری پیشنهادی روش ارایه شده است که بر کارآیی و سرعت پاسخ مورد نیاز اثر دارد. ضمنا روش مدل‌سازی چندعاملی ارایه‌شده فراهم‌کننده کارآیی و برهم کنش عامل‌ها مخصوصا در انتخاب انجمن‌ها از جمعیت مورد مطالعه و مشاوره در حل هم‌زمان پیش‌بینی لینک‌ها در هر انجمن و ادغام انجمن‌ها است. این ساختار فراهم‌آورنده قدرت پردازش‌موازی هم‌زمان، مشاوره در ادغام انجمن‌های مجزا برای ایجاد جامعه یا انجمن جدید و قدرت پیش‌بینی درونی و بیرونی لینک‌ها است که به دقت و صحت پیش‌بینی افزوده و در کاهش زمان آن کمک بسیار می‌کند.

مجموعه‌های داده‌ای مختلفی بر روی روش ارایه شده مورد آزمایش قرار گرفتند تا ضمن تایید صحت، میزان کارآیی آن مورد بررسی قرار گرفت. نتایج حاصل از آزمایش‌ها حکایت از افزایش مطلوب در دقت و صحت عملکرد روش دارد.

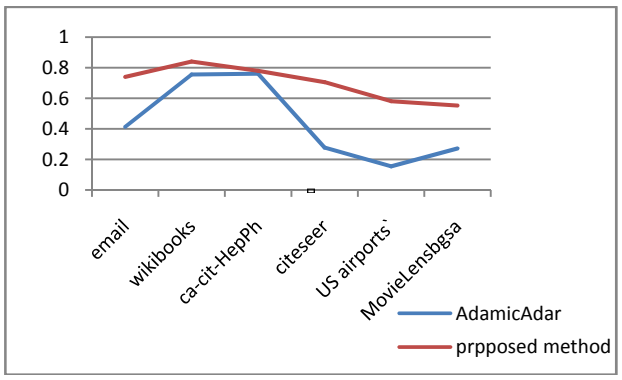
همچنین روش توزیعی پیشنهادی تداعی‌کننده یک مدل مناسب و یک الگوریتم مساعد برای رفع پاره‌ای از مشکلات در مساله مقیاس‌پذیری شبکه‌های اجتماعی براساس پیش‌بینی ساختاری است. تاکید ما در برنامه آینده تحقیقاتی خود به ارایه یک مدل فرمال پیش‌بینی لینک برای قالب کار ارایه شده است. همچنین تلاشی برای تناسب ساختار مقیاس‌پذیری آن از نظر انتقال و تحرک عامل‌ها صورت خواهیم داد تا برای پشتیبانی از داده‌های بزرگ از قابلیت مناسب برخوردار باشد. ضمنا بر روی مدل‌سازی دقیق‌تر و تناسب بیشتر عامل‌ها با معیار جاذبه جهت پشتیبانی از نیروی جاذبه متغیر و تطبیق آن با ساختار انجم‌ها سعی خواهد شد تا بتوان در کاهش مشکلات میزان جستجو عامل ها در شبکه و دقت آن راهی یافت.

جدول۲- میزان دقت مجموعه‌های داده‌ای مورد آزمایش

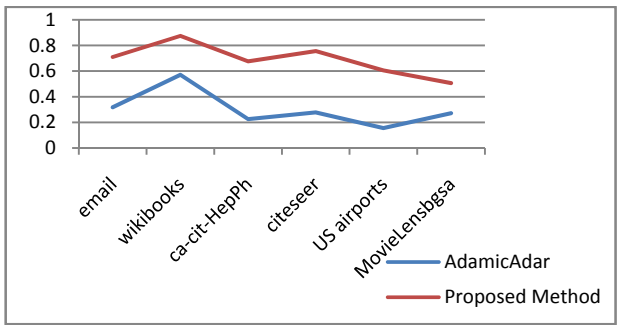
مجموعه داده‌ای / روش پیش‌بینی	روش سنتی	مدل پیشنهادی
email	۰.۴۱۲۸۱۱۴	۰.۷۴۰۰۰۶۵
wikibooks	۰.۷۵۵۳۹۵۷	۰.۸۴۰۰۷۱۹
ca-cit-HepPh	۰.۷۶۱۰۰۶۳	۰.۷۷۹۲۱۳۸
citeseer	۰.۲۷۷۶۴۳۵	۰.۷۰۴۸۰۰۲
US airports	۰.۱۵۵۰۳۲۵	۰.۵۸۰۴۵۷۳
MovieLensbgsa	۰.۲۷۱۵۴۴۷	۰.۵۵۲۷۶۴۴
میانگین	۰.۴۳	۰.۶۹

جدول ۳- میزان صحت مجموعه‌های داده‌ای مورد آزمایش

مجموعه داده‌ای / روش پیش‌بینی	روش سنتی	مدل پیشنهادی
email	۰.۳۱۷۵۱۵۳	۰.۷۰۹۵۴۹۲
wikibooks	۰.۵۷۱۱۵۳۱	۰.۸۷۴۱۸۳۱
ca-cit-HepPh	۰.۲۲۴۷۶۸۷	۰.۶۷۵۷۰۶۱
citeseer	۰.۷۴۴۷۷۳۵	۰.۷۵۶۰۲۶۹
US airports	۰.۱۱۵۷۵۳۸	۰.۶۰۴۶۳۷۴
MovieLensbgsa	۰.۱۴۵۵۵۴۷	۰.۵۰۶۵۰۵۲
میانگین	۰.۳۵	۰.۶۸



شکل ۶- نمودار مقایسه دقت بین روش سنتی و روش پیشنهادی



شکل ۷- نمودار مقایسه صحت بین روش سنتی و روش پیشنهادی

مراجع

[1] B. Furht, Handbook of Social Network Technologies and Applications, Springer, 2010.

[2] L. L. desdorff, "The Static And Dynamic Analysis of Network Data using Information Theory," Social Networks, vol. 13, pp. 301-345, 1991.

[3] L. Lü, and T. Zhou, "Link Prediction in Complex Networks: A Survey," Physica A: Statistical Mechanics and Its Applications, vol. 390, no. 6, pp. 1150–1170, 2011.

[4] D. Liben-Nowell, and J. Kleinberg, "The Link-prediction Problem for Social Networks," Journal of the American Society for Information Science and Technology, vol. 58, no.7, pp. 1019–1031, 2007.

[5] D. Yin, L. Hong, and B. D. Davison, "Structural Link Analysis and Prediction in Microblogs," Proceedings of the 20th ACM International Conference on Information and Knowledge Management, pp. 1163-1168, 2011.

[6] D. J. Hand, “Measuring Classifier Performance: a Coherent Alternative to the Area Under the ROC Curve,” Machine Learning, vol. 77, pp. 103–123, 2009.

[7] M. A. Ahmad, Z. Borbora, J. Srivastava, and N. Contractor, "Link Prediction Across Multiple Social Networks," IEEE International Conference on Data Mining Workshops (ICDMW), pp. 911–918, 2010.

[8] A. Socievole, F. De Rango, and S. Marano, "Link Prediction in Human Contact Networks Using Online Social Ties," 3rd International Conference on Cloud and Green Computing (CGC), pp. 305–312, 2013.

[9] Z. Jiawei, K. Xiangnan, and S. Yu. Philip, "Predicting Social Links for New Users across Aligned Heterogeneous Social Networks," Proc. 13th IEEE International Conference on Data Mining (ICDM '13), 2013.

[10] H. Yu, and et. al., "High-Quality Binary Protein Interaction Map of the Yeast Interactome Network," Science, vol. 322, pp. 104–110, 2008.

[11] M. P. H. Stumpf, and et. al., "Estimating the Size of the Human Interactome," Proc. Natl. Acad. Sci.U.S.A., vol. 105, pp. 6959–6964, 2008.

[12] L. A. N. Amaral, "A Truer Measure of our Ignorance," Proc. Natl. Acad. Sci. U.S.A. vol. 105, pp. 6795-6796, 2008.

[13] J. L. Schafer, and J. W. Graham, "Missing Data: Our View of the State of the Art," Psychological Methods, vol. 7, pp. 147-177, 2002.

[14] G. Kossinets, "Effects of Missing Data in Social Networks," Social Networks, vol. 28, no. 3, pp. 247–268, 2006.

[15] M. Fire, and et. al., "Link Prediction in Social Networks Using Computationally Efficient Topological Features, Privacy, Security, Risk and Trust," IEEE Third International Conference on Social Computing (SOCIALCOM), pp. 73 –80, 2011.

[16] J. Hopcroft, T. Lou, and J. Tang, "Who Will Follow You Back?: reciprocal relationship prediction," Proceedings of the 20th ACM International Conference on Information and Knowledge Management, pp. 1137–1146, 2011.

[17] S. Fortunato, "Community Detection in Graph," Physics Reports, vol. 486, pp. 75-174, 2010.

[18] X. Feng, J. Zhao, and K. Xu, "Link Prediction in Complex Networks: A Clustering Perspective," Springer Berlin/Heidelberg, vol. 85, no. 1, pp. 1-9, 2012.

[19] S. Soundarajan, J. "Hopcroft, Using Community Information to Improve the Precision of Link Prediction Methods," Proceedings of the 21st International Conference Companion on World Wide Web, ser. WWW’12 ompanion, pp. 607-608, 2012.

[20] J. Valverde-Rebaza, and A. Lopes, "Link Prediction in Complex Networks Based on Cluster Information," XXI Brazilian Symposium on Artificial Intelligence, ser. SBIA, pp. 92-101, 2012.

[21] Y. Dhote, and N. Mishra, "Survey And Analysis of Temporal Link Prediction in Online Social Networks," Advances in Computing, Communications and Informatics, pp. 1178 – 1183, 2013.

[22] D. Liben-Nowell, and J. Kleinberg, "The Link-prediction Problem for Social Networks," Journal of the American Society for Information Science and Technology, vol. 58, no. 7, pp. 1019–1031, 2007.

[23] L. Getoor, L., and C. P. Diehl, "Link Mining: A Survey," ACM SIGKDD Explorations Newsletter, vol. 7, no. 2, pp. 3–12, 2005.

[24] J. L. Gross, and J. Yellen, Handbook of Graph Theory, CRC, 1999.

[25] M. Fire, and et. al., "Link Prediction in Social Networks Using Computationally Efficient Topological Features, Security, Risk and Trust (PASSAT)," IEEE 3th International Conference on Social Computing (SocialCom), pp. 73–80, 2011.

[26] B. Bollobás, Random Graphs, Cambridge University press, 2001.

[27] P. Erdős, and A. Rényi, "On the Evolution of Random Graphs," Publication of the Mathematical Institute of the Hungarian Academy of Sciences, vol. 5, pp. 17–61, 1960.

[28] D. Watts, and S. Strogatz, "The Small World Problem," Collective Dynamics of Small-World Networks, vol. 393, pp. 440–442, 1998.

[29] B. Bollobás, Modern Graph Theory, Springer Verlag, 1998.

[30] M. Girvan, and M. E. J. Newman, "Community Structure in Social and Biological Networks," Proc. National Academy of Sciences, vol. 99, no. 12, pp. 7821–7826, 2002.

[31] S. Tsugawa, and H. Ohsaki, "Effectiveness of Link Prediction for Face-to-Face Behavioral Networks," PLoS ONE, vol. 8, no. 12, 2013.

[32] A. L. Barabâsi, and et. al., "Evolution of the Social Network of Scientific Collaborations," Physica A: Statistical Mechanics and Its Applications, vol. 311, no. 3, pp. 590–614, 2002.

[33] M. E. J. Newman, "Clustering and Preferential Attachment in Growing Networks," Physical Review, vol. 64, no. 2, 2001.

[34] T. Murata, and S. Moriyasu, "Link Prediction of Social Networks Based on Weighted Proximity Measures," IEEE/WIC/ACM International Conference On Web Intelligence, pp. 85–88, 2007.

[35] G. Salton, and M. J. McGill, Introduction to Modern Information Retrieval, McGraw-Hill Book Co., 1983.

[36] L. A. Adamic, and E. Adar, "Friends and Neighbors on the Web," Social Networks, vol. 25, no. 3, pp. 211–230, 2003.

[37] T. Sørensen, "A Method of Establishing Groups of Equal Amplitude in Plant Sociology Based on Similarity of Species and Its Application to Analyses of the Vegetation on Danish Commons," Biol. Skr., vol. 5, pp. 1–34, 1948.

[38] E. Ravasz, and et. al., "Hierarchical Organization of Modularity in Metabolic Networks," Science, vol. 297, no. 5586, pp. 1551–1555, 2002.

[39] E. A. Leicht, P. Holme, and M. E. J. Newman, 'Vertex Similarity in Networks," Physical Review, vol. 73, no. 2, pp. 026120, 2006.

[40] T. Zhou, L. Lü, and Y. C. Zhang, "Predicting Missing Links via Local Information," The European Physical Journal B-Condensed Matter and Complex Systems, vol. 71, no. 4, pp. 623–630, 2009.

[41] L. Katz, "A New Status Index Derived from Sociometric Analysis," Psychometrika, vol. 18, no. 1, pp. 39–43, 1953.

[42] P. Y. Chebotarev, and E. Shamis, "The Matrix-forest Theorem and Measuring Relations in Small Social Groups," Autom. Remote Control, vol. 61, pp. 1424–1450, 2001.

[43] R. N. Lichtenwalter, J. T. Lussier, and N. V. Chawla, "New Perspectives and Methods in Link Prediction," Proc. 16thACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 243–252, 2010.

[44] E. Rashedi, H. Nezamabadi-pour, and S. Saryazdi, "GSA: A Gravitational Search Algorithm," Information Sciences, vol. 179, no. 13, pp. 2232-2248, 2009.

[45] N. M. Sabri, M. Puteh, and M. R. Mahmood, "A Review of Gravitational Search Algorithm," Int. J. Advance. Soft Comput. Appl, vol. 5, no. 3, 2013.

[46] R. J. Stuart, P. Norvig, Artificial Intelligence: A Modern Approach, Prentice Hall, 2003.

[47] B. Brewington, and et. al., "Mobile agents for distributed information retrieval, Intelligent Information Agents," Springer-Verlag, pp. 355-395, 1999.

[48] E. Matthew, and G. M. Jardins, "Social Networks and Multi-agent Organizational Performance," American Association for Artificial Intelligence, pp. 32-37, 2004.

[49] S. Sucheta, "Using Community Information to Improve the Precision of Link Prediction Methods," 21st international conference companion on World Wide Web, pp. 607-608, 2012.

[50] A. H. Mohammad, Link Prediction In Social Networks, Indiana university, Social Network Data Analytics, pp. 243-275, 2011.

[51] E. Sherkat, M. Rahgozar, and M. Asadpour, "Ant Colony Approach to Link Prediction in Social Networks," JCSE, vol. 12, no. 1, pp. 1-9, 2014.

[52] N. K. Yazdani, "A Gravitational Search Algorithm for Multimodal Optimization," Elsevier, vol. 32, pp. 176-183, 2012.

[53] P. Jun, and et. al., "Application of an Effective Modified Gravitational Search Algorithm for the Coordinated Scheduling Problem in a Two-Stage Supply Chain," Springer, vol. 70, pp. 335-348, 2014.

[54] D. Mohammad, and H. Nezamabadi-pour, "Channel Assignment in Multi-Radio Wireless Mesh Networks using an Improved Gravitational Search Algorithm," Journal of Network and Computer Applications, vol. 38, pp. 163-171, 2014.

[55] T. Niknam, F. Bavafa, and M. Jabbari, "A Novel Self-Adaptive Learning Charged System Search Algorithm for Unit Commitment Problem," Journal of Intelligent and Fuzzy Systems, vol. 26, no. 1, pp. 439-449, 2014.

[56] J. N. Wang, and X. T. Li, "An Improved Gravitational Search Algorithm for Unconstrained Optimization," Smart

Materials and Intelligent Systems, vol. 143, pp. 409-413, 2010.

[57] R. Hassanzadeh, "A Multi-objective Gravitational Search Algorithm," Proc. Communication Systems and Network Conference, pp. 7–12, 2010 .

[58] E. Rashedi, H. Nezamabadi-pour, and S. Saryazdi, "BGSA: Binary Gravitational Search Algorithm," Nat. Comput., vol. 9, pp. 727-745, 2010.

[59] B. Shaw, V. Mukherjee, and S. P. Ghoshal, "A Novel Opposition-based Gravitational Search Algorithm for Combined Economic and Emission Dispatch Problems of Power Systems," Electrical Power Energy System, vol. 35, pp. 21-33, 2012.

[60] D. J. Hand, "Measuring Classifier Performance: a Coherent Alternative to the Area Under the ROC Curve," Machine Learning, vol. 77, no. 1, pp. 103-123, 2009.

[61] A. Acquisti, and R. Gross, "Predicting Social Security Numbers from Public Data," PNAS, vol. 106, no. 27, pp. 10975–10980, 2009.

[62] X. Chen, M. Vorvoreanu, K.P.C. Madhavan, "Mining Social Media Data for Understanding Students' Learning Experiences," Learning Technologies, vol. 7, no. 3, 2014.

[63] R. Diestel, Graph Theory, Springer-Verlag, 2010.

[64] F. Benevenuto, and T. Rodrigues, "Characterizing User Behavior in Online Social Networks," Proceedings of the 9th ACM SIGCOMM conference on Internet measurement conference, pp. 49-62, 2009.

[65] L. Leydesdorff, "The Static and Dynamic Analysis of Network Data using Information Theory," Elsevier Science Publishers B.V., 1991.

[66] B. Taskar, M. F. Wong, P. Abbeel, and D. Koller, "Link Prediction in Relational Data," Advances in Neural Information Processing Systems, pp. 659-666, 2003.

[67] M. W. David, "Evaluation: From Precision, Recall And F-Measure To Roc," Informedness, Markedness & Correlation, Journal of Machine Learning Technologies, vol. 2, no. 1, pp. 37-63, 2011.

[68] F. Katsuhide, I. Takayuki, and K. Mark, "Efficient Issue-grouping Approach for Multiple Interdependent Issues Negotiation Between Exaggerator Agents," Decision Support Systems, vol. 60, pp. 10–17, 2014.

[69] V. A. Pham, and A. Karmouch, "Mobile Software Agents: An Overview," IEEE Communications Magazine, vol. 36, no. 7, pp. 26-37, 1998.

[70] V. D. Blondel, J. L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast Unfolding of Communities in Large Networks," Journal of Statistical Mechanics: Theory and Experiment, vol. 8, no. 10, pp. 10008, 2008.

اسماعیل بسطامی لیسانس خود را در رشته مهندسی کامپیوتر نرم‌افزار از دانشگاه جامع امام حسین (ع) و فوق‌لیسانس خود را در رشته معماری کامپیوتر از دانشگاه شاهد اخذ نموده و اکنون محقق دانشگاه جامع امام حسین (ع) است. تحقیقات مورد علاقه نامبرده امنیت نرم‌افزار، داده‌کاوی و شبکه‌های اجتماعی است.

آدرس پست‌الکترونیکی ایشان عبارت است از:

esbastami@ihu.ac.ir

امین‌اله مه‌آبادی تحصیلات خود را در رشته مهندسی برق سخت‌افزار و معماری کامپیوتر به انجام رسانده و اکنون استادیار گروه مهندسی کامپیوتر و فناوری اطلاعات دانشگاه شاهد است. تحقیقات مورد علاقه نامبرده شبکه‌های اجتماعی، حمل و نقل هوشمند، پردازش تصویر، سیستم بر تراشه و طراحی ابزارهای شبیه‌سازی هوشمند است.

آدرس پست‌الکترونیکی ایشان عبارت است از:

mahabadi@shahed.ac.ir

اطلاعات بررسی مقاله:

تاریخ ارسال: ۱۳۹۴/۰۴/۰۵

تاریخ اصلاح: ۱۳۹۴/۰۵/۲۵

تاریخ قبول شدن: ۱۳۹۴/۰۷/۰۱

نویسنده مرتبط: دکتر امین‌اله مه‌آبادی، دانشکده فنی مهندسی، دانشگاه شاهد، تهران، ایران.

[1] Individual
[2] Entity
[3] Collaborating
[4] Competing
[5] Actor
[6] Social Networks Analysis (SNA)
[7] Social Science
[8] Approach
[9] Structural Analysis
[10] Sociology
[11] Mathematics
[12] Computer Science
[13] Economics
[14] Communication Science
[15] Psychology
[16] Cascade
[17] Link Prediction
[18] Accuracy
[19] Precision
[20] Recall
[21] Big Data
[22] Complex Network
[23] Gravitational Search Algorithm
[24] Proximity Measures
[25] Mutation
[26] Link Prediction
[27] Interaction
[28] Training Interval
[29] Test Interval
[30] Supervised
[31] Supervised
[32] Training
[33] Degree
[34] Snapshot

[35] Distance
[36] Preferential Attachment
[37] Transitivity
[38] Common Neighbors
[39] Jaccard's Coefficient
[40] Adamic and Adar
[41] Resource Allocation
[42] Resource
[43] Random Walk
[44] Katz
[45] PropFlow
[46] Matrix Forest Index
[47] Spanning Rooted Forest
[48] Forest
[49] Spanning Forest
[50] Spanning Sub-graph
[51] Kirchhoff Matrix
[52] heuristic
[53] Simulated Annealing
[54] Particle Swarm Optimization (PSO)
[55] Stochastic
[56] Exploration
[57] Exploitation
[58] Gravitational Search Algorithm (GSA)
[59] Position
[60] Inertial mass
[61] Active Gravitational Mass
[62] Passive Gravitational Mass
[63] Fitness Function
[64] Simulated Annealing
[65] Fitness Value
[66] Agent
[67] Multi-Agent System
[68] Security
[69] Portability
[70] Mobility
[71] Communication
[72] Resource Management
[73] Resource Discovery
[74] Control
[75] Data Management
[76] Louvain
[77] Accuracy
[78] Precision
[79] Recall
[80] scalable
[81] MapReduce

The CSI Journal on Computing Science and Information Technology
Vol. 13, No. 1, 2015

A Novel Link Prediction Method in Social Networks Based-on Gravitational Search Algorithm

Esmaeil Bastami **Aminollah Mahabadi**

Department of Electrical Engineering, Shahed University, Tehran, Iran

ABSTRACT

In this paper we present a novel distributed link prediction method for social networks that is scalable and using structural properties of the network for its predictions without need to any profile. This is an agent-oriented modeling which doing good community detection and using the gravitational search algorithm to select candidate links between communities. The experimental results of prediction on various data sets show that the proposed method is scalable with on average 69% precision and 68% accuracy. Also, better scalability with improvement in speed, precision and accuracy has been reported by selecting an appropriate or optimize CPU allocation.

Keywords: Link Prediction, Gravitational Search Algorithm (GSA), Social Networks, Agent-Oriented, Distributed Model.