

## مدلسازی انتشار اطلاعات در شبکه‌های کتابشناسی چندلایه

سما بابایی<sup>۱</sup>، کارشناس ارشد، سهیلا مولایی<sup>۲</sup>، کارشناس ارشد، مصطفی صالحی<sup>۳</sup>، 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 سیستم‌های اطلاعاتی را می‌توان به شکل شبکه‌ای چندلایه مدل کرد که هر لایه متشکل از تعدادی گره است و این گره‌ها با انواع مختلفی یال به یکدیگر مرتبط می‌شوند. اگرچه محققان در سال‌های اخیر به تحلیل انتشار اطلاعات در شبکه توجه نشان داده‌اند، اکثر این مطالعات به شبکه‌های تک‌لایه محدود بوده است. اما در دنیای واقعی به دلیل پیچیدگی روابط بین گره‌ها، عموماً شاهد وجود شبکه‌های چندلایه هستیم. کارهای پیشین، اغلب با ساده‌سازی‌های زیاد فضای مساله همراه هستند، مثلاً در بیشتر کارها تنوع گره‌ها و تاثیر متقابل آن‌ها نادیده گرفته شده است. روش پیشنهادی این مقاله با در نظر گرفتن تاثیر لایه‌های مختلف بر یکدیگر به پیش‌بینی انتشار در شبکه‌های چندلایه، می‌پردازد. مهم‌ترین ویژگی روش، این است که می‌تواند قدرت تاثیر تمامی لایه‌ها را مشخص کند و میزان این تاثیر را بدون توجه به شباهت یا تفاوت گره‌های هر لایه نسبت به هم، اندازه‌گیری نماید. مدل پیشنهادی بر روی دو مجموعه داده‌ی واقعی پیاده‌سازی و با سناریوهای مختلفی ارزیابی شده است. همچنین روند فعال شدن گره‌های مختلف را به دست آورده و با روند رشد واقعه‌شان در واقعیت مقایسه کردیم و نشان دادیم که روش پیشنهادی نسبت به روش‌های پیشین که تمامی گره‌ها را از یک نوع فرض می‌کردند، شباهت بیشتری به واقعیت دارد و می‌تواند راه مناسبی برای پیش‌بینی انتشار در شبکه‌های چندلایه باشد.

**واژه‌های کلیدی:** شبکه‌ی چندلایه، انتشار اطلاعات، شبکه‌ی چندگانه، شبکه‌ی به‌هم‌متصل، فرآیندهای انتشار

## Modeling Information Diffusion in Bibliographic Multilayer Networks

Sama Babaei, MSc<sup>1</sup>, Soheila Molaei, MSc<sup>2</sup>, Mostafa Salehi, PhD<sup>3</sup>

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

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

3- Faculty of New Sciences and Technologies, University of Tehran, Tehran, Iran, Email: [mostafa\\_salehi@ut.ac.ir](mailto: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:

تاریخ اصلاح مقاله:

تاریخ پذیرش مقاله:

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

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 روبه‌رو 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 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 آستانه‌ی 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  $\theta$  از همسایه‌هایش 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  $\theta_u$  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  $\theta_u$  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 نحوه‌ی تأثیر آن‌ها بر یکدیگر را اندازه‌گیری نماید. برای بررسی انتشار در این چارچوب، مفهومی به نام فرامسیر این‌طور 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  $\vec{R}_1 A_{i+1} \vec{R}_2 \dots \vec{R}_1 A_1$  نمایش داده می‌شود که یک رابطه ترکیب  $R = R_1 \circ R_2 \circ \dots \circ R_1$  بین انواع  $A_{i+1}$  و  $A_1$  را نشان می‌دهد و  $\circ$  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) عنوان می‌شود. به هر نمونه از یک فرامسیر، برای مثال شرکت کردن نویسنده باب در کنفرانس VLDB، به عنوان یک نمونه مسیر در 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 فرامسیره‌ها 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 آرایه شده در این مقاله عبارتند از:

- آرایه‌ی 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 DBLP ACM و

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 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، تاثیر 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 توعشان 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 بطور 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 مسیره‌ها مدل‌سازی 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 تاثیر 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

به دست آوردن ضرایب  $w$  بین گره  $i$  و همسایه‌هایش

Serial no. XX

روابط مقارن است؟

خیر

$$w_{ij} = \frac{|P_{(u,v)}^E|}{P_i^E + P_j^E - P_{ij}^E}$$

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 تاثیرات متقابلی که هر لایه بر دیگر لایه‌ها دارد، نادیده گرفته شود. همچنین تاثیر 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 تاثیرگذار باشند. هم‌چنین این تاثیرگذاری تحت روابط مختلف، شدت‌های مختلفی نیز دارد که لازم است در مدل‌سازی مورد توجه قرار بگیرند. بنابراین برای اینکه مدل‌سازی 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 لایه‌ای 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 (بال‌های درون لایه‌ای) منتقل شوند و هم می‌توانند از طریق ارتباط بین گره‌های ناهمگون (بال‌های بین لایه‌ای) انتشار یابند [۱۹]. برای هر نوع یال وزنی خاص در نظر گرفته می‌شود. در حقیقت هر فرامسیری می‌تواند باعث تاثیرگذاری در رسیدن فرآیند انتشاری به گره‌ای خاص بشود، یعنی هر دو گره‌ی متصل، تحت نوع رابطه‌ای که بین آن‌ها برقرار است، می‌توانند تاثیرات مختلفی بر

شکل ۲: روند نما پیش‌بینی و مدل‌سازی انتشار اطلاعات

هستیم. در واقع هر فرامسیر می تواند باعث اثرگذاری در قبول فرآیند انتشاری توسط گره های خاص بشود، یعنی هر دو گره ای که با یکدیگر در تماس هستند، تحت نوع رابطه ای که بین آن ها برقرار است، می توانند تاثیرات مختلفی از هم بگیرند. بنابراین باید تمام روابط در کنار هم دیده شوند و قدرت اثرگذاری همزمان تمامی آن ها با هم در نظر گرفته شود. حال LTM را که مدل مناسبی برای بررسی انتشار در شبکه های تک لایه است، به شبکه های چندلایه بسط می دهیم [۱۱]. یک شبکه ی چندلایه را مانند یک شبکه ی تک لایه با وزن های مختلف برای هر یال از هر لایه، در نظر می گیریم و یک مدل آستانه ی خطی ساده را برای به دست آوردن احتمال فعال شدن گره ها، به آن اعمال می کنیم. توجه داشته باشید که چنین فرضی برای شبکه های چندلایه، به منظور تک لایه در نظر گرفتن آن نیست، چرا که هنگامی که یک شبکه ی چندلایه به عنوان شبکه ی تک لایه در نظر گرفته می شود، تمام یال های موجود در شبکه ی چندلایه به یک یال تصویر می شوند و اثر متفاوتی که هر یال می تواند داشته باشد نادیده گرفته می شود. فرض می کنیم که درصدهای وزنی همسایه های فعال یک گره، احتمال فعال شدن گره را معین می کند و آستانه ی فعال شدن هر گره، دارای توزیع یکنواخت در بازه ی [۰،۱] است و وزن ارتباط به معنی وزن هر یال است که در ادامه این فرض ها را توضیح می دهیم.

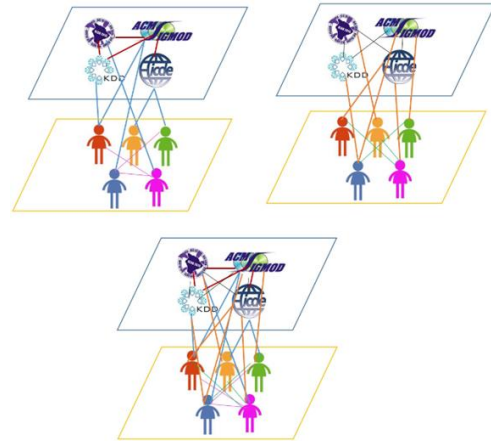
### ۳-۱- وزن ارتباطها

قدرت انتشار (وزن تاثیر) هر جفت گره از لایه های مختلف باید تخمین زده شود. به طور خاص از رابطه ی شباهت Pathsim [۲۴] برای محاسبه قدرت انتشار بین جفت گره های متقارن استفاده شده است. شباهت بین هر دو گره می تواند نشان دهنده ی قدرت تاثیرگذاری این دو گره بر یکدیگر باشد، چرا که هر چه دو گره بیشتر به یکدیگر شبیه باشند، توانایی تاثیرگذاری بیشتری دارند و هر چه شباهت آن ها به هم کمتر باشد، تاثیر کمتری بر یکدیگر می گذارند. در مقاله [۲۰] روشی مبتنی بر Pathsim ارائه شده است که شباهت نوع بعضی از گره ها به یکدیگر را نیز در بررسی شباهت گره هایی با نوع دیگر دخیل می کند، اما این مساله در بررسی انتشار به کار نمی آید، چرا که در بررسی انتشار، قدرت تعامل دو گره مهم است. اگر لایه ی  $E_k$  را در نظر بگیریم (فرامسیر متقارن)، وزن بین گره ی  $u$  و  $v$  به این صورت تعریف می شود [۲۴]:

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

$P_{(u,v)}^{E_k}$  مجموعه ای از فرامسیرهای متقارن (تعاملات متقارن) است که مربوط به لایه ی  $E_k$  است که از گره ی  $u$  شروع می شود و به گره ی  $v$  ختم می شود و همچنین  $| \cdot |$  نشان دهنده ی اندازه ی مجموعه است. برای محاسبه ی قدرت انتشار بین جفت گره های نامتقارن از رابطه ی شباهت جاکارد [۲۱] استفاده شده است. اگر لایه ی  $E_k$  را در نظر بگیریم (فرامسیر نامتقارن)، وزن بین گره ی  $u$  و  $v$  به این صورت تعریف می شود [۲۲]:

یکدیگر بگذارند. برای مثال در پخش شدن فرآیند چاقی، اگر فرد  $A$  با فرد  $B$  دوست صمیمی باشد، تاثیر بیشتری در چاقی فرد  $B$  دارد نسبت به زمانی که این دو، با هم همکار باشند. بنابراین، برای نشان دادن تاثیرات مختلف گره ها بر هم، لازم است هر فرامسیری، وزن اختصاصی داشته باشد. در مثال فوق، وزن رابطه ی دوستی، بیشتر از وزن رابطه ی همکار بودن است. اما اگر دو فرد هم با یکدیگر دوست صمیمی باشند و هم همکار، بالطبع تاثیری که فرد  $A$  در چاق شدن فرد  $B$  می گذارد، 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 تاثیر همزمان تمامی آن ها در نظر گرفته شود. گره ها از انواع متفاوتی هستند و روابط مختلفی بین آن ها وجود دارد. برای بررسی انتشار در چنین شبکه ای، تمام روابطی که بین هر گره با گره دیگر وجود دارد، با وزنی به خصوص در نظر گرفته می شوند و تاثیر آن رابطه به صورت منحصر به فرد در کنار تاثیر بقیه روابط، در نظر گرفته می شود. در واقع هر فرامسیر می تواند باعث اثرگذاری در قبول فرآیند انتشاری توسط گره ای خاص بشود.



شکل ۳- انتشار با در نظر گرفتن تاثیر تمامی انواع گره ها و روابطشان با یکدیگر

یعنی هر دو گره ای که با یکدیگر در تماس هستند، تحت نوع رابطه ای که بین آن ها برقرار است، می توانند تاثیرات مختلفی از هم بگیرند. بنابراین باید تمام روابط در کنار هم دیده شوند و قدرت اثرگذاری همزمان تمامی آن ها با هم در نظر گرفته شود. مفهوم این فرض در شکل ۳ مشاهده می شود.

در واقع در هر کدام از دو قسمت بالایی شکل ۳، هر یال نشان دهنده ی یک فرامسیر است که بین گره ها برقرار می باشد. به منظور در نظر گرفتن تاثیر تمام گره ها 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 وزن دار بر روی این فرامسیرها صورت بگیرد و همه ی آن ها با وزن مخصوص به خودشان، در کنار یکدیگر دیده شوند و بر روی هم تاثیر بگذارند. در چنین شبکه ای که شامل تمامی فرامسیرهاست و در قسمت پایینی شکل ۳ قابل مشاهده است، به دنبال احتمال آلوده شدن هر گره به فرآیند انتشاری مدنظر

تاثیر بگذارند. صورت کسر اول از رابطه ی LTM برای یک ارتباط به دست آمده است که قدرت تاثیر آن ارتباط توسط ضریب  $\beta_k$  لحاظ شده است. این رابطه ی LTM برای تمامی ارتباطات متقارن به کار رفته است و مجموع تمامی آن ها میزان تاثیر پذیری گره ی  $i$  از تمام همسایه های همگون فعال آن را نشان می دهد. صورت کسر دوم نیز مانند صورت کسر اول از رابطه ی LTM برای یک ارتباط به دست آمده است، با این تفاوت که این رابطه ی LTM برای تمامی ارتباطات نامتقارن به کار رفته است. پس می بایست مجموع تمامی آن ها که میزان تاثیر پذیری گره ی  $i$  از تمام همسایه های ناهمگون فعال خود را نشان می دهد، به عنوان صورت کسر دوم لحاظ شود. صورت کسر سوم تاثیر تمام گره های شبکه (چه با گره ی  $i$  همسایه باشند و چه نباشند) را نشان می دهد. مخرج هر سه کسر نیز برابر با مجموع تاثیرات کلی تمامی همسایه های همگون و

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

که  $P_{(u,v)}^{E_k}$  مجموعه ای از فرامسیره های نامتقارن (تعاملات نامتقارن) است که مربوط به لایه ی  $E_k$  است که از گره ی  $u$  شروع می شود و به گره ی  $v$  ختم می شود.

### ۳-۲- احتمال فعال شدن هر گره

شبکه ی چندلایه را یک شبکه با وزن های متفاوت برای هر لایه در نظر می گیریم، یعنی همه یال ها را با هم جمع می کنیم و برای یال های لایه های مختلف، وزن های متفاوتی در نظر می گیریم. با این فرض، احتمال فعال شدن شی  $i$  در مرحله ی زمانی  $t+1$  به این صورت محاسبه می شود:

$$p_i(t+1) = \frac{\sum_{k \in SYM} \beta_k \sum_{j \in \Gamma_i(i,k)} w_{ij,t}^{(k)} \delta(j,t)}{\sum_{k \in SYM} \beta_k \sum_{j \in \Gamma_i(i,k)} w_{ij,t}^{(k)} + \sum_{k \in ASYM} \beta_k \sum_{j \in \Gamma_i(i,k)} w_{ij,t}^{(k)} + \beta_0 N} + \frac{\sum_{k \in ASYM} \beta_k \sum_{j \in \Gamma_i(i,k)} w_{ij,t}^{(k)} \delta(j,t)}{\sum_{k \in SYM} \beta_k \sum_{j \in \Gamma_i(i,k)} w_{ij,t}^{(k)} + \sum_{k \in ASYM} \beta_k \sum_{j \in \Gamma_i(i,k)} w_{ij,t}^{(k)} + \beta_0 N} + \frac{\beta_0 \sum_j \delta(j,t)}{\sum_{k \in SYM} \beta_k \sum_{j \in \Gamma_i(i,k)} w_{ij,t}^{(k)} + \sum_{k \in ASYM} \beta_k \sum_{j \in \Gamma_i(i,k)} w_{ij,t}^{(k)} + \beta_0 N} \quad (4)$$

که  $\Gamma_i(i,k)$  بیان گر مجموعه ی همسایه های گره ی  $i$  در لایه ی  $E_k$  در مرحله ی زمانی  $t$  است و  $w_{ij,t}^{(k)}$  نشان دهنده ی وزن یال بین گره ی  $i$  و  $j$  در لایه ی  $k$  است و  $\delta(j,t)$  بیان گر این است که گره ی  $j$  در مرحله ی زمانی  $t$  فعال شده یا خیر.  $N$  تعداد کل گره های موجود در شبکه است. همچنین تاثیر کلی (مانند تاثیری که رسانه ها در انتشار اطلاعات دارند و در درون شبکه قابل ردیابی نیست) را با اضافه کردن فاکتور ثابت با وزن  $\beta_0$  نشان داده ایم.  $\beta_k$  نشان دهنده ی وزن فراسطحی برای لایه ی  $k$  است.

رابطه ی ۴ را کمی بیشتر تشریح خواهیم کرد. برای به دست آوردن احتمال فعال شدن گره ی  $i$  در مرحله ی زمانی بعدی ( $t+1$ ) می بایست نسبت پیشامد رخ داده به فضای نمونه را محاسبه کنیم که در این جا پیشامد مورد نظر، تاثیر کلی همسایه های فعال شده ی گره ی  $i$  است (با در نظر گرفتن تاثیر لایه ای هر همسایه) و فضای نمونه تاثیر کلی تمامی همسایه های گره ی  $i$  (با در نظر گرفتن تاثیر لایه ای هر همسایه) است. این که رابطه ی ۴ از مجموع سه کسر تشکیل شده، به این دلیل است که کسر اول نمایان گر نسبت تاثیر کلی همسایه های همگون فعال شده ی گره ی  $i$  به تاثیر تمامی همسایه های همگون و ناهمگون گره ی  $i$  می باشد و کسر دوم نمایان گر نسبت تاثیر کلی همسایه های ناهمگون فعال شده ی گره ی  $i$  به تاثیر تمامی همسایه های همگون و ناهمگون گره ی  $i$  می باشد. کسر سوم نیز نشان دهنده ی تاثیرات اجتناب ناپذیر تمام گره های شبکه است که می تواند مثلاً از طریق رسانه ها بر گره ی  $i$

شبه کد ۱: الگوریتم iLTM
<p><b>Input: Datasets</b></p> <p><b>Outputs: Activation probability of each inactive node</b></p> <p>To calculate the activation probability of node <math>i</math>: Choose a Database (in this case: DBLP): <b>for</b> any symmetric metpath(APA,AVA,etc.) do : Find the proper weight for any metapath by iLTM learning Algorithm, and multiple with: <b>for</b> any neighbor of node <math>i</math> in metapath <math>k</math> do : Calculate <math>w_{ij}</math> between <math>i</math> and each activated neighbor <math>j</math> with a similarity measure (in this case: PathSim Similarity)</p> <p><b>Plus by:</b> <b>for</b> any asymmetric metpath(APV,VPAPA,etc.) do : Find the proper weight for any metapath by iLTM learning Algorithm, and multiple with: <b>for</b> any neighbor of node <math>i</math> in metapath <math>k'</math> do : Calculate <math>w_{ij}</math> between <math>i</math> and each activated neighbor <math>j</math> with a similarity measure (in this case: Jaccard Similarity)</p> <p><b>Plus by:</b> One for any activated neighbor of node <math>i</math>.</p> <p><b>Devide by:</b> <b>for</b> any symmetric metpath(APA,AVA,etc.) do : Find the proper weight for any metapath by iLTM learning Algorithm, and multiple with: <b>for</b> any neighbor of node <math>i</math> in metapath <math>k</math> do : Calculate <math>w_{ij}</math> between <math>i</math> and each neighbor <math>j</math> with a similarity measure (in this case: PathSim Similarity)</p> <p><b>Plus by:</b> <b>for</b> any asymmetric metpath(APV,VPAPA,etc.) do : Find the proper weight for any metapath by iLTM learning Algorithm, and multiple with: <b>for</b> any neighbor of node <math>i</math> in metapath <math>k'</math> do : Calculate <math>w_{ij}</math> between <math>i</math> and each neighbor <math>j</math> with a similarity measure (in this case: Jaccard Similarity)</p> <p><b>Plus by:</b> One for any neighbor of node <math>i</math>.</p>

ناهمگون گره ی  $i$  با در نظر گرفتن قدرت تاثیرهای مختلف نوع ارتباطات

## ۳-۴- الگوریتم یادگیری برای iTLM

برای به دست آوردن بهترین پارامتر، از فرآیند یادگیری پارامترها استفاده می کنیم تا به تخمین بیشینه احتمال برسیم. یعنی پارامتر  $\beta_k$  برای هر لایه  $E_k$  را مورد یادگیری قرار می دهیم تا به بهترین  $\beta_k$  ای برسیم که در صورت قرار گرفتن در رابطه ی فوق، احتمال فعال شدن گره را بیشینه کند. در مدل iTLM با قرار دادن معادله ی فوق در معادله ی قبلی به معادله ی زیر می رسیم:

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

باید  $\beta_k$  را تحت فرآیند یادگیری طوری به دست آوریم که نتیجه ی لگاریتم احتمال فوق بیشینه شود، برای این کار باید گام یادگیری ( $\eta$ ) هوشمندانه انتخاب شود. طبق روش گرادیان نزولی  $\beta_k$  از به روزرسانی پیاپی رابطه ی بازگشتی زیر به دست می آید:

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

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

$$\begin{aligned} \frac{\partial \log L}{\partial \beta_k} &= \sum_t \sum_{u \in V_t} \frac{\sum_v w_{u,v}^{(k)} \delta(v,t)}{\sum_{k'} \beta_{k'} \left( \sum_v w_{u,v}^{(k')} \delta(v,t) \right)} \\ &\quad + \sum_t \sum_{u \in R_t} \frac{\sum_v w_{u,v}^{(k)} (1 - \delta(v,t))}{\sum_{k'} \beta_{k'} \left( \sum_v w_{u,v}^{(k')} (1 - \delta(v,t)) \right)} \\ &\quad - \sum_t \sum_u \frac{\sum_v w_{u,v}^{(k)}}{\sum_{k'} \beta_{k'} \sum_v w_{u,v}^{(k')}} \end{aligned} \quad (9)$$

برای به دست آوردن  $\beta_k$  های مثبت در نظر می گیریم:

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

$$\sum_t \sum_u \frac{\beta_k^{(old)} \sum_v w_{u,v}^{(k)}}{\sum_{k'} \beta_{k'}^{(old)} \sum_v w_{u,v}^{(k')}} \quad (11)$$

و بدین ترتیب خواهیم داشت:

$$\begin{aligned} \beta_k^{(new)} &= \sum_t \left\{ \frac{\sum_v w_{u,v}^{(k)} \delta(v,t)}{\sum_{k'} \beta_{k'}^{(old)} \left( \sum_v w_{u,v}^{(k')} \delta(v,t) \right)} \right. \\ &\quad \left. + \frac{\sum_v w_{u,v}^{(k)} (1 - \delta(v,t))}{\sum_{k'} \beta_{k'}^{(old)} \left( \sum_v w_{u,v}^{(k')} (1 - \delta(v,t)) \right)} \right\} \beta_k^{(old)} \\ &\quad / \sum_t \sum_u \frac{\sum_v w_{u,v}^{(k)}}{\sum_{k'} \beta_{k'}^{(old)} \sum_v w_{u,v}^{(k')}} \end{aligned} \quad (11)$$

متفاوت و تاثیر کلی تمامی گره های شبکه می باشد. الگوریتم پیشنهادی در شبه کد ۱ قابل مشاهده است.

## ۳-۳- آموزش پارامترها

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

برای هر فرآیند انتشار، فهرست گره های درگیر در انتشار دنباله ای

از مجموعه گره های فعال و زمان فعال شدنشان است:  $A = \{V_t\}_{t=1}^T$  که  $t$  نشان دهنده مراحل زمانی گسسته است. برای به دست آوردن بهترین پارامترها، از فرآیند یادگیری پارامترها استفاده می کنیم. پارامتر  $\beta_k$  برای هر لایه  $E_k$  را یادگرفته تا به بهترین  $\beta_k$  ای برسیم که در صورت قرار گرفتن در رابطه ی زیر، احتمال فعال شدن گره را بیشینه کند. در هر مرحله ی زمانی  $t$ ، احتمال فعال شدن  $p_u(t)$  برای هر گره  $u$  فعال نشده ( $u \in R_t$ )، از توزیع برنولی تبعیت می کند و  $R_t$  نشان دهنده ی مجموعه گره هایی است که تا زمان  $t-1$  فعال نشده اند. در واقع در مرحله ی زمانی  $t$ ، هر گره  $u \in R_t$  با احتمال  $p_u(t)$  فعال می شود و با احتمال  $1 - p_u(t)$  غیرفعال باقی می ماند. بنابراین احتمال مشاهده ی مجموعه گره های فعال شده  $V_t$  در زمان  $t$ ، و مجموعه گره های  $R_t$  که در زمان  $t$  فعال نشده اند به این صورت محاسبه می شود:

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

به دلیل مستقل بودن رخدادهای  $p_u(t)$  و  $1 - p_u(t)$  نسبت به خود و نیز نسبت به یکدیگر، رابطه ی زیر، از ضرب آن ها به دست آمده است. مستقل بودن هر یک از  $p_u(t)$  و  $1 - p_u(t)$  نسبت به خود، از آن جا نشأت می گیرد که در یک مرحله ی زمانی فعال شدن یا فعال نشدن تمام گره ها از هم مستقل هستند.

احتمال مشاهده فهرست گره های 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 این که احتمال مشاهده ی مجموع گره های فعال شده و نیز فعال نشده در مراحل زمانی مختلف از هم مستقل هستند، از ضرب استفاده شده است.

در عمل تعداد گره های فعال نشده، بسیار بیشتر از گره های فعال

شده است، اما تعدادشان را برابر در نظر می گیریم و وزن هر گره ی فعال

نشده را برابر با  $\frac{|R_t|}{|V_t|}$  در نظر می گیریم.

حال الگوریتم یادگیری برای این مدل را معرفی می کنیم. هدف این

است که بهترین  $\beta$  ی که می تواند احتمال را بیشینه کند، پیدا کنیم.

بدون از دست دادن عمومیت، الگوریتم را برای حالتی که یک انتشار

داریم، ارایه می دهیم.

## ۳-۵- پیش بینی عملیات

می توان از پارامتر  $\beta$  برای پیش بینی عملیات آینده با داشتن یک مجموعه از گره های فعال شده استفاده کرد. در حقیقت پارامتر  $\beta$  به دست آمده را در رابطه ی احتمال قرار می دهیم. این احتمال ها برای انجام رتبه بندی (چه کسی بیشتر از همه احتمال دارد که در مرحله زمانی بعدی فعال شود) و پیش بینی (پیش بینی تعداد کل گره های فعال شده در مرحله زمانی بعدی) به کار می رود. اگر احتمال فعال سازی به دست آمده برای هر گره، از مقدار آستانه ای (که این مقدار توسط الگوریتم های یادگیری طوری به دست می آید که بهترین پاسخ را بدهد) بیشتر باشد، آن گره را فعال به حساب آورده و از این پس این گره توانایی فعال سازی همسایه های غیر فعالش را دارد.

## ۴-۴- ارزیابی روش پیشنهادی

برای ارزیابی روش پیشنهادی و مقایسه ی نتایج آن با مدل های پیشین [۹]، پارامترهای یاد گرفته شده ی این مدل که از طریق الگوریتم های یادگیری به دست آمدند را بر روی مجموعه داده ی واقعی آزمایش می کنیم. برای این کار نیاز است مجموعه داده را معرفی کرده و نحوه ی ارزیابی را توضیح دهیم.

## ۴-۱- مجموعه داده

شبکه ی آزمون مورد استفاده در این مقاله، شبکه ای چندلایه است که دارای چندین نوع گره ی مختلف و چندین نوع یال مختلف می باشد. دو شبکه ی اطلاعاتی کتابشناسی واقعی DBLP<sup>۱</sup> و ACM<sup>۲</sup> (در حوزه ی علوم کامپیوتر) را انتخاب کردیم. مجموعه داده DBLP در بسیاری از مقالات [۲۳] این حوزه به عنوان شبکه تک لایه در نظر گرفته شده است. اما این شبکه ها گره هایی از نوع نویسنده، مقاله و کنفرانس دارند و می توان 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 گره های با نوع متفاوت (گره های نا همگون یا نامتقارن) نیز وجود دارد. برای مثال ارتباط بین نویسنده و کنفرانس به این معنی که یک نویسنده با نوشتن مقاله ای در یک کنفرانس شرکت کرده است به صورت (APV) نشان داده می شود. همچنین رابطه ی 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 می تواند از آن تاثیر بپذیرد. مشخصات شبکه های DBLP و ACM در جدول ۱ آورده شده است که شامل مقالات سال های ۱۹۳۷ تا ۲۰۱۱ می باشد.

جدول ۱- اطلاعات مجموعه دادگان

تعداد نویسندگان	تعداد مقالات	تعداد کنفرانس ها	
۹۱۶۹۸۸	۱۵۷۲۲۷۸	۶۷۱۳	DBLP
۹۷۴۰۱۸	۲۰۸۴۰۵۵	۷۲۱۰	ACM

## ۴-۲- تحلیل زمانی

الگوریتم پیشنهادی مقاله، برای دیتاهای فوق، با کامپیوتری با حافظه ی ۸ گیگابایت و پردازنده ی Core i7-4702MQ برای پیش بینی انتشار یک موضوع به طور میانگین حدود ۲۴ ساعت طول می کشد. در ابتدا، به دلیل استفاده از ساختار ماتریسی برای نگهداری شبکه ی مورد استفاده، با مشکل کمبود حافظه روبرو می شد که توانستیم با تبدیل این ساختار به ساختار گراف در زبان برنامه نویسی پایتون، این مشکل را برطرف نموده و بتوانیم در زمان خطی از این الگوریتم پاسخ بگیریم.

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

بهبود حاصل شده روی این الگوریتم نسبت به نمونه های قبلی، به دلیل در نظر گرفتن ساختار چندلایه ای و به هم متصل بودن شبکه است که چون گره های متفاوت از نظر نوع، به صورت موازی در کنار گره های هم نوع دیده می شوند، سربار زمانی به الگوریتم اضافه نمی کند و به طور موازی مورد محاسبه قرار می گیرد و بنابراین بر پیچیدگی زمانی افزایش نخواهد داشت.

همچنین بهبود حاصل شده روی این الگوریتم نسبت به نمونه های قبلی، به دلیل در نظر گرفتن ساختار چندلایه ای و به هم متصل بودن شبکه است که چون گره های متفاوت از نظر نوع، به صورت موازی در کنار گره های هم نوع دیده می شوند، سربار زمانی به الگوریتم اضافه نمی کند و به طور موازی مورد محاسبه قرار می گیرد و بنابراین بر پیچیدگی زمانی افزایش نخواهد داشت.

## ۴-۳- معیار ارزیابی

برای مجموعه داده ی معرفی شده، یک مجموعه ی آموزش و یک مجموعه ی آزمایش داریم که با استفاده از این ها می توان انتشار موضوع  $T_c$  در مرحله زمانی  $t+1$  را پیش بینی کرد.

در واقع، احتمال فعال شدن هر نویسنده در مرحله زمانی بعدی به دست می آید و اگر این احتمال از یک مقدار آستانه که توسط الگوریتم های یادگیری، به دست آمده بیشتر باشد، آن نویسنده را در آن زمان فعال اعلام می کنیم و این نویسنده توانایی فعال سازی نویسنده های همسایه ی خود در مرحله زمانی بعدی را دارا می باشد. سپس تمام نویسنده هایی که در هر مرحله، احتمال فعال شدنشان بیشتر از مقدار آستانه بوده را فعال در نظر گرفته و در پایان بازه ی زمانی، تعداد آن هایی که به درستی فعال شناخته شده اند و کسانی که فعال بوده اند، اما الگوریتم ما آن ها را فعال به حساب نیاورده است، را محاسبه کرده و با واقعیت مقایسه می کنیم و معیارهای دقت و فراخوانی را به دست آورده



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

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

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 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 MLTM-R 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

و با استفاده از امتیاز F1، میانگین این دو معیار را سنجیده و با روش های پیشین [۹] مقایسه می کنیم. بسیاری از مقالات [۲۵] این حوزه نیز از این روش ها برای ارزیابی استفاده کرده اند. 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 می کند و در واقع کارایی الگوریتم را نشان می دهد [۲۶]، [۲۷]. در واقع در این معیار تعداد گره های فعال که درست برچسب خورده است، بر تعداد تمام گره های فعال واقعی تقسیم می شود و بدین ترتیب تخمینی از نتایج درست الگوریتم به دست می آید.

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

در رابطه ی ۱۲ عبارتهای مختلف به شکل زیر تعریف می شود:

True Positive (TP): گره های فعال که الگوریتم آن ها را به درستی فعال در نظر گرفته است.

True Negative (TN): گره های غیرفعال که الگوریتم آن ها را به درستی

غیرفعال در نظر گرفته است.

False Positive (FP): گره های فعال که الگوریتم آن ها را به اشتباه

غیرفعال در نظر گرفته است.

همچنین معیار فراخوانی با رابطه ی ۱۳ محاسبه می شود [۲۶]، [۲۷].

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

که در آن False Negative (FN)، گره های غیرفعال است که الگوریتم آن ها را به اشتباه فعال در نظر گرفته است. معیار سوم، امتیاز F1، نیز از رابطه ی ۱۴ محاسبه می شود:

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

## ۵- نتایج

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

### ۵-۱- تاثیر فرآیند انتشاری - انتشار موضوعات مختلف

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

جدول ۴- تاثیر 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

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.

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

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

جدول ۳ برای موضوع Data Mining موجود است و مشاهده می‌شود 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 رسانده‌اند، نمایان‌گر نحوه‌ی رشد آن موضوع (برای مثال موضوع داده‌کاوی (Data Mining)) در شبکه‌ی کتابشناسی ACM می‌باشد. 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	MLTM-R [۹]	۰/۶۸	۰/۴۳	۰/۵۲
	iLTM	۰/۸۹	۰/۵۹	۰/۷۱
Machine Learning	MLTM-R [۹]	۰/۷۴	۰/۶۰	۰/۶۹
	iLTM	۰/۸۶	۰/۷۲	۰/۷۹
SVM	MLTM-R [۹]	۰/۷۴	۰/۴۶	۰/۵۷
	iLTM	۰/۷۹	۰/۶۱	۰/۶۹
Neural Network	MLTM-R [۹]	۰/۹۴	۰/۳۳	۰/۵۰
	iLTM	۰/۹۱	۰/۴۷	۰/۶۲
Naïve Bayes	MLTM-R [۹]	۰/۵۶	۰/۴۵	۰/۵۰
	iLTM	۰/۸۲	۰/۷۵	۰/۷۸

#### ۵-۲- تاثیر فرامسیرهای مختلف و ترکیب‌های آن‌ها

برای نشان دادن کارایی روش پیشنهادی از منظری دیگر، می‌توان تاثیر 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، بدلیل داشتن

مجموعه دادگان به طور میانگین حدود ۱۴ درصد و در دیگری ۱۶ درصد افزایش داشته که این امر نشان‌دهنده‌ی کارایی بیشتر روش پیشنهادی ما، نسبت به روش پیشین است.

برای گسترش این موضوع می‌توان مجموعه دادگانی در حوزه‌های دیگر، مانند شبکه‌ی افراد و طیور را در نظر گرفت و در این شبکه مراحل انتشار بیماری را مدل‌سازی و روند گسترش بیماری را مشاهده کرد و برای جلوگیری از انتشار آن، تمهیداتی اندیشید. مثلاً می‌توان خارج کردن گره با بیشترین ارتباط از شبکه و یا ایمن کردن چنین گره‌ای، یا حذف تعدادی یال با توجه به قدرت انتشاری هر لایه، برای جلوگیری از بیشتر آلوده شدن گره‌ها، را نام برد. همچنین می‌توان به راه‌هایی برای تغییر مسیر انتشار و سازماندهی حرکت فرآیند انتشاری در چنین شبکه‌ای اندیشید و از پخش وقایع بد جلوگیری کرد و یا به پخش اطلاعات مناسب، کمک کرد و هدف‌گذاری برای انتشار ایجاد کرد، بدین صورت که با قصد قبلی، گره‌های اولیه را طوری در شبکه‌ی چندلایه انتخاب کرد که در کوتاهترین زمان ممکن، انتشار تمامی گره‌ها و یا حداقل کسر مناسبی از آن‌ها را درگیر کند.

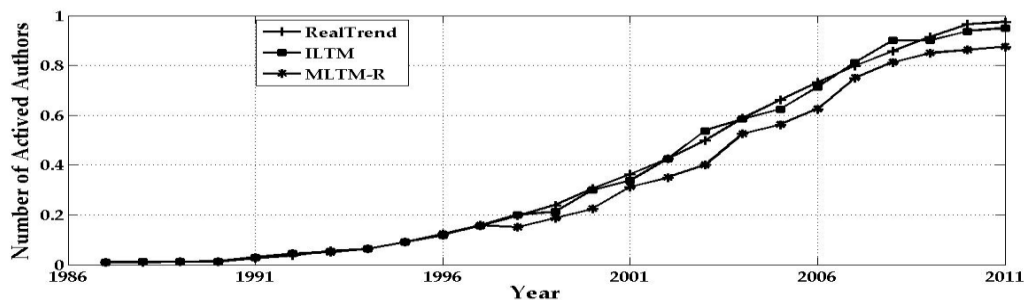
## مراجع

- [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.

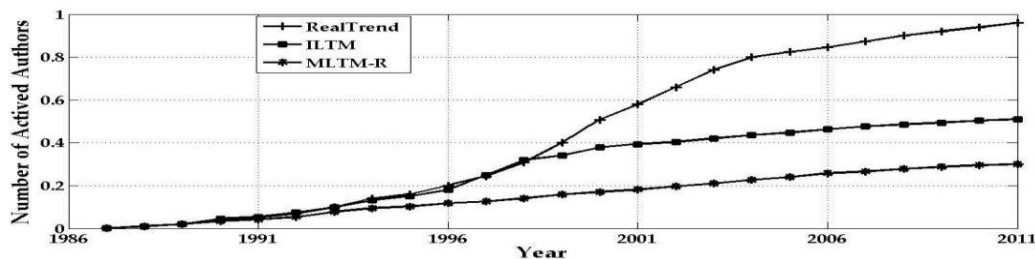
همچنین بر روی مجموعه‌ی داده‌ی ACM همان‌طور که در جدول ۵ مشاهده می‌شود، وقتی تنها یک فرامسیر متقارن را مدنظر قرار دهیم – که در این‌جا فرامسیر نویسنده-مقاله-نویسنده مدنظر است- دقت و فراخوانی و در نتیجه امتیاز F1 بالاتری خواهیم داشت نسبت به حالتی که فرامسیر نامتقارن به تنهایی مورد استفاده قرار گیرد. اما 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، هنگامی که تاثیر 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.

## ۶- نتیجه‌گیری و پیشنهادها

در بخش اول مفاهیم شبکه‌های چندلایه و انتشار را توضیح دادیم. در بخش دوم، مدل‌های انتشار موجود بررسی شدند. در بخش سوم، روشی نوین برای پیش‌بینی انتشار اطلاعات در شبکه‌های 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 می‌تواند نتایج قابل اعتمادتری به دست آورد. در این مقاله، توسعه‌ی مدل آستانه‌ی خطی (LTM) ارائه شد و آن را به شبکه‌های چندلایه تعمیم دادیم. این مدل به شبکه‌ی اطلاعاتی کتابشناسی اعمال شد. مقایسه‌ی روش پیشنهادی و روش پیشین [۹]، نشان داد که نتایج روش پیشنهادی شباهت بیشتری به واقعیت دارد، چرا که نمودارهای نشان‌دهنده‌ی روند رشد تعداد افرادی که به یک موضوع آلوده شده‌اند، بیان‌گر این مطلب هستند که روش پیشنهادی ما نسبت به روش پیشین [۹]، فاصله کمتری از واقعیت دارد. در حقیقت در هر دو مجموعه داده‌ی 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 می‌توان نتیجه گرفت که در شبکه‌ی کتابشناسی که مدل‌مان را به آن اعمال کردیم، نویسندگان مقالات در انتخاب موضوع مقاله‌ی خود، نه تنها از دیگر نویسندگان تاثیر می‌پذیرند، بلکه کنفرانس‌ها نیز می‌توانند بر تصمیم‌گیری آن‌ها 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



شکل ۴- مقایسه روند رشد موضوع داده‌کاوی بر روی مجموعه داده DBLP در واقعیت و پیش‌بینی‌های انجام شده



شکل ۵- مقایسه روند رشد موضوع داده‌کاوی بر روی مجموعه داده ACM در واقعیت و پیش‌بینی‌های انجام شده

- [20] سهیلا مولائی، سما بابایی، مصطفی صالحی، "جستجوی شباهت در شبکه‌های ناهمگن بر مبنای فرامسیرهای وزن‌دار." بیست و یکمین کنفرانس انجمن کامپیوتر ایران، تهران، اسفند ۱۳۹۴، ۲۰-۱۸.
- [21] J. Santisteban, Tejada-Cárcamo, Unilateral Jaccard Similarity Coefficient, GSB@ SIGIR. (2015) 23-27. doi:http://ceur-ws.org/Vol-1393/.
- [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.
- [23] سمیه توکلی، افسانه فاطمی. "تشکیل تیم دوهدفه 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.
- [25] 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.
- [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.
- [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. Kivela, 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.

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

<sup>1</sup> Venue-Author-Venue

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