

محمد مهدی دلیری خمایی^۱، علیرضا رضوانیان^۲ و محمدرضا میبیدی^۳

^۱دانشگاه آزاد اسلامی، واحد قزوین، گروه مهندسی نرم افزار، قزوین، ایران، m.daliri@qiau.ac.ir

^۲دانشگاه صنعتی امیرکبیر، آزمایشگاه محاسبات نرم، گروه مهندسی نرم افزار، تهران، ایران، a.rezvanian@aut.ac.ir

^۳دانشگاه صنعتی امیرکبیر، آزمایشگاه محاسبات نرم، گروه مهندسی نرم افزار، تهران، ایران، mmeybodi@aut.ac.ir

چکیده - با توجه به ارتباطات میان کاربران در شبکه های اجتماعی آنلاین، انتشار اقلام مختلف اطلاعاتی نیز بسیار مورد توجه قرار گرفته است، به طوریکه توجه بسیاری را در بازاریابی مدرن در این شبکه ها به این مسئله جلب نموده است. تا کنون مدل های مختلفی برای مدلسازی انتشار در شبکه های اجتماعی ارائه شده است که در اکثر این مدل ها از حالت ساده ای به صورت کاربران فعال و غیرفعال استفاده شده است. با توجه به این نکته که در دنیای واقعی، نحوه تاثیرپذیری افراد در انتشار اطلاعات به صورت مرحله ای صورت میگیرد، در این مقاله از این ایده برای پیشنهاد یک مدل جدید برای انتشار به صورت مرحله ای استفاده شده است. برای مدل پیشنهادی از اتوماتای سلولی نامنظم بهره برده شده است. که در این مدل حالت های مختلفی از شروع انتشار تا پذیرش توسط کاربر به عنوان وضعیتی از سلول های اتوماتای سلولی نامنظم در نظر گرفته شده است. به منظور ارزیابی مدل پیشنهادی آزمایش های مختلفی بر روی دادگان استاندارد شبکه های اجتماعی معروف صورت گرفته است و براساس پارامترهای مختلفی نیز مورد مقایسه قرار گرفته است. نتایج شبیه سازی حاکی از برتری نسبی مدل پیشنهادی در مقایسه با مدل انتشار آستانه خطی می باشد.

کلید واژه - شبکه های اجتماعی، گسترش انتشار، مدل انتشار، اتوماتای سلولی نامنظم

مقدمه

استفاده کردند. آنها از الگوریتم ژنتیک برای جستجو و انتخاب ترکیب بهینه از مشخصه فیزیکی شبکه استفاده نموده اند. ین و همکاران [۶] یک الگوریتم فرا ابتکاری بر پایه الگوریتم ژنتیک برای انتخاب نودهای اولیه پیشنهاد داده اند. اکثر الگوریتم های اشاره شده به مدل انتشار بستگی شدیدی دارند. به عبارت دیگر کاهش زمان اجرا در مدل انتشار، منجر به اکتساب نتایج در زمان بهتری می شود. بدین گونه کارایی فرایند انتشار از مقیاس پذیری بهتری بهره می گیرد.

علاوه بر الگوریتم های انتخاب نودهای اولیه در فرایند انتشار، بخشی از تحقیقات بر روی مدل های انتشار تمرکز یافته است. یکی از چالش ها در مدل انتشار، اینست که چگونه می توان فرایند انتشار را مناسب تر شبیه سازی نمود. یکی از اولین مدل های انتشار مدل باس می باشد که برای شبیه سازی فرایند تاثیرپذیری یک محصول بکار رفته است. اما این مدل صرفاً بر روی روند انتشار تاکید داشته و نحوه انتخاب کاربران اولیه را در نظر نمی گیرد همچنین این مدل فاقد شبیه سازی دنیای واقعی می باشد. کیمپی و همکاران دو مدل پایه ای از فرایند انتشار را معرفی نمودند. مدل آستانه خطی^۴ و مدل آبشاری مستقل^۵ که بعدها بصورت گسترده ای مورد استفاده قرار گرفته است. این دو مدل شبیه سازی بهتری در مقایسه با مدل باس ارائه می کنند.

در این مقاله یک مدل جدید برای انتشار بر مبنای اتوماتای سلولی نامنظم ارائه شده است که در مدل پیشنهادی برای هر کاربر مقادیر مختلفی از آستانه تاثیر در نظر گرفته می شود تا بتوان به حالت نزدیکی از واقعیت دست یافت. در واقع هدف از این مقاله، در ابتدا ارائه یک مدل حقیقی از فرایند انتشار که بتواند فرایند انتشار را در دنیای واقعی شبیه سازی کند دوم تسریع فرایند انتشار بطوریکه برای

اخیرا با گسترش خدمات آنلاین بر روی اینترنت مانند وب سایتهای شبکه های اجتماعی، مطالعه بر روی شبکه های اجتماعی از جنبه های مختلفی بیش از گذشته مورد توجه قرار گرفته است. تقریباً انواع وب سایت های شبکه های اجتماعی با داشتن یک ساختار اجتماعی از روابط میان کاربران، سعی بر تاثیر به برقراری ارتباط، اشتراک تصاویر و انتشار اطلاعات دارند. ویژگی هایی چون رغبت کاربران به اشتراک گذاری اطلاعات در وب سایت های شبکه های اجتماعی، نظر مدیران بازاریابی را برای ترویج محصول بخود بیش از پیش جلب می نماید. [۱] بازاریابی اشاعه ای، اولین بار توسط دومینگوس^۱ و همکاران از جنبه داده کاوی معرفی شد. بازاریابی ویروسی به طور فرمال در گراف به صورت مسئله پیشینه سازی انتشار معرفی می شود. اولین بار کیمپی و همکاران [۲] این مسئله را فرموله نموده و ثابت کردند که مسئله دارای پیچیدگی سخت بوده و یک الگوریتم حریصانه با ضریب تقریب $1 - \frac{1}{e}$ پیشنهاد نمودند. ایده اصلی الگوریتم آنها روش حریصانه بوده سعی در انتخاب نودهایی با بیشترین تاثیرپذیری در هر تکرار را دارد. این روش دارای بار محاسباتی بالایی می باشد. برای حل این مسئله لسکووک و همکاران [۳] (الگوریتم CELF)^۲ را با استفاده از ویژگی زیرپیمانه ای^۳ پیشنهاد کرده اند. الگوریتم CELF تعداد دفعات اجرای الگوریتم مدل انتشار را کاهش می دهد. با اینحال چن وی و همکاران [۴] نشان دادند که انتخاب مجموعه کوچکی از نود ها ساعت ها بطول می انجامد. حال چنانچه زمان در فرایند انتشار بهبود یابد این مدل می تواند در بازاریابی اشاعه ای مورد مطالعه قرار گیرد. فورست استوندهال [۵] از معیار فیزیکی شبکه برای چینش نودهای اولیه

بازاریابی اشاعه ای مناسب باشد. در این مقاله یک مدل انتشار چندگانه با استفاده از اتوماتای سلولی نامنظم پیشنهاد شده است که در این حالت، کلیه ی نودهای شبکه به صورت یک سلول در نظر گرفته میشود. همچنین کلیه سلول ها وضعیت خود را با توجه به قوانین طراحی شده در حالات تعیین شده در زمان های گسسته تغییر می دهند. مدل پیشنهادی با اقتباس از فرآیند انتشار در واقعیت عمل میکند. در واقع در ابتدا ممکن است یک کاربر فعال شده باشد (یکایید محصول به وی معرفی شده باشد یا یک خبر به وی برسد) اما لزوماً بدین معنی نیست که این کاربر آن را دنبال کند، بلکه حتی ممکن است بعنوان ترویج دهنده (تبلیغ کننده) محصول باشد. همچنین ممکن است که ترویج دهنده محصول این محصول را بوسیله اشتراک گذاری به دیگر کاربران در جهت خرید یا عدم خرید اشاعه دهد. به هر حال با گذشت زمان، کاربر فعال شده یا دنبال کننده ممکن است محصول مورد نظر را خریداری کند. مدل پیشنهادی بر روی دادگان استاندارد و معروف شبکه های اجتماعی مورد ارزیابی قرار گرفته است و در مقایسه با مدل های دیگر مورد بحث و تحلیل قرار گرفته است.

این مقاله در ادامه به این صورت سازماندهی شده است: در بخش دوم اتوماتای سلولی نامتقارن معرفی می شود. در بخش سوم مدل پیشنهادی انتشار چند گانه مبتنی بر اتوماتای سلولی همراه با جزئیات شرح داده شده است در بخش چهارم الگوریتم پیشنهادی معرفی گردید در بخش پنجم به معرفی دادگان آزمایشی و شبیه سازی اختصاص یافته است. بخش ششم نتیجه گیری و کارهای آینده به این مقاله پایان میدهد

۲- اتوماتای سلولی نامنظم

در این مقاله کاربرد مدل جدیدی تحت عنوان اتوماتای سلولی (CA) می باشد در مدلسازی انتشار مورد مطالعه و ارزیابی قرار می گیرد و نشان داده می شود که این مدل می تواند مدل مناسب تری نسبت به مدل های ذکر شده فوق باشد.

اتوماتای سلولی شبکه ای از سلولها است که هر کدام می تواند K وضعیت داشته باشد در حالت یک بعدی، اگر سلول i دارای دو همسایه $i-1$ و $i+1$ باشد در این صورت وضعیت سلول i در زمان $t+1$ یعنی $a_i^{(t+1)}$ بر اساس تابع φ زیر بدست می آید.

$$a_i^{(t+1)} = \varphi(a_{i-1}^{(t)}, a_i^{(t)}, a_{i+1}^{(t)}) \quad (1)$$

تابع $a_i^{(t+1)}$ را قانون اتوماتای سلولی می نامیم. ایده همسایگی در اتوماتای سلولی را می توان بگونه ای بسط داد که دو همسایه دیگر و یا بیشتر را شامل شود. اگر چهار سلول در طرفین یک سلول همسایه محسوب شوند (دو سلول در هر طرف) اتوماتای سلولی دارای شعاع همسایگی (۲) مساوی ۲ می باشد. سلولها در اتوماتای سلولی می توانند با هر ابعادی قرار گیرند

ویژگی اساسی اتوماتای سلولی عبارتند از: فضا و زمان بصورت گسسته می باشد، هر سلول تعداد محدودی وضعیت ممکن را اختیار می کند، تمام سلول ها یکسان می باشند، بروز در آوردن سلولها بشکل همگام می باشد، قوانین بصورت قطعی اعمال می شوند، قانون

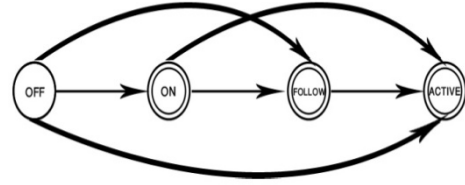
در هر سلول فقط بستگی به مقدار همسایگان اطراف آن دارد. و مقدار جدید هر سلول فقط بستگی به مقادیر تعداد محدود (معمولاً یک مرحله) از مراحل قبل دارد. قوانین به صورت تابعی از وضعیت همسایگی ها می باشد؛ در گروهی از قوانین مقدار یک سلول در مرحله بعدی به مقادیر همسایه ها و در گروهی دیگر، تنها به مجموع همسایه ها بستگی دارد [۷].

اتوماتای سلولی نامنظم در اواسط دهه ۸۰ در مرجع [۸] معرفی شده است بصورت غیر رسمی، اتوماتای سلولی نامنظم دارای یک پیکربندی از نقاط در فضا بدون هیچ محدودیت از پیش تعریف شده می باشد. هر نقطه در این فضا شامل تعداد نقاطی تحت عنوان همسایه می باشد. چند ضلعی ورونی برای تعیین ارتباط و مثلث بندی دلانی برای تقسیم فضا و تعیین همسایگان هر نود نمونه مثال هایی از اتوماتای سلولی نامنظم می باشد. [۹] دیاگرام ورونی روشی برای تقسیم فضا به تعدادی ناحیه می باشد. در این دیاگرام به هر مجموعه ای از نقاط (که دامنه ها، سایت ها و یا ژنراتورها نامیده می شوند) ناحیه ای اختصاص داده می شود. این نواحی سلول های ورونی نامیده می شود. برای یک مجموعه از نقاط دیاگرام ورونی سطح را به مناطقی تقسیم بندی می کند که برای هر نقطه از مجموعه نقاط یک منطقه تعریف می شود. به طوری که تمام نقاط این منطقه به نقطه تولید کننده آن منطقه نزدیکتر می باشد. در حالی که مثلث بندی دلانی مثلث بندی نقاط در نمودار ورونی بوده بطوری که دایره محاط از مثلث بندی نقاط تهی می باشد. در این مقاله، ما از اتوماتای سلولی نامنظم بعنوان یک گراف جهت دار استفاده نموده ایم بطوریکه هر راس در این گراف شامل یک سلول از اتوماتای سلولی نامنظم می باشد.

۳. مدل پیشنهادی

مدل پیشنهادی از چهار وضعیت تشکیل شده است که بیانگر سطح تاثیر نودها در شبکه می باشد. در واقع رویکرد اصلی این مدل تاثیر پذیری متفاوت افراد از یکدیگر می باشد. همانگونه که در شکل ۱ مشاهده می کنید در این مدل چهار وضعیت که بترتیب عبارتند از غیر فعال، فعال، پیروی کننده، و تطبیق یافته تعریف شده است. ایده اصلی در این مدل بدین صورت می باشد که یک فرد می توان در مواجهه با افرادی که قبلاً تاثیر گرفته اند سطح تاثیر مختلفی داشته باشد. مثلاً شرکتی را در نظر بگیرد که برای فرایند بازاریابی به عده ای محدود محصولات خود را رایگان ارائه می کند و انتظار دارد که این افراد در شبکه با توجه به فرایند تاثیر محصول مورد نظر را اشاعه دهند، در این مورد مسئله مهم سطح تاثیر پذیری افراد برای تطبیق با محصول می باشد. در واقع دسته ای از افراد هستند که از محصول آگاهی لازم را پیدا می کند اما آگاهی و رغبت آنها به اندازه ای نیست که ترغیب به خرید محصول شوند که از این دسته در مدل بعنوان اعضای فعال یاد می شود، اما دسته ای دیگری نیز وجود دارند که دارای آگاهی بوده ولی نیاز به تحریک بیشتری برای خرید محصول دارند در حالتی هم می توان افرادی را در نظر گرفت که این افراد به

محض ترغیب شدن به محصول آن را خریداری می نمایند که از این دسته بعنوان افراد تطبیق یافته یاد می شود.



شکل ۱: مدل پیشنهاد شده چند گانه همراه با وضعیت‌های قرار گرفتن هر نود یک گراف جهت دار $G(v, E)$ که v معرف مجموعه تمام نودها در شبکه اجتماعی و E مجموعه تمام یال‌ها نشان داده می شود. برای هر یال (u, v) مقدار احتمال وجود دارد، که نشانگر تاثیر کاربر u بر کاربر v می باشد. مدل انتشار پیشنهادی به عنوان ورودی تعدادی از کاربران را به عنوان کاربران اولیه دریافت می کند و خروجی مدل کاربران فعال شده در پایان فرایند انتشار بر روی شبکه می باشد. در مدل پیشنهادی از آتاماتای سلولی نامنظم استفاده شده است که توسط $l = (S, A, N, R, Sum)$ توصیف شده است. بطوریکه مجموعه S مجموعه تمام حالت‌های سلول که در شکل ۲-۲ ارائه شده می باشد. در اینجا مجموعه $S = \{0 - 0.5, 0.5 - 0.6, 0.6 - 0.8, 0.8 - 1\}$ بدین معنی که اگر سلولی مقدار آستانه آن در اجرای الگوریتم بین ۰ تا ۰.۵ باشد این نود غیر فعال بود و اگر ۰.۵ و کمتر از ۰.۶ باشد این نود فعال شده^۶ و اگر بین ۰.۶ تا ۰.۸ باشد این نود بعنوان نود پیروی کننده محصول^۷ و اگر ۰.۸ تا ۱ بعنوان تطبیق یافته محصول^۸ در نظر گرفته می شود. در سایر حالات نودها بعنوان نودهای غیر فعال در نظر گرفته می شوند. همانگونه که زمان پیش می رود نودها بیشتر و بیشتر پیشنهادات همسایگان خود را قبول می کنند و فرایند تاثیر پذیری ادامه می یابد. اجتماع تمام حالات نودها در شبکه می باشد. همچنین $A = S^n$ که n تعداد نودها می باشد. در هنگام شروع همه نودها غیر فعال می باشند بنابراین $A_0 = \{0, 0, 0, 0 \dots 0\}$. در زمان t وقتی که $A_t = A_{t-1}$ مدل اجرای خود را متوقف می کند زیرا هیچ تغییری در وضعیت نودها ظاهر نمی شود. برای هر سلول مجموعه نودهای ورودی تعریف می کنیم. بعنوان مثال مجموعه نودهای ورودی برای نود v_1 بصورت $\Omega_{v_1} = \{v_2, v_3, v_9 \dots\}$ مولفه N مجموعه کل نودهای ورودی می باشد $N = \{\Omega_{v_1}, \Omega_{v_2}, \dots\}$ مولفه Sum مجموعه کل نودهای فعال شده می باشد. قوانین آتاماتای سلولی نامنظم برای انتشار می باشد که بصورت (۲) تعریف می شود:

$$S_{t+1}(v_i) = \begin{cases} off & S_t(v_i) = off, \sum_{j \in \Omega(i)} (S_t(j)F_t(v_{ij}) - T(t)) < 0.5 \\ on & S_t(v_i) = off, 0.5 \leq \sum_{j \in \Omega(i)} (S_t(j)F_t(v_{ij}) - T(t)) < 0.6 \\ follow & S_t(v_i) = off \text{ or } S_t(v_i) = on, 0.6 \leq \sum_{j \in \Omega(i)} (S_t(j)F_t(v_{ij}) - T(t)) < 0.8 \\ active & S_t(v_i) = off \text{ or } S_t(v_i) = on \text{ or } S_t(v_i) = follow \\ & 0.8 \leq \sum_{j \in \Omega(i)} (S_t(j)F_t(v_{ij}) - T(t)) = 1 \end{cases} \quad (2)$$

قوانین بروزرسانی بالا نشان دهنده نحوی قرار گیری نودها در وضعیت های تعریف شده آنها می باشد. در قوانین بالا $S_t(v_i)$ معرف وضعیت نود v_i در زمان t می باشد، همچنین تابع F_{ij} مقدار تاثیر نود j بر روی نود i را نشان می دهد. فرض ما در این مقاله این است که وقتی یک نود در وضعیت شامل محصول قرار می گیرد حالت خود را حفظ می کند. تابع $T(\Delta t)$ بعنوان تابع کاهشی محسوب میشود که یک تابع غیر صعودی می باشد. در واقع تابع کاهشی $T(\Delta t)$ برای شبیه سازی این واقعیت که تاثیر در طول زمان بطور تدریجی کاهش می یابد. به عبارت دیگر یک کاربر با احتمال بیشتری پیشنهادی که از دوستان خود در زمان اولیه به او داده می شود را قبول می کند. در غیر اینصورت احتمال اینکه به نوبه خود یک محصول را بپذیرد در طول زمان کاهش پیدا می کند. ما تابع کاهش تاثیر را برای شبیه سازی و ارزیابی رویه انتشار بصورت زیر معرفی می کنیم:

$$T(\Delta t) = F \left(\frac{1}{1 + e^{-\Delta t}} \right) \quad (3)$$

در فرمول (۳) مقدار F مقدار تاثیر بر روی یال‌ها می باشد.

۴. الگوریتم پیشنهادی

بر اساس مدل انتشار چندگانه پیشنهادی با استفاده از آتاماتای سلولی نامنظم، الگوریتمی مبتنی بر آتاماتای سلولی ارائه می گردد. شبه کد الگوریتم بشرح زیر می باشد. همانطور که در الگوریتم شماره ۱ مشاهده می شود ورودی الگوریتم یک گراف جهت دار وزن دار و مجموعه نودهای فعال اولیه که با نماد S نشان داده شده می باشد همچنین K تعداد رئوس در کل شبکه است. خروجی الگوریتم تعداد نودهای فعال شده در اتمام الگوریتم می باشد. در مرحله ارزش دهی آغازین متغیر زمان را t برابر صفر و مجموعه A را بعنوان مجموعه وضعیت نودها در طول اجرای الگوریتم تعریف می کنیم در واقع در این مجموعه تغییر وضعیت لحاظ می شود در ابتدا وضعیت نودها off بوده یعنی هیچ نودی توسط الگوریتم فعال یا عبارت دیگر تغییر وضعیت نداده است. خطوط ۱ تا ۶ الگوریتم برای جستجوی نودهای ورودی به هر نود بکار رفته است. پس از اتمام این مرحله در خطوط ۸ تا ۱۰ وضعیت تمام نودهای اولیه فعال را بروز رسانی می نماییم. در گام بعدی الگوریتم تا زمانی ادامه می یابد که هیچ نودی تغییر وضعیت ندهد زیرا عدم تغییر وضعیت به معنی پایداری و تعادل شبکه می باشد در این مرحله برای هر نود بررسی می شود که آیا نود در حال بررسی می تواند وضعیت خود تغییر دهد یا خیر اگر این نود وضعیت خود را تغییر دهد، وضعیت آن در مجموعه A و در زمان تغییر مثلا A_t بروز می شود شرح مراحل شامل خطوط ۱۱ تا ۱۹ الگوریتم می باشد.

بعنوان مدیران شبکه مدیریت می گردد. این مجموعه دادگان در تاریخ ۳ ژوئن ۲۰۰۸ جمع آوری شده است. مجموعه داده چهارم Gnutella که یک شبکه اشتراک گذاری فایل می باشد که بخشی از این شبکه در اگوست ۲۰۰۲ جمع آوری شده است. جزئیات مربوط به هریک از این دادگان در جدول (۱) ارائه شده است.

جدول ۱: شرح جزئیات مربوط به مجموعه دادگان آزمایشی

دادگان آزمایشی	رئوس	یالها	ضریب خوشگی	قطر	چگالی	متوسط درجه
Squeakfoundation	۶۹۷	۵۰۰۹	۰.۴۴	۶	۰.۰۱۳	۰.۴۴۰
Robots.net	۱۶۹۸	۳,۴۹۲	۰.۲۲	۱۱	۰.۰۰۱	۵.۸۵۰
Wiki-vote	۷۱۱۵	۱۰۳۶۸۹	۰.۱۴۰۹	۷	۰.۰۰۲	۲۵.۰
Gnutella	۸۷۱۷	۳۱۵۲۵	۰.۰۰۶۷	۱۰	۰.۰۰۰۴	۴.۷۲

در آزمایش اول، هدف بررسی تاثیر تعداد نودهای اولیه در انتشار در یک شبکه اجتماعی می باشد. بدین منظور از دادگان معرفی شده برای آزمایش استفاده شده است. تعداد نودهای اولیه از ۵٪ تا ۲۵٪ از نودهای هر شبکه به عنوان نودهای اولیه به طور تصادفی در نظر گرفته شده است. در مدل انتشار پیشنهادی که جزئیات آن در بخش ۲ توضیح داده شد با گذشت زمان نودهای غیرفعال موجود در شبکه، از نودهای اولیه همسایگی تاثیر پذیرفته و انتشار در شبکه گسترش پیدا می کند. این گسترش تا زمانی که دیگر تغییری در وضعیت شبکه حاصل نشود ادامه می یابد. نتایج انتشار برای تعداد نودهای اولیه با شرایط مذکور برای مدل آستانه خطی با عنوان LTM و برای مدل پیشنهادی با عنوان MVICA در جداول (۲)، (۳)، (۴)، (۵) نمایش داده شده است. در نتایج گزارش شده میانگین و انحراف معیار زمان لازم برای انتشار و نرخ انتشار (نرخ پوشش) با توجه به تعداد متغیر نودهای اولیه ذکر شده است.

جدول ۲: نتایج مقایسات مدل های انتشار MVICA و LTM برای

شبکه Squeakfoundation

تعداد نودهای اولیه	MVICA		LTM	
	زمان انتشار	نرخ پوشش	زمان انتشار	نرخ پوشش
۳۵	0.4 ± 2.7	58.3 ± 9.7	2.7 ± 0.6	56.3 ± 7.9
۷۰	3.0 ± 0.4	138.1 ± 34.6	2.9 ± 0.4	112.2 ± 19.2
۱۰۵	3.1 ± 0.3	220.8 ± 70.0	3.3 ± 0.5	202.1 ± 25.5
۱۴۰	3.5 ± 0.6	314.4 ± 69.5	3.2 ± 0.4	270.70 ± 79.1
۱۷۵	3.4 ± 0.5	349.2 ± 76.8	3.4 ± 0.6	345.2 ± 84.7

Algorithm 1: proposed cellular automata based algorithm

Input:

$G=(V,E)$ is a Directed graph
 $N=|V|$ // number of vertices
 $K=|S|$ //number of seed set.

Output:

Number of active node's according to proposed model

Pseudo-code:

Initialize $t=1$; $A=\{\text{off}, \text{off}, \dots, \text{off}\}$

01:Begin Algorithm

02:For $i=1$ to n

03: For $j=1$ to n

04: If $((V_i, V_j) \in E)$ then

05: $\Omega_{V_j} \leftarrow V_i$

06: End for

07: End for

08:End For

09:For each V in S

10: $V.state = \text{active}$

11:End for

12:Repeat

13:If $V_j.state$ Is Equal Off Or $V_j.state$ Is Equal On or $V_j.state$ Is Equal Follow) Then

14:Automaton V_i Compute

$(\sum_{j \in \Omega(i)} (S_t(j)F_t(v_{i,j}) - T(t)))$

15: If $(V_j.state \text{ is changed })$ Then

16: Update $V_j State$

17: Update A_t

18: End if

19:End if

20:Until A_{t-1} Equal to A_t

21:End Algorithm

۵. آزمایشات

برای شبیه سازی از چهار نوع مجموعه داده استاندارد مربوط به شبکه های اجتماعی استفاده شده است. اولین مجموعه داده squeakfoundation، یک وب سایت شبکه اجتماعی آنلاین است که جزئیات مربوط به این شبکه اجتماعی در آدرس <http://people.squeakfoundation.org> قابل مشاهده است. مجموعه داده های مورد استفاده در این شبکه در تاریخ ۱۴ دسامبر سال ۲۰۰۷ به صورت ایستا جمع آوری شده است. مجموعه داده دوم یک شبکه اجتماعی تخصصی مربوط به رباتیک می باشد که از طریق سایت <http://robots.net> قابل مشاهده است، که این شبکه در ۱۰ می ۲۰۰۸ جمع آوری شده است. مجموعه داده سوم Wiki می باشد که یک سایت دایره المعارف رایگان می باشد که توسط کاربران خاصی

جدول ۳: نتایج مقایسات مدل‌های انتشار MVICA و LTM برای

شبکه Robot

LTM		MVICA		تعداد نودهای اولیه
نرخ پوشش	زمان انتشار	نرخ پوشش	زمان انتشار	
10.85 ± 6.9	2.8 ± 0.6	4.06 ± 125.1	0.6 ± 2.07	۸۵
269.2 ± 65.3	3.3 ± 0.4	264 ± 45.9	0.4 ± 3.3	۱۷۰
351 ± 51.7	3.1 ± 0.7	416.8 ± 76.5	3.3 ± 0.4	۲۵۵
507.9 ± 58.5	3.5 ± 0.5	523 ± 95.4	3.4 ± 0.5	۳۴۰
621.4 ± 56.4	3.7 ± 0.4	114.8 ± 652.7	3.5 ± 0.5	۴۲۵

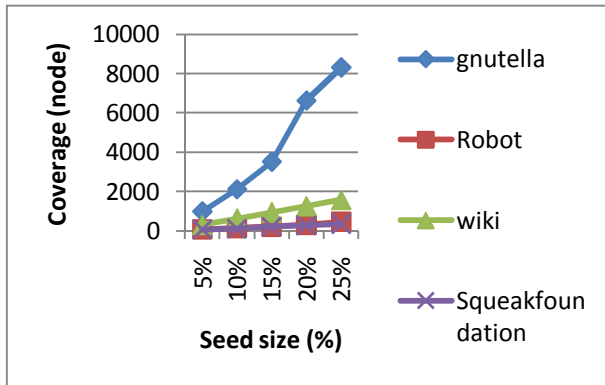
جدول ۴: نتایج مربوط به انتشار برای شبکه wiki-vote

LTM		MVICA		تعداد نودهای اولیه
نرخ پوشش	زمان انتشار	نرخ پوشش	زمان انتشار	
31.0 ± 2.0	2.1 ± 0.3	30.93 ± 2.4	0.2 ± 2.07	۳۰۵
621.8 ± 3.0	2.17 ± 0.3	620.9 ± 3.5	0.2 ± 2.07	۶۱۱
938.2 ± 14.3	2.27 ± 0.4	934.3 ± 3.4	2.13 ± 0.3	۹۱۵
1254 ± 4.3	2.4 ± 0.5	1256.0 ± 4.4	2.23 ± 0.4	۱۲۲۴
1572.3 ± 8.4	2.47 ± 0.5	1574.7 ± 8.4	2.38 ± 0.4	۱۵۳۰

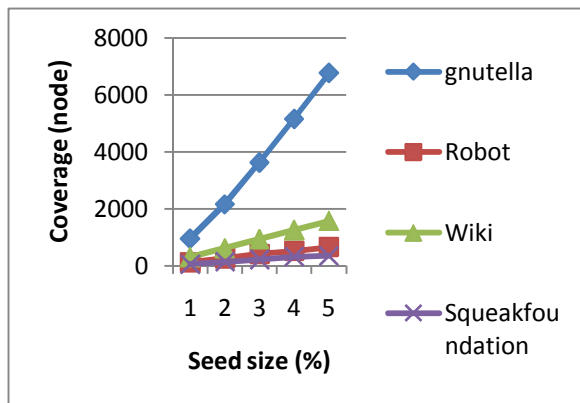
جدول ۵: نتایج مربوط به انتشار برای شبکه Gnutella

LTM		MVICA		تعداد نودهای اولیه
نرخ پوشش	زمان انتشار	نرخ پوشش	زمان انتشار	
979.4 ± 28.6	6.97 ± 2.2	53.2 ± 961.6	1.3 ± 7.2	۴۳۶
210.84 ± 98.2	9.97 ± 2.0	65.2 ± 2165.3	1.2 ± 7.8	۸۷۲
350.56 ± 145.3	12.7 ± 2.6	97.9 ± 3426.6	8.5 ± 0.9	۱۳۰۸
6620.0 ± 904.3	34.57 ± 14.1	94.1 ± 5153.7	8.4 ± 0.8	۱۷۱۷
8305.3 ± 101.6	18.7 ± 3.06	171.9 ± 6768.4	8.1 ± 0.9	۲۱۲۶

نمودار نودهای فعال شده برحسب درصد تعداد نودهای اولیه برای شبکه‌های مختلف به ترتیب در شکل ۷ و ۸ برای نودهای فعال شده تحت مدل انتشار آستانه خطی و مدل انتشار چندگانه پیشنهادی مبتنی بر اتاماتای سلولی نامنظم نشان داده شده است.



شکل ۶: مقایسه نرخ پوشش انتشار برحسب درصد نودهای فعال شده اولیه (۵ تا ۲۵ درصد) برای LTM



شکل ۷: مقایسه نرخ پوشش انتشار برحسب درصد نودهای فعال شده اولیه برای MVICA

با توجه به نتایج ارائه شده در جداول ۲ تا ۵ و شکل‌های ۷ و ۸ برای شبکه‌های مختلف با ویژگی‌های مختلفی چون تعداد نودها، چگالی نودها نتایج متفاوتی بدست آمده است. در نتایج حاصله در مدل پیشنهادی در کلیه حالات به جز زمانیکه از تعداد نودهای اولیه بالایی استفاده می‌شود و شبکه خیلی خلوت است، شاهد نرخ پوشش بالاتری نسبت به مدل آستانه خطی می‌باشیم. چنانچه در جدول شماره ۲ تا ۵ مشاهده می‌شود با افزایش تعداد نودهای اولیه میزان انتشار به طور کلی نیز افزایش می‌یابد، در حالیکه برای زمان انتشار مقایسه با مدل آستانه خطی از میزان کمتری برخوردار است که از دلایل آن می‌توان به وجود ساختار ریز پیمانه ای اشاره کرد. در واقع از آنجائیکه انتشار در روش پیشنهادی نسبت به مدل‌های آستانه خطی، شبیه‌سازی نزدیک‌تری به پدیده انتشار در دنیای واقعی دارد، این مورد در مدل پیشنهادی زمان بیشتری را برای انتقال از حالتی به حالت دیگر را صرف خواهد کرد که این مورد در مدل آستانه خطی صرفاً برای ۲ حالت وجود دارد. اختلاف زمان انتشار در مدل انتشار چندگانه بسیار نزدیک به مدل آستانه خطی و قابل چشم پوشی می‌باشد. با توجه به نتایج بدست آمده ارتباط معکوسی بین دو بین چگالی گراف و زمان انتشار وجود دارد، به طوریکه با کاهش چگالی گراف شاهد افزایش زمان انتشار بوده‌ایم. همچنین با افزایش تعداد رئوس در شبکه‌های مختلف شاهد افزایش نرخ انتشار در شبکه نیز

[9] D. Stevens, "Integration of an Irregular Cellular Automata Approach and Geographic Information Systems for High-Resolution Modeling of Urban Growth", Msc. Thesis, Department of Geography, University of Toronto, 2003.

هستیم. برای معیارهای دیگری مانند ضریب خوشگی و قطر شبکه نمی‌توان ارتباطی با نحوه انتشار ارائه نمود.

۶: نتیجه گیری

انتشار اطلاعات در شبکه اجتماعی یکی از مسائل مهم در بازاریابی اشاعه ای می باشد. ارائه یه مدل نزدیک به واقعیت در این زمینه امری ضروری برای تحلیل شبکه های اجتماعی محسوب می شود. در این مقاله یک مدل انتشار مدل انتشار چند گانه بر پایه اتاماتای سلولی نامنظم ارائه شده است. در مدل پیشنهادی حائلهایی از تاثیرپذیری کاربران در هر سلول و قواعدی برای تغییر حالت هر سلول طراحی شده است. آزمایشات متنوعی برای مدل پیشنهادی بر روی شبکه های اجتماعی استاندارد صورت گرفت که نشاندهنده ی بهبود نسبی مدل پیشنهادی بود. در آزمایشات صورت گرفته اهمیت تعداد نودهای اولیه در نتایج حاصله مشهود می باشد. هم چنین از ویژگی های مدل پیشنهادی میتوان به نزدیک بودن به دنیای واقعی، مقیاس پذیری و سرعت بالای الگوریتم اشاره کرد. برای مدل پیشنهادی در آینده می توان توسعه هایی نیز درباره ماهیت نظرات در نظر گرفت.

۷. مراجع

- [1] M. Richardson, P. Domingos, "Mining knowledge-sharing sites for viral marketing", Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining, July 23-26, 2002
- [2] D. Kempe, J. Kleinberg, É. Tardos, "Maximizing the spread of influence through a social network", Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining, August 24-27, 2003
- [3] J. Leskovec, A. Krause, C. Guestrin, C. Faloutsos, J. VanBriesen, and N. S. Glance. "Cost-effective outbreak detection in networks". In Proceedings of the 13th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, pages 420–429, 2007.
- [4] W. Chen, C. Wang, Y. Wang, " Scalable influence maximization for prevalent viral marketing in large-scale social networks", Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining, July 25-28, 2010
- [5] F. Stonedahl, W. Rand, U. Wilensky, "Evolving viral marketing strategies", Proceedings of GECCO'10 Proceedings of the 12th annual conference on Genetic and evolutionary
- [6] Y. Gui-sheng, W. Ji-jie, D. Hong bin, L. Jia, , "Intelligent Viral Marketing Algorithm over Online Social Network," Networking and Distributed Computing (ICNDC), pp.319-323, 21-24 Sept. 2011
- [7] L. Gordillo and V. Luna, "Parallel Sort on a Linear Array of Cellular Automata", IEEE Trans. Comput, Vol 2, pp. 1904-1910, 1994.
- [8] H. Couclelis, "Cellular Worlds - a framework for modeling micro-macro dynamics", Environment and Planning, pp. 585-596, 1985.

Domingos Et al ^۱
 Cost Effective Lazy Forward ^۲
 Submodular ^۳
 Linear threshold modele ^۴
 Independent Cascade Model ^۵
 On ^۶
 Follow ^۷
 Active ^۸