

مدل سازی انتشار اطلاعات در شبکه های کتابشناسی چندلایه

سما بابایی^۱، کارشناس ارشد؛ سهیلا مولایی^۲، کارشناس ارشد؛ مصطفی صالحی^۳، استادیار

۱- دانشکده علوم و فنون نوین - دانشگاه تهران - تهران - ایران - babaei.sama@ut.ac.ir

۲- دانشکده علوم و فنون نوین - دانشگاه تهران - تهران - ایران - soheila.molaei@ut.ac.ir

۳- دانشکده علوم و فنون نوین - دانشگاه تهران - تهران - ایران - mostafa_salehi@ut.ac.ir

چکیده: بسیاری از سیستم های اطلاعاتی را می توان به شکل شبکه ای چندلایه مدل کرد که هر لایه متشکل از تعدادی گره است و این گره ها با انواع مختلفی یال به یکدیگر مرتبط می شوند. اگرچه محققان در سال های اخیر به تحلیل انتشار اطلاعات در شبکه توجه نشان داده اند، اکثر این مطالعات به شبکه های تک لایه محدود بوده است. اما در دنیای واقعی به دلیل پیچیدگی روابط بین گره ها، عموماً شاهد وجود شبکه های چندلایه هستیم. کارهای پیشین، اغلب با ساده سازی های زیاد فضای مساله همراه هستند، مثلاً در بیشتر کارها تنوع گره ها و تأثیر متقابل آن ها نادیده گرفته شده است. روش پیشنهادی این مقاله با در نظر گرفتن تأثیر لایه های مختلف بر یکدیگر به پیش بینی انتشار در شبکه های چندلایه، می پردازد. مهم ترین ویژگی روش، این است که می تواند قدرت تأثیر تمامی لایه ها را مشخص کند و میزان این تأثیر را بدون توجه به شباهت یا تفاوت گره های هر لایه نسبت به هم، اندازه گیری نماید. مدل پیشنهادی بر روی دو مجموعه داده ی واقعی پیاده سازی و با سناریوهای مختلفی ارزیابی شده است. همچنین روند فعال شدن گره های مختلف را به دست آورده و با روند رشد واقعیشان در واقعیت مقایسه کردیم و نشان دادیم که روش پیشنهادی نسبت به روش های پیشین که تمامی گره ها را از یک نوع فرض می کردند، شباهت بیشتری به واقعیت دارد و می تواند راه مناسبی برای پیش بینی انتشار در شبکه های چندلایه باشد.

واژه های کلیدی: شبکه ی چندلایه، انتشار اطلاعات، شبکه ی چندگانه، شبکه ی به هم متصل، فرآیندهای انتشار

Modeling Information Diffusion in Bibliographic Multilayer Networks

Sama Babaei¹, MSc; Soheila Molaei², MSc; Mostafa Salehi³, Assistant Professor

1- Faculty of New Sciences and Technologies, University of Tehran, Tehran, Iran, Email: babaei.sama@ut.ac.ir

2- Faculty of New Sciences and Technologies, University of Tehran, Tehran, Iran, Email: soheila.molaei@ut.ac.ir

3- Faculty of New Sciences and Technologies, University of Tehran, Tehran, Iran, Email: mostafa_salehi@ut.ac.ir

Abstract: Nowadays, most of information systems can be modeled as multilayer networks which each layer includes some nodes connected to each other by different types of links. Information diffusion in networks is the subject that researchers considered recently and they analyzed and modeled this process type in the networks. Although most of researches in this field have focused on single layer networks, but in the real world, because of the complexity of relations, most systems must be modeled as multilayer networks. In the previous works, there are much simplification in problem space, like projection all layer into one layer or negligence the mutual effect of nodes in different layers. So a new effective model for analyzing diffusion in multilayer networks is needed. This method is focused on predicting diffusion in multilayer networks, with considering mutual effect of different layers on each other. The most important specification of this proposed method, is the ability to specify power of all layers and measuring this power regardless node's similarity or difference. In fact, this model can determine the diffusion power of each type of nodes. The model is applied on two real bibliographic information networks, and experimentally demonstrated the effectiveness of this model compared with other diffusion models.

Keywords: Multilayer Networks, Information Diffusion, Multiplex Networks, Interconnected Networks, Diffusion Processes

تاریخ ارسال مقاله: ۱۳۹۶/۰۳/۰۹

تاریخ اصلاح مقاله: ۱۳۹۶/۰۷/۰۴ و ۱۳۹۶/۰۷/۱۹

تاریخ پذیرش مقاله: ۱۳۹۶/۰۹/۰۴

نام نویسنده مسئول: مصطفی صالحی

نشانی نویسنده مسئول: ایران، تهران، خیابان کارگر شمالی، بعد از پل جلال آل احمد، روبروی کوچه دهم، دانشکده علوم و فنون نوین، اتاق ۳۳۷.

۱- مقدمه

روابط مختلف، شدت‌های مختلفی نیز دارد که لازم است در مدل‌سازی مورد توجه قرار بگیرند.

در این مقاله، مدل LTM یا مدل آستانه‌ای خطی [۱۰] را برای شبکه‌های چندلایه بسط می‌دهیم تا بتوانیم روند پخش اطلاعات در این شبکه‌ها را پیگیری کنیم. مدل آستانه‌ای خطی یکی از مدل‌های انتشار مبتنی بر تصمیم است، یعنی بنابر تعدادی ویژگی، هر گره می‌تواند تصمیم بگیرد که در فرآیند انتشار درگیر بشود و یا خیر. گره‌هایی که در انتشار دخیل می‌شوند، اصطلاحاً به آن فرآیند آلوده می‌شوند. در مدل LTM هر گره یک مقدار آستانه $\theta \in [0, 1]$ را انتخاب می‌کند و اگر و فقط اگر حداقل کسر θ از همسایه‌های تاکنون رفتاری را پذیرفته باشند، او نیز این رفتار را قبول می‌کند. این مدل برای زمان‌های گسسته طراحی شده است و احتمال فعال شدن گره u در زمان $t+1$ را بر اساس مجموعه گره‌های فعال شده در زمان t محاسبه می‌کند. در واقع، در LTM، هر گره u اگر و تنها اگر وزن کلی همسایه‌های فعال شده‌اش از یک آستانه θ_u بیشتر شود، فعال می‌شود [۱۰]:

$$\sum_{v \in \Gamma(u)} w_{u,v} \delta(v, t) > \theta_u \quad (1)$$

که در این رابطه، $\Gamma(u)$ مجموعه گره‌های همسایه گره u را نشان می‌دهد و $w_{u,v}$ بیانگر قدرت تأثیر گره v بر گره u است و اگر v در زمان t فعال شده باشد $\delta(v, t) = 1$ و در غیر این صورت برابر ۰ است و θ_u نیز آستانه سیستم است. بر طبق مدل LTM، واتز نقش آستانه‌ها و ساختار شبکه بر روی انتشار اطلاعات را مورد مطالعه قرار داده است [۱۰]. مدل آستانه واتز برای شبکه‌های چندگانه نیز تعمیم داده شده است [۱۱]، [۱۲]. در بسیاری از این تعمیم‌ها با تصویر کردن تمامی لایه‌ها به یک لایه، تأثیر لایه‌های مختلف در فرآیند انتشار نادیده گرفته شده است.

اما روش پیشنهادی این مقاله، می‌تواند بدون توجه به اینکه نوع گره‌های هر لایه به یکدیگر شباهت دارند یا خیر، قدرت تأثیر تمامی لایه‌ها بر روی یکدیگر را مشخص کند و نحوه تأثیر آن‌ها بر یکدیگر را اندازه‌گیری نماید. برای بررسی انتشار در این چارچوب، مفهومی به نام فرامسیر این طور تعریف می‌شود [۸]:

تعریف ۱: فرامسیر P مسیری است که بر طرح کلی شبکه T_G (A, R) تعریف می‌شود که A نشان دهنده نوع گره‌ها و R نشان دهنده روابط بین گره‌ها (نوع یال‌ها) است و به شکل $A_1 \vec{R}_1 A_2 \vec{R}_2 \dots \vec{R}_l A_{l+1}$ نمایش داده می‌شود که یک رابطه ترکیب $R = R_1 \circ R_2 \circ \dots \circ R_l$ بین انواع A_1 و A_{l+1} را نشان می‌دهد و o نشان دهنده عملگر ترکیب بر ارتباطات است [۱۳].

برای مثال در شبکه‌ی کتابشناسی شکل ۱، رابطه‌ی شرکت کردن یک نویسنده در دو کنفرانس، به صورت فرامسیر VAV (کنفرانس، نویسنده، کنفرانس^۱) عنوان می‌شود. به هر نمونه از یک فرامسیر، برای

بسیاری از مجموعه اشیاء به هم پیوسته را می‌توان به صورت یک شبکه در نظر گرفت. پیوند بین این اشیاء، نشان‌دهنده‌ی کانال‌های ارتباطی و بستری برای گسترش اطلاعات بین آن‌ها است. اطلاعات می‌توانند از طریق این پیوندها انتقال پیدا کنند، منتشر شوند و از یک شی (گره) به همسایه‌هایش برسند. به عنوان مثال مجموعه‌ای از افراد را می‌توان به عنوان گره‌های یک شبکه در نظر گرفت که با هم ارتباط دارند. ارتباط بین این افراد می‌تواند از هر نوعی نظیر ارتباط خانوادگی، دوستی، همکاری و غیره باشد. انتشار اطلاعات موضوعی است که در زمینه‌های زیادی از جمله شبکه‌های پیچیده [۱]، [۲]، شبکه‌های اجتماعی [۳]، [۴]، شبکه‌های حوزه سلامت [۵] و غیره مورد بررسی قرار گرفته است که شبکه‌های مورد استفاده در آنها اغلب چندگانه هستند. در اغلب شبکه‌ها تنها روابط مابین گره‌ها مورد بررسی قرار گرفته است و اطلاعات اضافه در مورد ویژگی‌های زمانی و زمینه‌ای تعاملات نادیده گرفته شده است. اما اخیراً ویژگی چندلایگی در سیستم‌های اطلاعاتی و در نظر داشتن آن در تحلیل‌ها، مورد توجه محققان این حوزه قرار گرفته است. در واقع اکثر شبکه‌های جهان واقعی، چندلایه هستند. شبکه‌های چندلایه، شبکه‌هایی شامل گره‌ها و یال‌ها از انواع مختلف بوده، که این یال‌ها می‌توانند نشانگر انواع مختلفی از روابط باشند [۵-۷].

فرآیندهای انتشار (برای مثال انتشار یک شایعه و یا یک بیماری) یکی از پدیده‌های بنیادی است که در شبکه‌های چندلایه رخ می‌دهد، برای مثال می‌توان به انتشار اطلاعات در بین کاربران یک شبکه‌ی اجتماعی برخط یا انتشار بیماری بین افراد در یک شبکه‌ی ارتباط مستقیم افراد اشاره کرد. با این حال، اگرچه انتشار اطلاعات در دهه‌های اخیر توجه زیادی به بحث انتشار اطلاعات در شبکه‌های تک لایه داشته است، اما این بحث در شبکه‌های چندلایه مسأله‌ای نوظهور است و با چالش‌های فراوانی روبرو است.

برای مدل‌سازی انتشار اطلاعات در شبکه‌های چندلایه، تاکنون روش‌هایی پیشنهاد داده شده است. در یکی از کارها [۸] ماهیت یکسان بودن نوع گره‌ها و متعدد بودن نوع یال‌های شبکه مورد توجه قرار گرفته است و به این نوع شبکه، شبکه‌ی ناهمگن نام داده‌اند. در ادامه‌ی آن در مقاله‌ی [۹] انتشار اطلاعات بر روی چنین شبکه‌ای مدل‌سازی شده است. در بسیاری از کارهای پیشین، شبکه‌ی چندلایه به یک شبکه‌ی تک‌لایه تصویر شده است و ویژگی‌های متفاوت هر لایه نادیده گرفته شده است. در بخشی دیگر از کارها، هر لایه به طور جداگانه و مستقل دیده شده است که باعث می‌شود تأثیرات متقابلی که هر لایه بر دیگر لایه‌ها دارد، نادیده گرفته شود. همچنین در کارهای گذشته، تأثیر گره‌های متفاوت در انتشار بر روی یکدیگر دیده نشده و این مورد هم باعث می‌شود بخشی از اطلاعات از دست برود. چرا که گره‌ها هم می‌توانند بر گره‌های هم‌نوع خود اثر بگذارند و هم بر گره‌هایی از نوع دیگر تأثیرگذار باشند. هم‌چنین این تأثیرگذاری تحت

کارگیری روش جدید را اعلام نموده و آن‌ها را بررسی می‌کنیم و در بخش ششم نتیجه‌گیری کل مقاله انجام شده را مشاهده خواهید کرد.

۲- کارهای مرتبط

کارهای مرتبط این حوزه، انتشار اطلاعات در بسترهای متعدد را مورد بررسی قرار داده‌اند، این بسترها عبارتند از شبکه‌های ساده، شبکه‌های چندگانه، شبکه‌های به هم متصل و شبکه‌های ناهمگن.

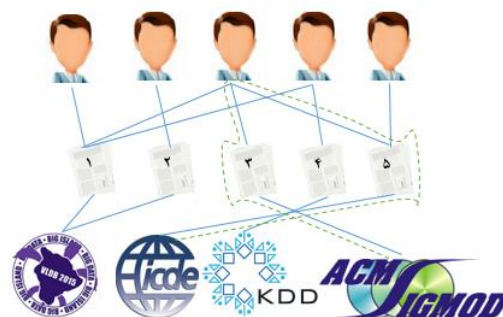
در مقاله [۱۴] یک مدل آستانه وابسته به محتوا ارائه شده است، که در آن هر نوع یال با تعصب نسبی در انتشار محتوای داده شده مرتبط شده است. در حقیقت برای هر فرآیند انتشاری خاص، یک ضریب منحصر به فرد به هر نوع یال نسبت داده می‌شود که نقش آن نوع یال در انتشار آن نوع فرآیند را نشان می‌دهد. در این مقاله مدل آستانه خطی برای شبکه‌های چندگانه تعمیم داده شده است. شبکه‌های چندگانه زیردسته‌ای از شبکه‌های چندلایه بوده که در آنها گره‌ها از یک نوع بوده و یال‌ها از انواع مختلفی هستند.

همچنین در مقاله [۱۵] برای شرایط و اندازه انتشار سراسری در یک شبکه‌ی چندگانه با انواع مختلف تعاملات، با استفاده از مدل آستانه واتز انجام شده است. در این مطالعه فرض شده است که یک گره آلوده می‌شود اگر بخشی از گره‌های آلوده شده همسایگان در هر نوعی از یال از آستانه بیشتر باشند. همچنین در این مقاله نشان داده شده است که شبکه‌های چندگانه بستر مناسبتری نسبت به شبکه‌های ساده برای انتشار سراسری هستند.

در مقاله [۱۶] از مدل آستانه خطی وابسته به محتوا برای شبکه چندگانه استفاده شده است تا برای هر نوع یال وزن به خصوصی در نظر گرفته شود و همچنین با استفاده از ضریبی، تأثیر محتوای منتشر شده را نیز در نظر گرفته است. در این کار نویسندگان با استفاده از تابع تولید، شرایط و احتمال و اندازه مورد انتظار، انتشار سراسری را به دست آورده‌اند. در واقع، این مقاله نشان می‌دهد که با افزایش خوشه‌بندی، احتمال انتشار سراسری و همچنین اندازه این انتشار کاهش می‌یابد. در حالیکه گرایش به انتشار وابسته به تغییرات میانگین درجات است. با مقایسه پویایی انتشارهای پیچیده در شبکه چندگانه و شبکه ساده تصویر شده آن، به این نتیجه رسیده‌اند که اگر تمامی یال‌ها با تنوعشان در نظر گرفته نشوند، به نتایج غیر دقیق و نامناسبی در مورد انتشار دست پیدا می‌کنیم.

برخی از محققین این حوزه شبکه‌های چندلایه را شبکه‌های اطلاعات ناهمگن نیز نامیده‌اند. تفاوت شبکه چندلایه و ناهمگن به‌طور کامل در مقاله [۱۷] بررسی شده است. در مقاله [۹] چارچوب جدیدی برای پیش‌بینی انتشار در شبکه‌های اطلاعات ناهمگن مطرح شده است. برای این منظور از تعمیم مدل انتشاری آستانه‌ی خطی بر روی این شبکه‌ها استفاده شده است. در واقع یک مدل برای انتشار اطلاعات در شبکه‌های ناهمگن (با تمایز قدرت عبور اطلاعات از نوع رابطه‌های مختلف) مطرح شده است و انتشار اطلاعات در انواع مسیرها مدل‌سازی

مثال شرکت کردن نویسنده باب در کنفرانس VLDB، به عنوان یک نمونه مسی‌رد نظر گرفته می‌شود. فرامسیری متقارن است که گره‌های دو طرف آن همجنس باشند، مانند VAV. همچنین فرامسیری نامتقارن است که گره‌های دو طرفش ناهمجنس باشند، مانند APV.



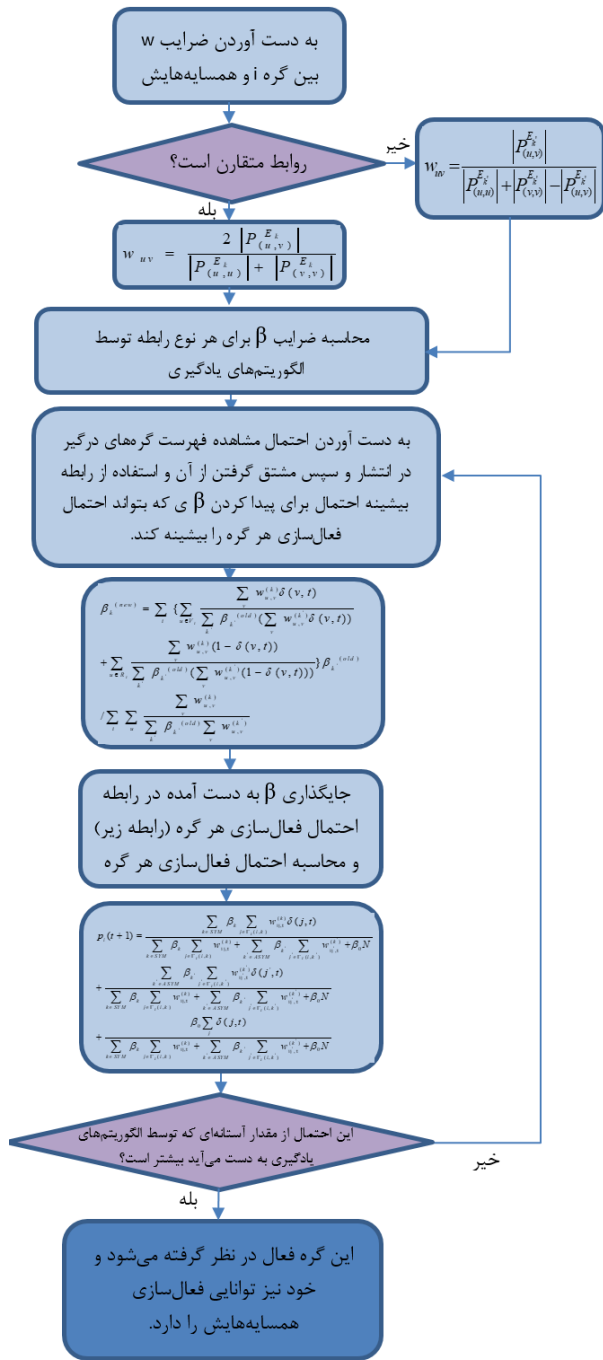
شکل ۱ - شبکه کتابشناسی DBLP، شامل انواع گره‌ها نظیر نویسنده، مقاله و کنفرانس که می‌توانند توسط فرامسیرهایی به یکدیگر متصل باشند.

با استفاده از این مفهوم، انتشار اطلاعات در شبکه‌ی چندلایه، با تمایز قدرت عبور اطلاعات از نوع رابطه‌های مختلف بررسی می‌شود. تعمیمی از مدل LTM با استفاده از این فرامسیرها می‌تواند قدرت انتشار در هر نوع فرامسیر و بنابراین در هر نوع رابطه موجود بین گره‌ها، را معین کند.

برای ارزیابی روش پیشنهادی، از دو مجموعه داده در حوزه شبکه کتابشناسی استفاده کرده‌ایم و روند حرکت فرآیند انتشاری بر روی این مجموعه داده را مورد پیش‌بینی قرار داده‌ایم و آن را با واقعیت مقایسه کرده‌ایم. نتایج به دست آمده نشان می‌دهد که تحت روش پیشنهادی نسبت به روش‌های پیشین، پیش‌بینی حدود ۱۰٪ بیشتر به واقعیت نزدیک شده است. مهم‌ترین نوآوری‌ها و دستاوردهای ارائه شده در این مقاله عبارتند از:

- ارزیابی مدلی برای پیش‌بینی انتشار در شبکه‌های چندلایه با در نظر گرفتن تفاوت نوع گره‌ها و همچنین یال‌ها
- یادگیری پارامترهای مدل تحت الگوریتم‌های یادگیری و آزمایش این پارامترها بر روی مجموعه داده‌ی واقعی برای به دست آوردن احتمال آلوده شدن هر گره به فرآیند انتشاری
- نشان دادن کارایی بیشتر مدل پیشنهادی نسبت به مدل‌های پیشین [۹] با آزمایش بر روی دو مجموعه داده‌ی واقعی ACM و DBLP

ساختار این مقاله در ادامه به شرح زیر است: در بخش دوم، به بررسی کارهای پیشین مرتبط با این حوزه می‌پردازیم و همچنین مدل‌های انتشار موجود بررسی می‌شوند. در بخش سوم، راهکار پیشنهادی را توضیح می‌دهیم و در بخش چهارم، مجموعه داده و روش‌های ارزیابی راهکار جدید را معرفی می‌کنیم. در پنجم نیز، نتایج حاصل از به

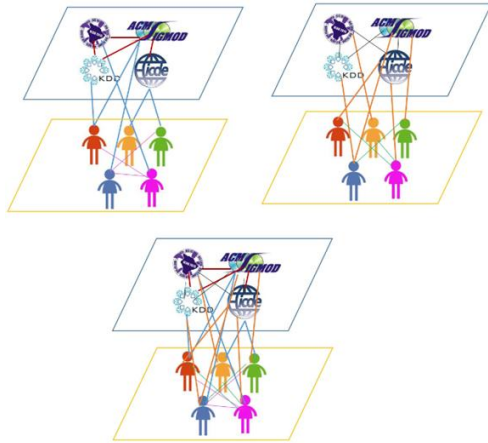


شکل ۲: روندنما پیش‌بینی و مدل‌سازی انتشار اطلاعات

می‌شود. از آنجایی که شبکه‌ی ناهمگن در دنیای واقعی به وفور وجود دارد، مساله مهم این است که بدانیم چگونه می‌توان از انواع روابط مختلف شبکه برای پیش‌بینی رفتار انتشار آینده بر روی شبکه استفاده کرد. از این جهت، اگر تنها تأثیر یکی از نوع رابطه‌ها مدل شود، مقدار زیادی از اطلاعات در دیگر نوع رابطه‌ها را از دست خواهیم داد. بر خلاف شبکه‌های ساده (که تنها دارای یک نوع گره و یک نوع یال هستند)، هر نوع رابطه با دیگر رابطه‌ها در ارتباط است و این ارتباط نقش مهمی در انتشار اطلاعات در کل شبکه دارد. اطلاعات می‌تواند توسط انواع مختلف روابط همزمان منتشر شود. بنابراین بهتر است قدرتهای انتشار مختلف برای هر نوع از روابط در نظر گرفته شود. باید در نظر داشت که اطلاعات از هر کدام از نوع رابطه‌ها می‌تواند پخش شود اما هر نوع رابطه نقش متفاوتی در فرآیند انتشار دارد.

در بسیاری از کارهای مرتبط، شبکه‌ی چندلایه به یک شبکه‌ی تک‌لایه تصویر شده است و ویژگی‌های مختلف هر لایه نادیده گرفته شده است. در بسیاری دیگر، هر لایه به طور جداگانه دیده شده است و باعث شده تأثیرات متقابلی که هر لایه بر دیگر لایه‌ها دارد، نادیده گرفته شود. همچنین تأثیر گره‌های متفاوت در انتشار بر روی یکدیگر دیده نشده [۹] و باعث از بین رفتن اطلاعات شده است، چرا که گره‌ها هم می‌توانند بر گره‌های هم‌نوع خود اثر بگذارند و هم بر گره‌هایی از نوع دیگر تأثیرگذار باشند. همچنین این تأثیرگذاری تحت روابط مختلف، شدت‌های مختلفی نیز دارد که لازم است در مدل‌سازی مورد توجه قرار بگیرند. بنابراین برای اینکه مدل‌سازی

تمام روابطی که بین هر گره با گره دیگر وجود دارد، با وزنی به خصوص در نظر گرفته می‌شوند و تأثیر آن رابطه به صورت منحصر به فرد در کنار تأثیر بقیه روابط، در نظر گرفته می‌شود. در واقع هر فرامسیر می‌تواند باعث اثرگذاری در قبول فرآیند انتشاری توسط گره‌ای خاص بشود.



شکل ۳ - انتشار با در نظر گرفتن تأثیر تمامی انواع گره‌ها و روابطشان با یکدیگر

یعنی هر دو گره‌ای که با یکدیگر در تماس هستند، تحت نوع رابطه‌ای که بین آن‌ها برقرار است، می‌توانند تأثیرات مختلفی از هم بگیرند. بنابراین باید تمام روابط در کنار هم دیده شوند و قدرت اثرگذاری همزمان تمامی آن‌ها با هم در نظر گرفته شود. مفهوم این فرض در شکل ۳ مشاهده می‌شود.

در واقع در هر کدام از دو قسمت بالایی شکل ۳، هر یال نشان‌دهنده‌ی یک فرامسیر است که بین گره‌ها برقرار می‌باشد. به منظور در نظر گرفتن تأثیر تمام گره‌ها بر یکدیگر، لازم است برای هر کدام از این فرامسیرها، یک وزن خاص در نظر گرفته شود تا وجه تمایزی بین انواع فرامسیرها وجود داشته باشد. سپس جمعیتی وزن‌دار بر روی این فرامسیرها صورت بگیرد و همه‌ی آن‌ها با وزن مخصوص به خودشان، در کنار یکدیگر دیده شوند و بر روی هم تأثیر بگذارند. در چنین شبکه‌ای که شامل تمامی فرامسیرهاست و در قسمت پایینی شکل ۳ قابل مشاهده است، به دنبال احتمال آلوده شدن هر گره به فرآیند انتشاری مدنظر هستیم. در واقع هر فرامسیر می‌تواند باعث اثرگذاری در قبول فرآیند انتشاری توسط گره‌ای خاص بشود، یعنی هر دو گره‌ای که با یکدیگر در تماس هستند، تحت نوع رابطه‌ای که بین آن‌ها برقرار است، می‌توانند تأثیرات مختلفی از هم بگیرند. بنابراین باید تمام روابط در کنار هم دیده شوند و قدرت اثرگذاری همزمان تمامی آن‌ها با هم در نظر گرفته شود.

حال LTM را که مدل مناسبی برای بررسی انتشار در شبکه‌های تک‌لایه است، به شبکه‌های چندلایه بسط می‌دهیم [۱۱]. یک شبکه‌ی چندلایه را مانند یک شبکه‌ی تک‌لایه با وزن‌های مختلف برای هر یال از هر لایه، در نظر می‌گیریم و یک مدل آستانه‌ی خطی ساده را برای به دست آوردن احتمال فعال شدن گره‌ها، به آن اعمال می‌کنیم. توجه

پیشنهادی با واقعیت همخوانی بیشتری داشته باشد، لازم است که شبکه بدون ساده‌سازی و کاملاً چندلایه در نظر گرفته شود و در فرآیند انتشار، تمام نوع گره‌های موجود در شبکه و تمام نوع روابط موجود در شبکه در کنار هم دیده شوند.

۳- راهکار پیشنهادی

در این بخش می‌خواهیم مدلی ارائه دهیم تا بتوان روند رشد گره‌های درگیر در انتشار را مشاهده کنیم. مدل‌های انتشار زیادی تاکنون مطرح شده است که می‌توان به مدل آستانه‌ی خطی [۱۰]، مدل آبخاری مستقل [۱۵]، مدل آبخاری کاهنده [۱۸]، مدل آستانه عمومی [۱۸] اشاره کرد. در اینجا بر روی مدل آستانه‌ی خطی (LTM) متمرکز می‌شویم و آن را به شبکه‌های چندلایه تعمیم می‌دهیم. مدل آستانه‌ی خطی یکی از مدل‌های انتشار مبتنی بر تصمیم است، یعنی بنا بر تعدادی ویژگی، هر گره می‌تواند تصمیم بگیرد که در فرآیند انتشار درگیر بشود و یا خیر. گره‌هایی که در انتشار دخیل می‌شوند، اصطلاحاً به آن فرآیند آلوده می‌شوند که این فرآیند انتشار می‌تواند بیماری باشد و یا شایعه، ویروس، اخبار و اطلاعات و غیره را شامل شود. مراحل این مدل‌سازی در شکل ۲ به تصویر کشیده شده است.

در این‌جا لایه‌های شبکه روی گره‌ها تعریف شده‌اند، یعنی مجموعه‌ای از گره‌های همگون یک لایه شبکه را تشکیل می‌دهند. برای مثال به شکل ۱ توجه کنید. نویسندگان یک لایه، مقالات لایه‌های دیگر و کنفرانس‌ها لایه سوم را تشکیل می‌دهند. برای مثال کنفرانس ACM SIGMOD می‌تواند تحت فرامسیر VPAPV به کنفرانس ICDE (با نمونه مسیر v4-p3-a3-p5-v2) متصل شود. اطلاعات می‌توانند هم از طریق ارتباط بین گره‌های همگون (یال‌های درون لایه‌ای) منتقل شوند و هم می‌توانند از طریق ارتباط بین گره‌های ناهمگون (یال‌های بین لایه‌ای) انتشار یابند [۱۹]. برای هر نوع یال وزنی خاص در نظر گرفته می‌شود. در حقیقت هر فرامسیری می‌تواند باعث تأثیرگذاری در رسیدن فرآیند انتشاری به گره‌ای خاص بشود، یعنی هر دو گره‌ی متصل، تحت نوع رابطه‌ای که بین آن‌ها برقرار است، می‌توانند تأثیرات مختلفی بر یکدیگر بگذارند. برای مثال در پخش شدن فرآیند چاقی، اگر فرد A با فرد B دوست صمیمی باشد، تأثیر بیشتری در چاقی فرد B دارد نسبت به زمانی که این دو، با هم همکار باشند. بنابراین، برای نشان دادن تأثیرات مختلف گره‌ها بر هم، لازم است هر فرامسیری، وزن اختصاصی داشته باشد. در مثال فوق، وزن رابطه‌ی دوستی، بیشتر از وزن رابطه‌ی همکار بودن است. اما اگر دو فرد هم با یکدیگر دوست صمیمی باشند و هم همکار، بالطبع تأثیری که فرد A در چاق شدن فرد B می‌گذارد، بیشتر از حالتی است که فقط با هم دوست صمیمی هستند یا فقط با هم همکار هستند. پس لازم است تمام روابط با هم در شبکه دیده شوند و قدرت تأثیر همزمان تمامی آن‌ها در نظر گرفته شود. گره‌ها از انواع متفاوتی هستند و روابط مختلفی بین آن‌ها وجود دارد. برای بررسی انتشار در چنین شبکه‌ای،

۴۳- احتمال فعال شدن هر گره

شبکه‌ی چندلایه را یک شبکه با وزن‌های متفاوت برای هر لایه در نظر می‌گیریم، یعنی همه یال‌ها را با هم جمع می‌کنیم و برای یال‌های لایه‌های مختلف، وزن‌های متفاوتی در نظر می‌گیریم. با این فرض، احتمال فعال شدن شیء i در مرحله‌ی زمانی $t+1$ به این صورت محاسبه می‌شود:

$$p_i(t+1) = \frac{\sum_{k \in SYM} \beta_k \sum_{j \in \Gamma_t(i,k)} w_{ij,t}^{(k)} \delta(j,t)}{\sum_{k \in SYM} \beta_k \sum_{j \in \Gamma_t(i,k)} w_{ij,t}^{(k)} + \sum_{k \in ASYM} \beta_k \sum_{j \in \Gamma_t(i,k)} w_{ij,t}^{(k)} + \beta_0 N} + \frac{\sum_{k' \in ASYM} \beta_{k'} \sum_{j \in \Gamma_t(i,k')} w_{ij,t}^{(k')} \delta(j',t)}{\sum_{k \in SYM} \beta_k \sum_{j \in \Gamma_t(i,k)} w_{ij,t}^{(k)} + \sum_{k' \in ASYM} \beta_{k'} \sum_{j \in \Gamma_t(i,k')} w_{ij,t}^{(k')} + \beta_0 N} + \frac{\beta_0 \sum_j \delta(j,t)}{\sum_{k \in SYM} \beta_k \sum_{j \in \Gamma_t(i,k)} w_{ij,t}^{(k)} + \sum_{k' \in ASYM} \beta_{k'} \sum_{j \in \Gamma_t(i,k')} w_{ij,t}^{(k')} + \beta_0 N} \quad (4)$$

که $\Gamma_t(i,k)$ بیان‌گر مجموعه‌ی همسایه‌های گره‌ی i در لایه‌ی E_k در مرحله‌ی زمانی t است و $w_{ij,t}^{(k)}$ نشان‌دهنده‌ی وزن یال بین گره‌ی i و j در لایه‌ی k است و $\delta(j,t)$ بیان‌گر این است که گره‌ی j در مرحله‌ی زمانی t فعال شده یا خیر. N تعداد کل گره‌های موجود در شبکه است. همچنین تأثیر کلی (مانند تأثیری که رسانه‌ها در انتشار اطلاعات دارند و در درون شبکه قابل ردیابی نیست) را با اضافه کردن فاکتور ثابت با وزن β_0 نشان داده‌ایم. β_k نشان دهنده‌ی وزن فراسطحی برای لایه‌ی k است.

رابطه‌ی ۴ را کمی بیشتر تشریح خواهیم کرد. برای به دست آوردن احتمال فعال شدن گره‌ی i در مرحله‌ی زمانی بعدی ($t+1$) می‌بایست نسبت پیشامد رخ داده به فضای نمونه را محاسبه کنیم که در این جا پیشامد مورد نظر، تأثیر کلی همسایه‌های فعال شده‌ی گره‌ی i است (با در نظر گرفتن تأثیر لایه‌ای هر همسایه) و فضای نمونه تأثیر کلی تمامی همسایه‌های گره‌ی i (با در نظر گرفتن تأثیر لایه‌ای هر همسایه) است.

این‌که رابطه‌ی ۴ از مجموع سه کسر تشکیل شده، به این دلیل است که کسر اول نمایان‌گر نسبت تأثیر کلی همسایه‌های همگون فعال شده‌ی گره‌ی i به تأثیر تمامی همسایه‌های همگون و ناهمگون گره‌ی i می‌باشد و کسر دوم نمایان‌گر نسبت تأثیر کلی همسایه‌های ناهمگون فعال شده‌ی گره‌ی i به تأثیر تمامی همسایه‌های همگون و ناهمگون گره‌ی i می‌باشد. کسر سوم نیز نشان‌دهنده‌ی تأثیرات اجتناب‌ناپذیر تمام گره‌های شبکه است که می‌تواند مثلاً از طریق رسانه‌ها بر گره‌ی i تأثیر بگذارند. صورت کسر اول از رابطه‌ی LTM برای یک ارتباط به دست آمده است که قدرت تأثیر آن ارتباط توسط ضریب β_k لحاظ شده است. این رابطه‌ی LTM برای تمامی ارتباطات متقارن به کار رفته است و مجموع تمامی آن‌ها میزان تأثیرپذیری گره‌ی i از تمام همسایه‌های همگون فعال آن را نشان می‌دهد. صورت کسر دوم نیز مانند صورت کسر اول از رابطه‌ی LTM برای یک ارتباط به دست

داشته باشید که چنین فرضی برای شبکه‌های چندلایه، به منظور تک‌لایه در نظر گرفتن آن نیست، چرا که هنگامی که یک شبکه‌ی چندلایه به عنوان شبکه‌ی تک‌لایه در نظر گرفته می‌شود، تمام یال‌های موجود در شبکه‌ی چندلایه به یک یال تصویر می‌شوند و اثر متفاوتی که هر یال می‌تواند داشته باشد نادیده گرفته می‌شود. فرض می‌کنیم که درصدهای وزنی همسایه‌های فعال یک گره، احتمال فعال شدن گره را معین می‌کند و آستانه‌ی فعال شدن هر گره، دارای توزیع یکنواخت در بازه‌ی $[0,1]$ است و وزن ارتباط به معنی وزن هر یال است که در ادامه این فرض‌ها را توضیح می‌دهیم.

۴۳- وزن ارتباطها

قدرت انتشار (وزن تأثیر) هر جفت گره از لایه‌های مختلف باید تخمین زده شود. به طور خاص از رابطه‌ی شباهت Pathsim [۲۴] برای محاسبه قدرت انتشار بین جفت گره‌های متقارن استفاده شده است. شباهت بین هر دو گره می‌تواند نشان‌دهنده‌ی قدرت تأثیرگذاری این دو گره بر یکدیگر باشد، چرا که هر چه دو گره بیشتر به یکدیگر شبیه باشند، توانایی تأثیرگذاری بیشتری دارند و هرچه شباهت آن‌ها به هم کمتر باشد، تأثیر کمتری بر یکدیگر می‌گذارند. در مقاله [۲۰] روشی مبتنی بر Pathsim ارائه شده است که شباهت نوع بعضی از گره‌ها به یکدیگر را نیز در بررسی شباهت گره‌هایی با نوع دیگر دخیل می‌کند، اما این مساله در بررسی انتشار به کار نمی‌آید، چرا که در بررسی انتشار، قدرت تعامل دو گره مهم است. اگر لایه‌ی E_k را در نظر بگیریم (فرامسیر متقارن)، وزن بین گره‌ی u و v به این صورت تعریف می‌شود [۲۴]:

$$w_{uv} = \frac{2|P_{(u,v)}^{E_k}|}{|P_{(u,u)}^{E_k}| + |P_{(v,v)}^{E_k}|} \quad (2)$$

$P_{(u,v)}^{E_k}$ مجموعه‌ای از فرامسیرهای متقارن (تعاملات متقارن) است که مربوط به لایه‌ی E_k است که از گره‌ی u شروع می‌شود و به گره‌ی v ختم می‌شود و همچنین $| \cdot |$ نشان‌دهنده‌ی اندازه‌ی مجموعه است. برای محاسبه‌ی قدرت انتشار بین جفت گره‌های نامتقارن از رابطه‌ی شباهت جاکارد [۲۱] استفاده شده است. اگر لایه‌ی E'_k را در نظر بگیریم (فرامسیر نامتقارن)، وزن بین گره‌ی u و v به این صورت تعریف می‌شود [۲۲]:

$$w_{uv} = \frac{|P_{(u,v)}^{E'_k}|}{|P_{(u,u)}^{E'_k}| + |P_{(v,v)}^{E'_k}| - |P_{(u,v)}^{E'_k}|} \quad (3)$$

که $P_{(u,v)}^{E'_k}$ مجموعه‌ای از فرامسیرهای نامتقارن (تعاملات نامتقارن) است که مربوط به لایه‌ی E'_k است که از گره‌ی u شروع می‌شود و به گره‌ی v ختم می‌شود.

۳-۳ آموزش پارامترها

موضوع مورد مطالعه در این بخش، این است که چگونه پارامترهای این مدل را با استفاده از فهرست گره‌های درگیر در انتشار در زمان گذشته، از طریق الگوریتم‌های یادگیری به دست آوریم.

برای هر فرآیند انتشار، فهرست گره‌های درگیر در انتشار دنباله‌ای از مجموعه گره‌های فعال و زمان فعال شدنشان است: $A = \{V_t\}_{t=1}^T$ که t نشان‌دهنده مراحل زمانی گسسته است. برای به دست آوردن بهترین پارامترها، از فرآیند یادگیری پارامترها استفاده می‌کنیم. پارامتر β_k برای هر لایه‌ی E_k را یادگرفته تا به بهترین β_k ای برسیم که در صورت قرار گرفتن در رابطه‌ی زیر، احتمال فعال شدن گره را بیشینه کند. در هر مرحله‌ی زمانی t ، احتمال فعال شدن $p_u(t)$ برای هر گره‌ی u فعال نشده ($u \in R_t$)، از توزیع برنولی تبعیت می‌کند و R_t نشان‌دهنده‌ی مجموعه گره‌هایی است که تا زمان $t-1$ فعال نشده‌اند. در واقع در مرحله‌ی زمانی t ، هر گره‌ی $u \in R_t$ با احتمال $p_u(t)$ فعال می‌شود و با احتمال $1-p_u(t)$ غیرفعال باقی می‌ماند. بنابراین احتمال مشاهده‌ی مجموعه گره‌های فعال شده V_t در زمان t ، و مجموعه گره‌های R_t که در زمان t فعال نشده‌اند به این صورت محاسبه می‌شود:

$$p(V_t, R_t) = \prod_{u \in V_t} p_u(t) \prod_{u \in R_t} (1 - p_u(t)) \quad (5)$$

به دلیل مستقل بودن رخدادهای $p_u(t)$ و $1-p_u(t)$ نسبت به خود و نیز نسبت به یکدیگر، رابطه‌ی زیر، از ضرب آن‌ها به دست آمده است. مستقل بودن هر یک از $p_u(t)$ و $1-p_u(t)$ نسبت به خود، از آن‌جا نشأت می‌گیرد که در یک مرحله‌ی زمانی فعال شدن یا فعال نشدن تمام گره‌ها از هم مستقل هستند.

احتمال مشاهده فهرست گره‌های درگیر در انتشار یک انتشار برابر است با:

$$L = \prod_{t=1}^T p(V_t, R_t) \quad (6)$$

در این رابطه نیز به علت این‌که احتمال مشاهده‌ی مجموع گره‌های فعال شده و نیز فعال نشده در مراحل زمانی مختلف از هم مستقل هستند، از ضرب استفاده شده است.

در عمل تعداد گره‌های فعال نشده، بسیار بیشتر از گره‌های فعال شده است، اما تعدادشان را برابر در نظر می‌گیریم و وزن هر گره‌ی فعال نشده را برابر با $\frac{|R_t|}{|V_t|}$ در نظر می‌گیریم.

حال الگوریتم یادگیری برای این مدل را معرفی می‌کنیم. هدف این است که بهترین β ی که می‌تواند احتمال را بیشینه کند، پیدا کنیم. بدون از دست دادن عمومیت، الگوریتم را برای حالتی که یک انتشار داریم، ارائه می‌دهیم.

۳-۴ الگوریتم یادگیری برای iLTM

برای به دست آوردن بهترین پارامتر، از فرآیند یادگیری پارامترها استفاده می‌کنیم تا به تخمین بیشینه احتمال برسیم. یعنی پارامتر β_k

آمده است، با این تفاوت که این رابطه‌ی LTM برای تمامی ارتباطات نامتقارن به کار رفته است. پس می‌بایست مجموع تمامی آن‌ها که میزان تأثیرپذیری گره‌ی i از تمام همسایه‌های ناهمگون فعال خود را نشان می‌دهد، به عنوان صورت کسر دوم لحاظ شود. صورت کسر سوم تأثیر تمام گره‌های شبکه (چه با گره‌ی i همسایه باشند و چه نباشند) را نشان می‌دهد. مخرج هر سه کسر نیز برابر با مجموع تأثیرات کلی تمامی همسایه‌های همگون و ناهمگون گره‌ی i با در نظر گرفتن قدرت تأثیرهای مختلف نوع ارتباطات متفاوت و تأثیر کلی تمامی گره‌های شبکه می‌باشد. الگوریتم پیشنهادی در شبه کد ۱ قابل مشاهده است.

شبه کد ۱: الگوریتم iLTM

Input: Datasets

Outputs: Activation probability of each inactive node

To calculate the activation probability of node i :

Choose a Database (in this case: DBLP):

for any symmetric metpath(APA,AVA,etc.) do :

Find the proper weight for any metapath by iLTM learning Algorithm, and multiple with:

for any neighbor of node i in metapath k do :

Calculate $w_{i,j}$ between i and each activated neighbor j with a similarity measure (in this case: PathSim Similarity)

Plus by:

for any asymmetric metpath(APV,VPAPA,etc.) do :

Find the proper weight for any metapath by iLTM learning Algorithm, and multiple with:

for any neighbor of node i in metapath k' do :

Calculate $w_{i,j}$ between i and each activated neighbor j with a similarity measure (in this case: Jaccard Similarity)

Plus by:

One for any activated neighbor of node i .

Devide by:

for any symmetric metpath(APA,AVA,etc.) do :

Find the proper weight for any metapath by iLTM learning Algorithm, and multiple with:

for any neighbor of node i in metapath k do :

Calculate $w_{i,j}$ between i and each neighbor j with a similarity measure (in this case: PathSim Similarity)

Plus by:

for any asymmetric metpath(APV,VPAPA,etc.) do :

Find the proper weight for any metapath by iLTM learning Algorithm, and multiple with:

for any neighbor of node i in metapath k' do :

Calculate $w_{i,j}$ between i and each neighbor j with a similarity measure (in this case: Jaccard Similarity)

Plus by:

One for any neighbor of node i .

۳-۵ پیش‌بینی عملیات

می‌توان از پارامتر β برای پیش‌بینی عملیات آینده با داشتن یک مجموعه از گره‌های فعال شده استفاده کرد. در حقیقت پارامتر β به دست آمده را در رابطه‌ی احتمال قرار می‌دهیم. این احتمال‌ها برای انجام رتبه‌بندی (چه کسی بیشتر از همه احتمال دارد که در مرحله زمانی بعدی فعال شود) و پیش‌بینی (پیش‌بینی تعداد کل گره‌های فعال شده در مرحله زمانی بعدی) به کار می‌رود. اگر احتمال فعال‌سازی به دست آمده برای هر گره، از مقدار آستانه‌ای (که این مقدار توسط الگوریتم‌های یادگیری طوری به دست می‌آید که بهترین پاسخ را بدهد) بیشتر باشد، آن گره را فعال به حساب آورده و از این پس این گره توانایی فعال‌سازی همسایه‌های غیر فعالش را دارد.

۴-۱ ارزیابی روش پیشنهادی

برای ارزیابی روش پیشنهادی و مقایسه‌ی نتایج آن با مدل‌های پیشین [۹]، پارامترهای یاد گرفته شده‌ی این مدل که از طریق الگوریتم‌های یادگیری به دست آمدند را بر روی مجموعه داده‌ی واقعی آزمایش می‌کنیم. برای این کار نیاز است مجموعه داده را معرفی کرده و نحوه‌ی ارزیابی را توضیح دهیم.

۴-۲ مجموعه داده

شبکه‌ی آزمون مورد استفاده در این مقاله، شبکه‌ای چندلایه است که دارای چندین نوع گره‌ی مختلف و چندین نوع یال مختلف می‌باشد. دو شبکه‌ی اطلاعاتی کتابشناسی واقعی DBLP^۱ و ACM^۲ (در حوزه‌ی علوم کامپیوتر) را انتخاب کردیم. مجموعه داده DBLP در بسیاری از مقالات [۲۳] این حوزه به عنوان شبکه تک‌لایه در نظر گرفته شده است. اما این شبکه‌ها گره‌هایی از نوع نویسنده، مقاله و کنفرانس دارند و می‌توان آن را به عنوان شبکه چندلایه نیز مورد استفاده قرار داد [۹]. ارتباطات بین نویسندگان بر اساس روش فرامسیر به این صورت است: همکاری در نویسندگی (APA)، داشتن همکار مشترک (APAPA) و شرکت در کنفرانس مشترک (APVPA). همچنین ارتباطاتی بین گره‌های با نوع متفاوت (گره‌های ناهمگون یا نامتقارن) نیز وجود دارد. برای مثال ارتباط بین نویسنده و کنفرانس به این معنی که یک نویسنده با نوشتن مقاله‌ای در یک کنفرانس شرکت کرده است به صورت (APV) نشان داده می‌شود. همچنین رابطه‌ی نامتقارن (APAPV) به این معنی است که یک فرد توسط یکی از همکارانش در نوشتن مقاله، که قبلاً در یک کنفرانس شرکت کرده است، با آن کنفرانس رابطه دارد و می‌تواند از آن تأثیر بپذیرد. مشخصات شبکه‌های DBLP و ACM در جدول ۱ آورده شده است که شامل مقالات سال‌های ۱۹۳۷ تا ۲۰۱۱ می‌باشد.

جدول ۱- اطلاعات مجموعه دادگان

تعداد کنفرانس‌ها	تعداد مقالات	تعداد نویسندگان	
۶۷۱۳	۱۵۷۲۲۷۸	۹۱۶۹۸۸	DBLP
۷۲۱۰	۲۰۸۴۰۵۵	۹۷۴۰۱۸	ACM

برای هر لایه‌ی E_k را مورد یادگیری قرار می‌دهیم تا به بهترین β_k برسیم که در صورت قرار گرفتن در رابطه‌ی فوق، احتمال فعال شدن گره را بیشینه کند. در مدل iLTM با قرار دادن معادله‌ی فوق در معادله‌ی قبلی به معادله‌ی زیر می‌رسیم:

$$\log L = \sum_t \left\{ \sum_{u \in V_t} \log(p_u(t)) + \sum_{u \in V_t} \log(1 - p_u(t)) \right\} \\ = \sum_t \left\{ \log \frac{\sum_{k \in SYM} \beta_k \sum_v w_{u,v}^{(k)} \delta(v,t) + \sum_{k \in ASYM} \beta_k \sum_v w_{u,v}^{(k)} \delta(v,t)}{\sum_{k \in SYM} \beta_k \sum_v w_{u,v}^{(k)} + \sum_{k \in ASYM} \beta_k \sum_v w_{u,v}^{(k)}}} \right. \\ \left. + \log \frac{\sum_{k \in SYM} \beta_k \sum_v w_{u,v}^{(k)} (1 - \delta(v,t)) + \sum_{k \in ASYM} \beta_k \sum_v w_{u,v}^{(k)} (1 - \delta(v,t))}{\sum_{k \in SYM} \beta_k \sum_v w_{u,v}^{(k)} + \sum_{k \in ASYM} \beta_k \sum_v w_{u,v}^{(k)}}} \right\} \quad (7)$$

باید β_k را تحت فرآیند یادگیری طوری به دست آوریم که نتیجه‌ی لگاریتم احتمال فوق بیشینه شود، برای این کار باید گام یادگیری (η) هوشمندانه انتخاب شود. طبق روش گرادین نزولی β_k از به روزرسانی پیاپی رابطه‌ی بازگشتی زیر به دست می‌آید:

$$\beta_k^{(new)} = \beta_k^{(old)} + \eta \frac{\partial \log L}{\partial \beta_k} \quad (8)$$

که در این رابطه $\frac{\partial \log L}{\partial \beta_k}$ مشتق اول تابع $\log L$ است:

$$\frac{\partial \log L}{\partial \beta_k} = \sum_t \sum_{u \in V_t} \frac{\sum_v w_{u,v}^{(k)} \delta(v,t)}{\sum_k \beta_k (\sum_v w_{u,v}^{(k)} \delta(v,t))} \\ + \sum_t \sum_{u \in R_t} \frac{\sum_v w_{u,v}^{(k)} (1 - \delta(v,t))}{\sum_k \beta_k (\sum_v w_{u,v}^{(k)} (1 - \delta(v,t)))} \\ - \sum_t \sum_u \frac{\sum_v w_{u,v}^{(k)}}{\sum_k \beta_k \sum_v w_{u,v}^{(k)}} \quad (9)$$

برای به دست آوردن β_k های مثبت در نظر می‌گیریم:

$$\eta = \frac{\beta_k^{(old)}}{\sum_t \sum_u \frac{\sum_v w_{u,v}^{(k)}}{\sum_k \beta_k^{(old)} \sum_v w_{u,v}^{(k)}}} \quad (10)$$

و بدین ترتیب خواهیم داشت:

$$\beta_k^{(new)} = \sum_t \left\{ \frac{\sum_v w_{u,v}^{(k)} \delta(v,t)}{\sum_k \beta_k^{(old)} (\sum_v w_{u,v}^{(k)} \delta(v,t))} \right. \\ \left. + \frac{\sum_v w_{u,v}^{(k)} (1 - \delta(v,t))}{\sum_k \beta_k^{(old)} (\sum_v w_{u,v}^{(k)} (1 - \delta(v,t)))} \right\} \beta_k^{(old)} \\ / \sum_t \sum_u \frac{\sum_v w_{u,v}^{(k)}}{\sum_k \beta_k^{(old)} \sum_v w_{u,v}^{(k)}} \quad (11)$$

۴-۴ تحلیل زمانی

محاسبه کرده و با واقعیت مقایسه می‌کنیم و معیارهای دقت و فراخوانی را به دست آورده و با استفاده از امتیاز F1، میانگین این دو معیار را سنجیده و با روش‌های پیشین [۹] مقایسه می‌کنیم. بسیاری از مقالات [۲۵] این حوزه نیز از این روش‌ها برای ارزیابی استفاده کرده‌اند. بالطبع، هر چه میزان این میانگین در یک مدل بالاتر باشد، به این معنی است که آن مدل قدرت بیشتری در فراخوانی و دقت نسبت به واقعیت دارد.

در این پژوهش، معیارهای دقت، فراخوانی و امتیاز F1 مورد استفاده قرار گرفته است. معیار دقت، نسبت تعداد نتایج درست به دست آمده نسبت به کل نتایج درست را بیان می‌کند و در واقع کارایی الگوریتم را نشان می‌دهد [۲۶]، [۲۷]. در واقع در این معیار تعداد گره‌های فعال که درست برچسب خورده است، بر تعداد تمام گره‌های فعال واقعی تقسیم می‌شود و بدین ترتیب تخمینی از نتایج درست الگوریتم به دست می‌آید.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (12)$$

در رابطه‌ی ۱۲ عبارتهای مختلف به شکل زیر تعریف می‌شود:
True Positive (TP): گره‌های فعال که الگوریتم آن‌ها را به درستی فعال در نظر گرفته است.

True Negative (TN): گره‌های غیرفعال که الگوریتم آن‌ها را به درستی غیرفعال در نظر گرفته است.

False Positive (FP): گره‌های فعال که الگوریتم آن‌ها را به اشتباه غیرفعال در نظر گرفته است.

همچنین معیار فراخوانی با رابطه‌ی ۱۳ محاسبه می‌شود [۲۶]، [۲۷].

$$\text{Recall} = \frac{TP}{TP + FN} \quad (13)$$

که در آن False Negative (FN)، گره‌های غیرفعال است که الگوریتم آن‌ها را به اشتباه فعال در نظر گرفته است. معیار سوم، امتیاز F1، نیز از رابطه‌ی ۱۴ محاسبه می‌شود:

$$F\text{-Score} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (14)$$

۵ نتایج

با اعمال مدل ارائه شده به این مجموعه داده، نتایجی به دست آمد که در ادامه به آن اشاره می‌کنیم.

۵-۴ تأثیر فرآیند انتشاری - انتشار موضوعات مختلف

انتشار چندین موضوع مختلف بر روی هر دو شبکه‌ی مورد بررسی قرار گرفت و نتایج آن در مقایسه با نتایج گرفته شده از طریق روش مقاله‌ی در جداول ۲ و ۳ آمده است. موضوعات Data Mining و Machine Learning از آن جهت انتخاب شده‌اند که در این مجموعه داده، تکرار مناسبی داشته‌اند و حجم قابل توجهی از داده برای مقایسه و نتیجه‌گیری در دسترس است. موضوعات SVM، Neural Network و Naive Bayes نیز از آن جهت انتخاب شده‌اند که موضوعات زیر

الگوریتم پیشنهادی مقاله، برای دیتاهای فوق، با کامپیوتری با حافظه‌ی ۸ گیگابایت و پردازنده‌ی Core i7- 4702MQ برای پیش‌بینی انتشار یک موضوع به طور میانگین حدود ۲۴ ساعت طول می‌کشد. در ابتدا، به دلیل استفاده از ساختار ماتریسی برای نگهداری شبکه‌ی مورد استفاده، با مشکل کمبود حافظه روبرو می‌شد که توانستیم با تبدیل این ساختار به ساختار گراف در زبان برنامه‌نویسی پایتون، این مشکل را برطرف نموده و بتوانیم در زمان خطی از این الگوریتم پاسخ بگیریم.

با فرض اینکه تعداد گره‌های شبکه، n گره باشد و میانگین تعداد مسیرهای بین هر دو گره d باشد، پیچیدگی زمانی محاسبه شباهت بین هر دو گره برابر $O(nd)$ می‌باشد [۲۴] که اگر شباهت بین تمام جفت گره‌ها را محاسبه کنیم، هزینه آن $O(n^2d)$ می‌شود. همچنین همانطور که در رابطه (۴) مشاهده می‌شود، تعداد فرامسیرهای موجود k می‌باشد و با در نظر گرفتن عملیات مختلف صورت گرفته در این رابطه، پیچیدگی زمانی رابطه (۴) برابر با $O(n^2dk)$ می‌باشد.

بهبود حاصل شده روی این الگوریتم نسبت به نمونه‌های قبلی، به دلیل در نظر گرفتن ساختار چندلایه‌ای و به هم متصل بودن شبکه است که چون گره‌های متفاوت از نظر نوع، به صورت موازی در کنار گره‌های هم‌نوع دیده می‌شوند، سربار زمانی به الگوریتم اضافه نمی‌کند و به طور موازی مورد محاسبه قرار می‌گیرد و بنابراین بر پیچیدگی زمانی افزایش نخواهد داشت.

همچنین بهبود حاصل شده روی این الگوریتم نسبت به نمونه‌های قبلی، به دلیل در نظر گرفتن ساختار چندلایه‌ای و به هم متصل بودن شبکه است که چون گره‌های متفاوت از نظر نوع، به صورت موازی در کنار گره‌های هم‌نوع دیده می‌شوند، سربار زمانی به الگوریتم اضافه نمی‌کند و به طور موازی مورد محاسبه قرار می‌گیرد و بنابراین بر پیچیدگی زمانی افزایش نخواهد داشت.

۴-۴ معیار ارزیابی

برای مجموعه داده‌ی معرفی شده، یک مجموعه‌ی آموزش و یک مجموعه‌ی آزمایش داریم که با استفاده از این‌ها می‌توان انتشار موضوع T_c در مرحله زمانی $t+1$ را پیش‌بینی کرد.

در واقع، احتمال فعال شدن هر نویسنده در مرحله زمانی بعدی به دست می‌آید و اگر این احتمال از یک مقدار آستانه که توسط الگوریتم‌های یادگیری، به دست آمده بیشتر باشد، آن نویسنده را در آن زمان فعال اعلام می‌کنیم و این نویسنده توانایی فعال‌سازی نویسنده‌های همسایه‌ی خود در مرحله زمانی بعدی را دارا می‌باشد. سپس تمام نویسنده‌هایی که در هر مرحله، احتمال فعال شدنشان بیشتر از مقدار آستانه بوده را فعال در نظر گرفته و در پایان بازه‌ی زمانی، تعداد آن‌هایی که به درستی فعال شناخته شده‌اند و کسانی که فعال بوده‌اند، اما الگوریتم ما آن‌ها را فعال به حساب نیاورده است، را

مفروض مقاله به چاپ رسانده اند، ذخیره شده است. سپس تعداد افرادی که پیش بینی کرده ایم که در مورد همان موضوع در این سال ها مقاله به چاپ رسانده اند نیز، ذخیره شده است. خط ترسیم شده در شکل ۴ نیز برای پیش بینی با روش مقاله ی [۹] یعنی MLTM-R، رسم شده است. مشاهده می شود که پیش بینی های روش پیشنهادی ما یعنی (iLTM)، تفاوت اندکی با واقعیت دارد و نحوه ی پخش اطلاعات را نسبت به روش پیشین (MLTM-R)، به واقعیت شبیه تر به دست آورده است. چرا که پیش بینی تعداد افرادی که به این موضوع آلوده شده اند، به آن چه در واقعیت رخ داده است، شبیه تر است. همچنین انتشار این موضوعات در مجموعه داده ی ACM نیز مورد بررسی قرار گرفته است که نتایج آن در جدول ۳ برای موضوع Data

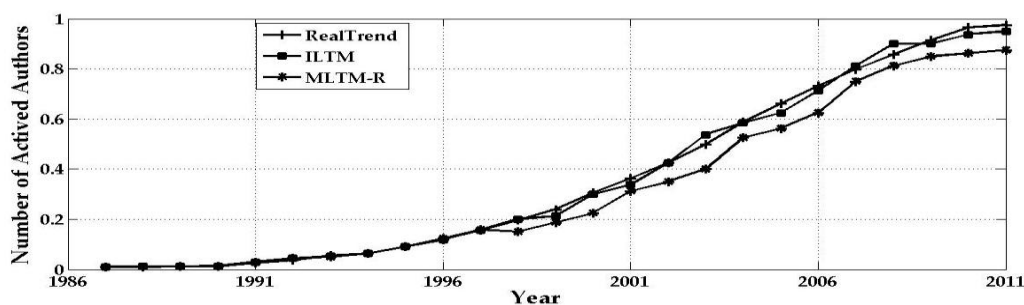
موضوع	روش	دقت	فراخوانی	امتیاز F1
Data Mining	[۹]MLTM-R	۰/۷۴	۰/۵۴	۰/۶۳
	iLTM	۰/۷۹	۰/۷۱	۰/۷۵
Machine Learning	[۹]MLTM-R	۰/۶۱	۰/۵۶	۰/۵۸
	iLTM	۰/۷۸	۰/۷۵	۰/۷۷
SVM	[۹]MLTM-R	۰/۶۰	۰/۵۴	۰/۵۶
	iLTM	۰/۷۹	۰/۶۱	۰/۶۹
Neural Network	[۹]MLTM-R	۰/۵۹	۰/۴۷	۰/۴۹
	iLTM	۰/۸۰	۰/۶۷	۰/۷۳
Naïve Bayes	[۹]MLTM-R	۰/۶۶	۰/۴۲	۰/۵۲
	iLTM	۰/۸۴	۰/۶۱	۰/۷۱

Mining موجود است و مشاهده می شود که به طور میانگین امتیاز F1 حدود ۱۶ درصد دارای بهبود بوده است. در نمودار شکل ۵، نمودار تجمعی تعداد افرادی که در چندین سال مختلف از سال ۱۹۸۶ تا ۲۰۱۱، راجع به یک موضوع مقاله به چاپ رسانده اند، نمایان گر نحوه ی رشد آن موضوع (برای مثال موضوع داده کاوی (Data Mining)) در شبکه ی کتابشناسی ACM می باشد.

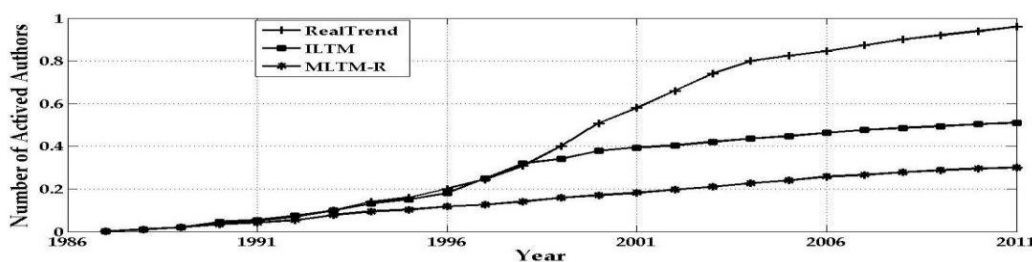
مجموعه ی موضوع جامع Machine Learning هستند و می توان برای نتیجه گیری به آن ها استناد کرد. جدول ۲، نشان دهنده ی معیارهای ارزیابی پیش بینی انتشار چندین موضوع در مجموعه داده ی DBLP می باشد. همان طور که قابل مشاهده است، در تمامی موضوعات، روش پیشنهادی این مقاله، دارای امتیاز F1 بهتری نسبت به روش های پیشین [۹] که معماری کامل شبکه را در نظر نمی گرفتند، می باشد. در واقع امتیاز F1، به عنوان میانگین دقت و فراخوانی، در مدل پیشنهادی نسبت به روش پیشین [۹]، دارای افزایش میانگین ۱۴ درصدی بوده است.

جدول ۲- معیارهای ارزیابی پیش بینی انتشار موضوعات مختلف در شبکه DBLP

همچنین نتایج این جدول نشان می دهند که هر چه یک موضوع، جامع تر باشد و ابعاد بیشتری را شامل بشود، نتایج به واقعیت شباهت بیشتری دارند. در حقیقت هر چه داده ی مورد نظر کامل تر باشد، نتایج بهتری به دست می آید. در جدول ۲، موضوع Machine Learning موضوعی جامع است که خود شامل موضوعات SVM، Neural Network، Naive Bayes و غیره می شود. همان طور که مشاهده می شود، میزان شباهت پیش بینی رشد این موضوع به واقعیت بیشتر از میزان شباهت موضوعات زیرمجموعه اش نسبت به واقعیت است، چرا که میزان دقت و فراخوانی و امتیاز F1 آن بیشتر از تمامی زیرمجموعه هایش شده است. در شکل ۴، نمودار تجمعی تعداد افرادی که در سال های مختلف، راجع به یک موضوع مقاله منتشر کرده اند، نمایان گر نحوه ی رشد آن موضوع (برای مثال موضوع داده کاوی (Data Mining)) در شبکه کتابشناسی DBLP می باشد. یعنی هر سال تعداد افرادی که در مورد موضوع



شکل ۴- مقایسه روند رشد موضوع داده کاوی بر روی مجموعه داده DBLP در واقعیت و پیش بینی های انجام شده



شکل ۵- مقایسه روند رشد موضوع داده کاوی بر روی مجموعه داده ACM در واقعیت و پیش بینی های انجام شده

فراخوانی و در نتیجه امتیاز F1 بالاتری خواهیم داشت نسبت به حالتی که فرامسیر نامتقارن به تنهایی مورد استفاده قرار گیرد. اما وقتی فرامسیر دیگری نیز به این یک فرامسیر اضافه می‌شود، چه متقارن باشد و چه نامتقارن، میزان دقت و فراخوانی آن و در نتیجه امتیاز F1 بیشتر می‌شود، از طرفی همان‌طور که در سطرهای ۶ الی ۱۱ جدول مشاهده می‌شود، بسته به این‌که این فرامسیر اضافه شده، چه نوع فرامسیری است، میزان این افزایش متفاوت می‌باشد. با توجه به اینکه فرامسیرهای نامتقارن، بخشی از اطلاعات پنهان را به همراه دارند، بنابراین وقتی آن‌ها را در کنار فرامسیرهای متقارن در فرآیند انتشار دخیل می‌کنیم، نتایج بهتری به دست می‌آوریم و این مساله در سطرهای ۸-۱۱ قابل مشاهده است. همچنین اگر تعداد فرامسیرها را به ۳ افزایش دهیم، میزان دقت و فراخوانی بالاتر رفته و نتایج قابل اعتمادتری به دست می‌آید. همان‌طور که در جدول قابل مشاهده است، هنگامی‌که تأثیر فرامسیرهای نامتقارن در کنار فرامسیرهای متقارن دیده می‌شود، نتایج بهتری رویت می‌شود، چرا که میزان دقت و فراخوانی و بنابراین امتیاز F1 این روش، از تمامی حالات پیشین بیشتر شده است.

جدول ۴- تأثیر فرامسیرهای مختلف در پیش‌بینی انتشار موضوعات مختلف بر روی مجموعه داده DBLP

NB	NN	SVM	ML	DM	
۰/۳۵	۰/۳۲	۰/۳۹	۰/۴۵	۰/۳۹	APA
۰/۳۷	۰/۳۵	۰/۴۴	۰/۳۶	۰/۵۳	AVA
۰/۲۷	۰/۲۹	۰/۳۳	۰/۳۱	۰/۴۲	APV
۰/۴۵	۰/۴۷	۰/۵۳	۰/۵۷	۰/۶۱	APAPV
۰/۵۲	۰/۴۹	۰/۵۶	۰/۵۸	۰/۶۳	APA-AVA[۹]
۰/۳۷	۰/۳۹	۰/۴۲	۰/۵۶	۰/۶۱	APV-APAPV
۰/۶۱	۰/۵۸	۰/۶۱	۰/۵۹	۰/۶۹	APA-APV
۰/۶۳	۰/۵۹	۰/۶۷	۰/۶۱	۰/۶۸	APA-APAPV
۰/۵۷	۰/۵۳	۰/۶۲	۰/۶۲	۰/۶۷	AVA-APV
۰/۵۹	۰/۵۵	۰/۶۵	۰/۶۹	۰/۶۵	AVA-APAPV
۰/۶۸	۰/۶۲	۰/۶۹	۰/۷۱	۰/۷۳	APA-AVA-APV
۰/۶۷	۰/۶۷	۰/۷۰	۰/۷۲	۰/۷۲	APA-AVA-APVPA
۰/۶۹	۰/۶۸	۰/۷۲	۰/۷۳	۰/۷۱	APV-APVPA-APA
۰/۶۵	۰/۶۴	۰/۷۱	۰/۷۲	۰/۷۰	APV-APVPA-AVA
۰/۷۱	۰/۷۳	۰/۶۹	۰/۷۷	۰/۷۵	APV-APVPA-APA-AVA

جدول ۵- تأثیر فرامسیرهای مختلف در پیش‌بینی انتشار موضوعات مختلف بر روی مجموعه داده ACM

همان‌طور که مشاهده می‌شود، پیش‌بینی‌های روش پیشنهادی ما یعنی (iLTM)، نسبت به روش (MLTM-R)، به واقعیت شبیه‌تر است و تطابق بیشتری با واقعیت دارد، چرا که روند رشد نمودار به واقعیت نزدیکی بیشتری دارد و یعنی پیش‌بینی تعداد افرادی که به این موضوع آلوده شده‌اند، به آن‌چه در واقعیت رخ داده است، شبیه‌تر است.

جدول ۳- معیارهای ارزیابی پیش‌بینی انتشار موضوعات مختلف در شبکه ACM

موضوع	روش	دقت	فراخوانی	امتیاز F1
Data Mining	iLTM	۰/۸۹	۰/۵۹	۰/۷۱
	[۹]MLTM-R	۰/۶۸	۰/۴۳	۰/۵۲
Machine Learning	iLTM	۰/۸۶	۰/۷۲	۰/۷۹
	[۹]MLTM-R	۰/۷۴	۰/۶۰	۰/۶۹
SVM	iLTM	۰/۷۹	۰/۶۱	۰/۶۹
	[۹]MLTM-R	۰/۷۴	۰/۴۶	۰/۵۷
Neural Network	iLTM	۰/۹۱	۰/۴۷	۰/۶۲
	[۹]MLTM-R	۰/۹۴	۰/۳۳	۰/۵۰
Naïve Bayes	iLTM	۰/۸۲	۰/۷۵	۰/۷۸
	[۹]MLTM-R	۰/۵۶	۰/۴۵	۰/۵۰

۴-۵ تأثیر فرامسیرهای مختلف و ترکیب‌های آن‌ها

برای نشان دادن کارایی روش پیشنهادی از منظری دیگر، می‌توان تأثیر فرامسیرهای مختلف و انواع مختلف ترکیب‌های آن‌ها را در پیش‌بینی انتشار به صورت جداگانه در نظر گرفت و نتایج را سنجید. برای این کار، ابتدا هر فرامسیر را به صورت تنها در نظر گرفته‌ایم و احتمال‌های فعال‌سازی گره‌ها را به دست آورده‌ایم و در نتیجه انتشار یک موضوع را با در نظر گرفتن تنها یک فرامسیر یا در حقیقت یک لایه‌ی منفرد (حالت تک‌لایه)، پیش‌بینی کرده‌ایم و امتیاز F1 را برای آن محاسبه می‌کنیم. این کار را برای هر یک از ۲ فرامسیر متقارن و ۲ فرامسیر نامتقارن انجام داده و سپس این مراحل را برای تمام ترکیب‌های ۲ تایی فرامسیرها تکرار می‌کنیم. به همین ترتیب مراحل برای ترکیب‌های ۳ تایی تکرار می‌شود و سپس تمام ۴ فرامسیر متقارن و نامتقارن را در کنار یکدیگر قرار داده و امتیاز F1 را محاسبه می‌کنیم. همان‌طور که در جدول ۴ مشاهده می‌شود، اگر از روش iLTM استفاده کنیم، بیشترین امتیاز F1 را داریم و این بدان معنی است که بیشترین شباهت به واقعیت را در بین تمام ترکیب‌های دیگر به دست آورده‌ایم. همچنین از نتایج جدول ۴ استنباط می‌شود که اگر یک فرامسیر نامتقارن به رابطه اضافه شود، به دلیل داشتن اطلاعات پنهان بین گره‌ها، نتیجه‌ی بهتری به دست می‌آید. یعنی روابط نامتقارن، هرچند به تنهایی از ارتباطات متقارن در پیش‌بینی انتشار ضعیف‌تر عمل کنند، اما وقتی در کنار ارتباطات متقارن قرار می‌گیرند، باعث بهبود پیش‌بینی می‌شوند.

همچنین بر روی مجموعه‌ی داده‌ی ACM همان‌طور که در جدول ۵ مشاهده می‌شود، وقتی تنها یک فرامسیر متقارن را مدنظر قرار دهیم -که در این‌جا فرامسیر نویسنده-مقاله-نویسنده مدنظر است- دقت و

برای گسترش این موضوع می توان مجموعه دادگانی در حوزه های دیگر، مانند شبکه های افراد و طیور را در نظر گرفت و در این شبکه مراحل انتشار بیماری را مدل سازی و روند گسترش بیماری را مشاهده کرد و برای جلوگیری از انتشار آن، تمهیداتی اندیشید. مثلاً می توان خارج کردن گره با بیشترین ارتباط از شبکه و یا ایمن کردن چنین گره ای، یا حذف تعدادی یال با توجه به قدرت انتشاری هر لایه، برای جلوگیری از بیشتر آلوده شدن گره ها، را نام برد. همچنین می توان به راه هایی برای تغییر مسیر انتشار و سازماندهی حرکت فرآیند انتشاری در چنین شبکه ای اندیشید و از پخش وقایع بد جلوگیری کرد و یا به پخش اطلاعات مناسب، کمک کرد و هدف گذاری برای انتشار ایجاد کرد، بدین صورت که با قصد قبلی، گره های اولیه را طوری در شبکه ای چندلایه انتخاب کرد که در کوتاهترین زمان ممکن، انتشار تمامی گره ها و یا حداقل کسر مناسبی از آن ها را درگیر کند.

NB	NN	SVM	ML	DM	
۰/۶۳	۰/۳۷	۰/۵۲	۰/۶۶	۰/۴۶	APA
۰/۵۳	۰/۳۲	۰/۴۳	۰/۵۸	۰/۳۸	AVA
۰/۴۹	۰/۱۹	۰/۳۳	۰/۵۵	۰/۲۶	APV
۰/۴۲	۰/۲۲	۰/۳۴	۰/۵۶	۰/۳۵	APAPV
۰/۵۰	۰/۵۰	۰/۵۷	۰/۶۹	۰/۵۲	APA-AVA[۹]
۰/۵۵	۰/۳۲	۰/۵۲	۰/۵۸	۰/۳۷	APV-APAPV
۰/۶۹	۰/۵۷	۰/۶۳	۰/۷۰	۰/۵۳	APA-APV
۰/۷۳	۰/۵۶	۰/۶۵	۰/۷۴	۰/۵۷	APA-APAPV
۰/۷۰	۰/۵۶	۰/۵۸	۰/۷۲	۰/۵۹	AVA-APV
۰/۷۴	۰/۵۷	۰/۵۹	۰/۷۶	۰/۵۶	AVA-APAPV
۰/۷۵	۰/۵۸	۰/۶۶	۰/۷۸	۰/۶۰	APA-AVA-APV
۰/۷۴	۰/۵۸	۰/۶۷	۰/۸۰	۰/۶۴	APA-AVA-APVPA
۰/۷۵	۰/۵۹	۰/۶۹	۰/۸۵	۰/۶۵	APV-APVPA-APA
۰/۷۶	۰/۶۰	۰/۶۷	۰/۸۷	۰/۶۸	APV-APVPA-AVA
۰/۷۸	۰/۶۲	۰/۶۹	۰/۷۹	۰/۷۱	APV-APVPA-APA-AVA

مراجع

- [1] C. Kirst, M. Timme, and D. Battaglia, "Dynamic information routing in complex networks," *Nat. Commun.*, vol. 7, p. 11061, 2016.
- [2] Z.-K. Zhang, C. Liu, X.-X. Zhan, X. Lu, C.-X. Zhang, and Y.-C. Zhang, "Dynamics of information diffusion and its applications on complex networks," *Phys. Rep.*, vol. 651, pp. 1-34, 2016.
- [3] S. Dhamal, K. J. Prabuchandran, and Y. Narahari, "Information Diffusion in Social Networks in Two Phases," *IEEE Trans. Netw. Sci. Eng.*, vol. 3, no. 4, pp. 197-210, Oct. 2016.
- [4] D. Margaris, C. Vassilakis, and P. Georgiadis, "Recommendation information diffusion in social networks considering user influence and semantics," *Soc. Netw. Anal. Min.*, vol. 6, no. 1, pp. 122-136, 2016.
- [5] Y. Hu, R. J. Song, and M. Chen, "Modeling for Information Diffusion in Online Social Networks via Hydrodynamics," *IEEE Access*, vol. 5, pp. 128-135, 2017.
- [6] J. Han, "Mining heterogeneous information networks by exploring the power of links," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 5808 LNAI, pp. 13-30, 2009.
- [7] Y. Sun, J. Han, P. Zhao, Z. Yin, H. Cheng, and T. Wu, "RankClus: Integrating Clustering with Ranking for Heterogeneous Information Network Analysis," *Proc. 12th Int. Conf. Extending Database Technol. Adv. Database Technol. - EDBT '09*, p. 565, 2009.
- [8] Y. Sun and J. Han, "Mining heterogeneous information networks: a structural analysis approach," *ACM SIGKDD Explor. Newsl.*, vol. 14, no. 2, pp. 20-28, 2013.
- [9] H. Gui, Y. Sun, J. Han, and G. Brova, "Modeling Topic Diffusion in Multi-Relational Bibliographic Information Networks," in *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management - CIKM '14*, pp. 649-658, 2014.
- [10] M. S. Granovetter, "Threshold Models of Collective Behavior," *Am. J. Sociol.*, vol. 83, no. 6, pp. 1420-1443, 1978.
- [11] N. Pathak, A. Banerjee, and J. Srivastava, "A Generalized Linear Threshold Model for Multiple Cascades," in *2010 IEEE International Conference on Data Mining*, pp. 965-970, 2010.
- [12] D. J. Watts, "A Simple Model of Global Cascades on Random Networks," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 99, no. 9, pp. 5766-5771, 2002.
- [13] Y. Sun and J. Han, "Mining Heterogeneous Information Networks: Principles and Methodologies," *Synth. Lect. Data Min. Knowl. Discov.*, vol. 3, no. 2, pp. 1-159, 2012.
- [14] C. D. Brummitt, K. M. Lee, and K. I. Goh, "Multiplexity-facilitated cascades in networks," *Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys.*, vol. 85, no. 4, 2012.

۶ نتیجه گیری و پیشنهادها

در بخش اول مفاهیم شبکه های چندلایه و انتشار را توضیح دادیم. در بخش دوم، مدل های انتشار موجود بررسی شدند. در بخش سوم، روشی نوین برای پیش بینی انتشار اطلاعات در شبکه های چندلایه ارائه شد. در بخش چهارم، مدلمان را ارزیابی کردیم و در بخش پنجم، نشان دادیم که الگوریتم پیشنهادی، کارایی مناسبی در برابر راهکارهای پیشین [۹] دارد و می تواند نتایج قابل اعتمادتری به دست آورد. در این مقاله، توسعه ای از مدل آستانه ای خطی (LTM) ارائه شد و آن را به شبکه های چندلایه تعمیم دادیم. این مدل به شبکه ای اطلاعاتی کتابشناسی اعمال شد. مقایسه ای روش پیشنهادی و روش پیشین [۹]، نشان داد که نتایج روش پیشنهادی شباهت بیشتری به واقعیت دارد، چرا که نمودارهای نشان دهنده ی روند رشد تعداد افرادی که به یک موضوع آلوده شده اند،

بیان گر این مطلب هستند که روش پیشنهادی ما نسبت به روش پیشین [۹]، فاصله کمتری از واقعیت دارد. در حقیقت در هر دو مجموعه داده ای آزمون مورد استفاده، میزان شباهت روند رشد موضوع در شبکه، بین روش پیشنهادی این مقاله و واقعیت بیشتر از میزان شباهت روش پیشین با واقعیت بوده است. بنابراین می توان نتیجه گرفت که در شبکه ای کتابشناسی که مدلمان را به آن اعمال کردیم، نویسندگان مقالات در انتخاب موضوع مقاله ی خود، نه تنها از دیگر نویسندگان تأثیر می پذیرند، بلکه کنفرانس ها نیز می توانند بر تصمیم گیری آن ها برای انتخاب موضوع مقاله، اثرگذار باشند. علاوه بر این، میزان شباهت نتایج دقت و فراخوانی روش پیشنهادی نسبت به واقعیت در مقایسه با روش پیشین، در یکی از مجموعه دادگان به طور میانگین حدود ۱۴ درصد و در دیگری ۱۶ درصد افزایش داشته که این امر نشان دهنده ی کارایی بیشتر روش پیشنهادی ما، نسبت به روش پیشین است.

- [22] A. Chao, R. L. Chazdon, R. K. Colwell, and T.-J. Shen, "A new statistical approach for assessing similarity of species composition with incidence and abundance data," *Ecol. Lett.*, vol. 8, no. 2, pp. 148–159, Dec. 2004.
- [۲۳] سمیه توکلی، افسانه فاطمی. "تشکیل تیم دوهفته در شبکه‌های اجتماعی." *مجله مهندسی برق دانشگاه تبریز*، جلد ۴۷، شماره ۲، تابستان ۱۳۹۶.
- [24] Y. Sun, J. Han, X. Yan, P. S. Yu, and T. Wu, "PathSim: Meta Path-Based Top-K Similarity Search in Heterogeneous Information Networks," *VLDB 2011*, vol. 3, no. 2, pp. 1–12, 2011.
- [۲۵] مریم باسره، ولی درهمی، سجاد ظریف زاده. "ارائه روشی برای استخراج خودکار عبارات کلیدی از اخبار وب پارسی." *مجله مهندسی برق دانشگاه تبریز*، جلد ۴۷، شماره ۳، پاییز ۱۳۹۶.
- [26] D. M. W. POWERS, "Evaluation: From Precision, Recall and F-Measure To Roc, Informedness, Markedness & Correlation," *J. Mach. Learn. Technol.*, vol. 2, no. 1, pp. 37–63, 2011.
- [27] P. Perruchet and R. Peereman, "The exploitation of distributional information in syllable processing," *J. Neurolinguistics*, vol. 17, no. 2–3, pp. 97–119, Mar. 2004.
- [15] C. Budak, D. Agrawal, and A. El Abbadi, "Limiting the spread of misinformation in social networks," in *Proceedings of the 20th international conference on World wide web - WWW '11*, p. 665, 2011.
- [16] Y. Zhuang, A. Arenas, and O. Yağan, "Clustering determines the dynamics of complex contagions in multiplex networks," *Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys.*, vol. 95, no. 1, 2017.
- [17] M. Salehi, R. Sharma, M. Marzolla, M. Magnani, P. Siyari, and D. Montesi, "Spreading Processes in Multilayer Networks," *IEEE Trans. Netw. Sci. Eng.*, vol. 2, no. 2, pp. 65–83, 2015.
- [18] D. Kempe, J. Kleinberg, and É. Tardos, "Influential Nodes in a Diffusion Model for Social Networks," pp. 1127–1138, 2005.
- [19] M. Kivelä, A. Arenas, M. Barthelemy, J. P. Gleeson, Y. Moreno, and M. A. Porter, "Multilayer networks," *J. Complex Networks*, vol. 2, no. 3, pp. 203–271, 2014.
- [۲۰] سهیلا مولائی، سما بابایی، مصطفی صالحی، "جستجوی شباهت در شبکه‌های ناهمگن بر مبنای فرامسیرهای وزن‌دار." *بیست و یکمین کنفرانس انجمن کامپیوتر ایران*، تهران، اسفند ۱۸–۲۰، ۱۳۹۴.
- [21] J. Santisteban, Tejada-Cárcamo, Unilateral Jaccard Similarity Coefficient, *GSB@ SIGIR*. (2015) 23–27. doi:http://ceur-ws.org/Vol-1393/.

زیرنویس‌ها

¹ Venue-Author-Venue

² <http://aminer.org/lab-datasets/citation/DBLP-citation-Feb21.tar.bz2>

³ <http://static.aminer.org/lab-datasets/citation/citation-acm-v8.txt.tgz>