Archive of SID

ارائه روشی جهت یافتن K مؤثرترین گره‌ها برای تبلیغ در شبکه‌های اجتماعی مبتنی بر ساختار شبکه و الگوهای متفاوت رفتاری گره‌ها (مورد مطالعه: پارسی یار)

مریم شهسواری[۱]، سید علیرضا هاشمی گلپایگانی[۲*]، لیلا اسماعیلی[۳]، سید محمدرضا فخری[۴]

[۱،۲،۳] دانشکده مهندسی کامپیوتر و فناوری اطلاعات، دانشگاه صنعتی امیرکبیر، تهران، ایران

[۴] دانشکده مهندسی کامپیوتر و فناوری اطلاعات، دانشگاه قم، قم، ایران

{maryamshahsavari, sa.hashemi, Leila.esmaeili}@aut.ac.ir, smrfakhri@gmail.com

چکیده

در سال‌های اخیر توجه بسیاری از بازاریاب‌ها بر موضوع انتشار اثر و بیشینه‌سازی آن در شبکه‌های اجتماعی معطوف بوده است. چرا که پژوهش‌ها نشان داده‌اند که کاربران شبکه‌های اجتماعی بر توصیه‌ی دوستان خود یا دیگر مصرف‌کنندگانی که محصول را قبلاً خریداری کرده‌اند اعتماد بیش‌تری دارند و کم‌تر تحت تأثیر تبلیغات رسانه‌ای قرار می‌گیرند. از طرف دیگر اثرات اجتماعی بر فرایند انتشار اطلاعات، تأثیر مستقیم دارد و در هر شبکه‌ی اجتماعی کاربران پرنفوذی وجود دارند که تأثیر زیادی بر نظرات دیگران و انتشار اطلاعات در کل شبکه دارند؛ شرکت‌ها در تلاش‌اند تا با هدف قرار دادن پرنفوذترین کاربران شبکه‌های اجتماعی، فرایند انتشار را کنترل کرده و با صرف هزینه‌ی اندکی برای بازاریابی آن را به بیش‌ترین حد ممکن برسانند. در این مقاله بیشینه‌سازی اثر در شبکه‌های اجتماعی و یافتن افراد پرنفوذ مورد توجه قرار گرفته است و یک مدل انتشار حدآستانه‌ی مبتنی بر مدل‌های رفتاری افراد در شبکه‌های اجتماعی ارائه شده است. از طرف دیگر می‌توان الگوریتم حریصانه را به عنوان متداول‌ترین راه‌حل ارائه‌شده برای حل مسئله‌ی بیشینه‌سازی اثر برشمرد که از زمان اجرای بالایی برخوردار می‌باشد. در این مقاله ایده‌ای برای کاهش زمان اجرای الگوریتم حریصانه ارائه شده است. ارزیابی آزمایش‌های انجام‌شده حاکی از بهبود روش پیشنهادی نسبت به الگوریتم پایه می‌باشد.

واژه‌های کلیدی

شبکه‌های اجتماعی، اثر اجتماعی، انتشار اثر، بیشینه‌سازی اثر، گره‌های پرنفوذ

۱- مقدمه

یک شبکه‌ی اجتماعی ساختاری اجتماعی است که افراد یا سازمان‌ها را به یکدیگر متصل می‌کند. مانند شبکه‌ی پست الکترونیکی، شبکه‌ی اجتماعی فیس‌بوک و شبکه‌ی هم‌کارانه‌ی علمی [۱]. شبکه‌های اجتماعی نقش مهمی در انتشار اطلاعات و رفتار دارند که از آن با نام اثر «دهان‌به‌دهان[۱]» در شبکه‌های اجتماعی یاد می‌شود [۲]. اثر دهان‌به‌دهان به معنای انتقال یک اثر از یک فرد به فرد دیگر در شبکه‌ی اجتماعی می‌باشد. اثر می‌تواند پیروی از هر نوع گرایش، طرز تفکر، رفتار و یا محصول جدید در شبکه‌ی اجتماعی باشد.

با توجه به ساختار قابل نگاشت شبکه‌های اجتماعی به گراف، امروزه تحلیل شبکه‌های اجتماعی[۲] (SNA) با استفاده از روش‌های تحلیل ریاضیاتی مورد توجه محققان در علوم مختلف قرار گرفته است [۳]. تحلیل ساختار روابط پیچیده میان گره‌ها در یک شبکه، بسیار مفید است. یکی از موارد مورد استفاده‌ی آن، اطلاع از چگونگی انتشار اطلاعات در شبکه می‌باشد [۱] که در کاربردهای گوناگون به ویژه بازاریابی مورد توجه است.

یکی از انواع بازاریابی در شبکه‌های اجتماعی، بازاریابی ویروسی می‌باشد که بر پایه‌ی اعتماد میان اشخاص است [۴]. تحقیقات نشان می‌دهند که افراد به اطلاعاتی که از روابط اجتماعی نزدیک خود دریافت می‌کنند اعتماد بیشتری دارند. به همین دلیل، بسیاری معتقد هستند که بازاریابی

[1] Word of Mouth

[3] Social Network Analysis

* نویسنده مسئول

www.SID.ir

Archive of SID

ویروسی مؤثرترین بازاریابی برای اهداف تجاری می‌باشد. وجود شبکه‌های اجتماعی مانند فیس‌بوک، مای‌اسپیس و توییتر فرصت‌های جدیدی برای بازاریابی ویروسی برخط فراهم می‌کند [۴].

حال این سناریو را در نظر بگیرید که یک شرکت کوچک، یک برنامه‌ی کاربردی برخط توسعه می‌دهد و قصد دارد تا آن را از طریق یک شبکه‌ی اجتماعی برخط تبلیغ کند. درعین‌حال بودجه‌ی محدودی در اختیار دارد و فقط می‌تواند تعداد کمی از افراد شبکه را انتخاب کرده و استفاده‌ی رایگان از نرم‌افزار را برای آن‌ها فراهم کند. این شرکت امیدوار است که افراد اولیه از نرم‌افزار جدید استقبال کرده و آن را به سایرین پیشنهاد کنند و سایرین هم به افراد دیگر و این روند ادامه پیدا کند تا جایی که بر اثر این تبلیغ دهان‌به‌دهان، عده‌ی زیادی از نرم‌افزار مطلع شده و آن را خریداری کنند. مسئله این است که افراد اولیه چه کسانی باشند تا اثرگذاری در شبکه به مقدار بیشینه برسد [۲].

مسئله‌ای که مطرح شد با عنوان مسئله‌ی بیشینه کردن تأثیر شناخته می‌شود. کمپ و همکارانش این مسئله را که برای اولین بار توسط ریچاردسون معرفی شد، بدین صورت تشریح می‌کنند [۲]: مسئله‌ی بیشینه‌سازی اثر یا انتشار، پیدا کردن k گره اولیه در یک شبکه‌ی اجتماعی می‌باشد، به‌گونه‌ای که با شروع فرایند انتشار از این k گره، در نهایت انتشار اثر بیشینه شود. روش‌هایی که تا کنون مطرح شده‌اند سعی بر بهبود زمان اجرای الگوریتم حریصانه با استفاده از یکی از مدل‌های انتشار داشته‌اند، اما هیچ یک از آن‌ها موضوع متفاوت بودن الگوهای رفتاری گره‌ها و میزان تمایل آن‌ها در موضوعات متفاوت را در نظر نگرفته‌اند.

در این مقاله مدل حدآستانه با در نظر گرفتن الگوهای متفاوت رفتاری گره‌ها و متفاوت بودن علاقه‌مندی گره‌ها در موضوعات مختلف بسط داده می‌شود. ازآنجاکه الگوریتم حریصانه با تقریب خوبی جواب نزدیک به بهینه را می‌دهد، بسیاری از پژوهش‌ها روش خود را بر اساس الگوریتم حریصانه مطرح کرده‌اند. ازجمله مشکلات الگوریتم حریصانه زمان اجرای بالای آن است. به همین دلیل در این مقاله روشی برای یافتن گره‌های پرنفوذ بر اساس الگوریتم حریصانه ارائه می‌شود که با در نظر گرفتن الگوهای متفاوت رفتاری و هم‌چنین علاقه‌مندی متفاوت گره‌ها به موضوعات مختلف، زمان اجرای الگوریتم حریصانه را به‌شدت کاهش می‌دهد. همچنین این روش قادر است جوابی نزدیک به جواب الگوریتم حریصانه‌ی پایه‌ی پیدا کردن k گره پرنفوذ به دست آورد.

در ادامه‌ی این مقاله در بخش ۲ مرور تحقیقات مرتبط آمده است. بخش ۳ به معرفی روش پیشنهادی می‌پردازد؛ چگونگی طراحی آزمایش‌ها، نحوه‌ی ارزیابی و نتایج حاصل از انجام آزمایش‌ها در بخش ۴ ارائه می‌شود. در پایان نیز نتیجه‌گیری و کارهای آتی بیان خواهد شد.

۲- مرور ادبیات مرتبط

یکی از روش‌های پیدا کردن k گره پرنفوذ در شبکه‌های اجتماعی الگوریتم حریصانه‌ی پیشنهادی کِمپ می‌باشد؛ اما روش‌های دیگری نیز برای پیدا کردن گره‌های پرنفوذ ارائه شده است که خلاصه آن در جدول ۱ آمده است.

طبق جدول ۱ بیش‌ترین تلاش‌ها در جهت بهبود زمان اجرای الگوریتم حریصانه و مقیاس‌پذیر بودن آن انجام شده‌اند. همچنین تنها در کار تحقیقاتی [۵] در سال ۲۰۱۴، مسئله بیشینه‌سازی اثر بر اساس علاقه کاربران مطرح شده است. در این مقاله یک الگوریتم دومرحله‌ای پیشنهاد داده شده است که ابتدا علاقه‌مندی کاربران در هرموضوع را به دست می‌آورد. سپس یک بردار علاقه به ازای هر موضوع به الگوریتم داده می‌شود. بر اساس بردار علاقه‌مندی یک احتمال انتشار به ازای هر موضوع برای هر یال محاسبه می‌شود، در مرحله‌ی بعدی از الگوریتم پیشنهاد شده در [۶]، برای محاسبه انتشار و انتخاب بهترین گره استفاده شده است.

ما در این مقاله قصد داریم تا با ارائه روشی جدید برای به دست آوردن علاقه‌مندی گره‌ها، مدل حدآستانه را بسط دهیم، به طوری که با در نظر گرفتن الگوهای متفاوت رفتاری برای گره‌ها و متفاوت بودن میزان علاقه‌مندی گره‌ها در موضوعات مختلف، انتشار را شبیه‌سازی نموده و با تغییر الگوریتم حریصانه میزان محاسبات آن را کاهش دهیم.

۳- چارچوب روش پیشنهادی

همان‌طور که پیش‌تر اشاره شد، برای پیدا کردن افراد پرنفوذ در یک شبکه‌ی اجتماعی از الگوریتم حریصانه‌ی پایه‌ای استفاده می‌شود که خروجی آن گره‌هایی هستند که با شروع فرایند انتشار از آن‌ها، انتشار محتوا در شبکه بیشینه خواهد بود؛ به منظور شبیه‌سازی انتشار اثر مدل‌های زیادی پیشنهاد شده‌اند. دو مدل متداول مدل حد آستانه و مدل آبشاری می‌باشند. در این پژوهش مدل حدآستانه توسعه داده شده است. در مدل حدآستانه طبق رابطه‌ی (۱) از حدآستانه‌ی فعال شدن هر گره و مجموع اثرات دریافتی آن از همسایه‌های فعالش برای بررسی فعال شدن یا نشدن یک گره استفاده می‌شود [۲]:

$$\sum_{w} b_{vw} \geq \theta_v \qquad (۱)$$

بنابراین مقدار حد آستانه‌ی θ_v نشان‌دهنده‌ی تمایل گره‌ها برای تطبیق با اثری است که همسایه‌های آن‌ها با آن تطبیق یافته‌اند. در هر دو مدل آبشاری و حد آستانه، یک مجموعه گره اولیه در نظر گرفته می‌شود که این مجموعه گره‌های فعال هستند و با S نمایش داده می‌شوند. تابعی که اثر مجموعه گره S را محاسبه می‌کند با $\sigma(S)$ نمایش داده می‌شود. درواقع $\sigma(S)$ نشان‌دهنده‌ی تعداد گره‌های فعال‌شده در پایان فرایند انتشار است که S مجموعه گره اولیه بوده است که فرایند انتشار با آن آغاز شده است. مسئله‌ی بیشینه‌سازی درواقع پیدا کردن k گره اولیه با هدف بیشینه‌سازی انتشار اثر است[۷]..

www.SID.ir

Archive of SID

جدول ۱: خلاصه ادبیات مرتبط (کاهش محاسبات نسبت به الگوریتم حریصانه = RGreddyT)

شماره مرجع	سال	مفروضات			هدف	راه‌حل
		مدل انتشار پایه	تفاوت الگوی رفتاری گره‌ها	تفاوت علایق گره‌ها در موضوعات مختلف		
[۶]	۲۰۰۷	مدل حدآستانه	خیر	خیر	RGreddyT	استفاده از ویژگی زیرپیمانه‌ای انتشار اثر در شبکه-های اجتماعی
[۹]	۲۰۰۷	مدل آبشاری مستقل	خیر	خیر	RGreddyT	الگوریتم حریصانه‌ی «پوشش مجموعه‌ای[1]»
[۱۰]	۲۰۰۸	مدل انتشار حرارت	خیر	خیر	RGreddyT	الگوریتم k گره برتر
						الگوریتم حریصانه‌ی kگام
						الگوریتم پیشرفته‌ی kگام
[۶]	۲۰۰۹	مدل آبشاری مستقل	خیر	خیر	RGreddyT	استفاده از ویژگی مدل آبشاری و حذف برخی یال‌ها
[۱۱]	۲۰۰۹	مدل حد آستانه	خیر	خیر	RGreddyT	ارائه یک الگوریتم بهینه با پیچیدگی زمانی چند-جمله‌ای
[۹]	۲۰۰۹	مدل حد آستانه	خیر	خیر	RGreddyT	انتخاب مبتنی بر پتانسیل[2]
[۱۲]	۲۰۱۰	مدل آبشاری مستقل	خیر	خیر	RGreddyT	استفاده از تشخیص اجتماعات
[۱۳]	۲۰۱۰	مدل حد آستانه	خیر	خیر	RGreddyT	ارائه یک الگوریتم با استفاده از مفاهیم بازی‌های تعاملی[3]
[۱۴]	۲۰۱۰	مدل حدآستانه	خیر	خیر	RGreddyT	رویکرد حریصانه‌ی مبتنی بر مرکزیت درجه و بینابینی[4]
[۱۵]	۲۰۱۲	مدل انتشار حرارت	خیر	خیر	RGreddyT	استفاده از ساختار اجتماعات برای شناسایی گره-های پرنفوذ
[۱۶]	۲۰۱۳	حدآستانه	خیر	خیر	RGreddyT	ارائه‌ی یک مدل بهینه‌سازی برای پیدا کردن بهترین مسیر انتشار اطلاعات
[۵]	۲۰۱۴	مدل آبشاری مستقل	خیر	بله	RGreddyT	پیشنهاد روشی برای به دست آوردن میزان علاقه-مندی کاربران و مبتنی بر الگوریتم [۷]
پیشنهادی	۲۰۱۶	مدل حدآستانه	بله	بله	RGreddyT	کنار گذاشتن برخی گره‌ها از پیمایش

[1] Set cover Greddy Algorithm

[2] Potential Based Selection

[3] Cooperative Game theory

[4] Betweenness

www.SID.ir

Archive of SID

کمپ و همکارانش اثبات کردند که این مسئله، NP-سخت می‌باشد و یک الگوریتم حریصانه‌ی تقریبی ارائه کردند که مؤلفه‌ی اصلی آن، محاسبه‌ی انتشار اثر است که کار سختی می‌باشد. درواقع بزرگ‌ترین مشکل الگوریتم حریصانه زمان اجرای آن است که در شبکه‌های بزرگ بسیار زیاد است.

پرواضح است که روند انتشار موضوعات مختلف در شبکه‌های اجتماعی یکسان نخواهد بود [۸]. محتوا می‌تواند تبلیغ کالا با موضوعات مختلف باشد و افراد علاقه‌مند به موضوع تبلیغ، پتانسیل بیش‌تری برای انتشار تبلیغ خواهند داشت. ایده‌ی اولیه‌ی این پژوهش تفاوت در الگوهای رفتاری افراد از حیث انتشار محتوا می‌باشد. فرض اولیه این است که علاقه‌مندی افراد نسبت به موضوعات مختلف متفاوت است و به‌تبع آن آستانه‌ی تحریک‌پذیری افراد نسبت به انتشار محتواهایی با موضوعات مختلف نیز متفاوت می‌باشد.

بدین منظور در این مقاله ابتدا به تحلیل علاقه‌مندی افراد به موضوعات مختلف پرداخته شده و بر اساس آن یک بردار حد آستانه به هر فرد تخصیص داده شده است که طبق مدل انتشار حدآستانه‌ی توسعه‌یافته، رفتار بازنشر هر فرد تابعی از بردار حد آستانه‌ی او خواهد بود؛ منظور از رفتار بازنشر، عکس‌العمل کاربر پس از دریافت محتوا در قبال انتشار آن می‌باشد. بردار حدآستانه برداری است که در هر مؤلفه از آن مقدار حدآستانه‌ی کاربر در یکی از موضوعات، مشخص شده است. اما سؤالی که در اینجا مطرح می‌شود این است که علاقه‌مندی افراد در یک شبکه‌ی اجتماعی را چطور می‌توان تفسیر نمود؟ در این پژوهش از روش خوشه‌بندی برای دسته‌بندی گره‌ها و استخراج علاقه‌مندی آن‌ها از روی داده‌های مجموعه داده شبکه‌ی اجتماعی سرویس وبلاگ‌نویسی پارسی‌بلاگ (پارسی‌یار) [۲۱] استفاده شده است تا با یک دقت نسبی بتوان عددی به میزان علاقه‌مندی گره‌ها منتسب نمود و این مقادیر مبنایی برای به دست آوردن حد آستانه‌ی افراد در موضوعات مختلف خواهند بود.

در مجموعه‌ی داده‌ی شبکه‌ی اجتماعی پارسی‌یار، ساختار شبکه‌ی ارتباطی افراد و اطلاعاتی در مورد عضویت افراد در گروه‌هایی با موضوعات مختلف وجود دارد.

۳-۱- استخراج و تفسیر علاقه‌مندی گره‌ها در شبکه‌ی اجتماعی

برنامه‌های تبلیغاتی علاوه بر دانستن اطلاعات صفحات شخصی کاربران، باید از فعالیت‌هایی که کاربران در شبکه‌های اجتماعی انجام می‌دهند نیز آگاهی داشته باشند [۱۷]. با ظهور فناوری‌های جدید، سامانه‌های توصیه‌گر ظهور پیدا کردند، که ایده‌ی اولیه‌ی این سامانه‌ها این است که محصولاتی را برای مشتری تبلیغ کنند که با علایق و ترجیحات او هم‌خوانی دارد. علایق و ترجیحات نیز یا از روی رتبه‌دهی‌های کاربران به‌طور صریح یا از روی تراکنش‌های قبلی کاربران، کوکی‌ها و گزارش‌های وبی به‌طور ضمنی به دست می‌آیند [۱۸]. در این پژوهش تلاش شده است تا با استفاده از اطلاعات کاربران علاقه‌مندی آن‌ها به‌طور ضمنی به دست آید.

عضویت افراد در گروه‌های مختلف حاکی از گرایش‌ها و طرز تفکر آن‌ها می‌باشد. در مجموعه داده‌ای که در این پژوهش مورداستفاده قرار گرفته است، اطلاعاتی در مورد عضویت و فعالیت‌های افراد در گروه‌هایی با موضوعات مختلف موجود می‌باشد. براساس مجموعه داده، اطلاعات زیر برای هر کاربر قابل استخراج است:

۱. تعداد گروه‌ها عضو با موضوع مشخص
۲. تعداد مباحث ثبت شده در گروه‌ها با موضوع مشخص
۳. تعداد نظرهای ثبت شده در گروه‌ها با موضوع مشخص
۴. متوسط تعداد سال‌های عضویت در گروه‌های مختلف

در گام بعد با توجه اطلاعات قابل‌استخراج، کمّی‌سازی علاقه‌مندی افراد انجام می‌شود، این عدد به عنوان معیاری برای تحلیل الگوی رفتاری در انتشار محتوا استفاده خواهد شد. در این مرحله به منظور کمّی سازی علایق کاربران از خوشه‌بندی استفاده می‌شود. خوشه‌بندی یک روش بدون-نظارت است که کاربردهای بسیاری دارد [۱۹]. در این مقاله از روش خوشه‌بندی K-means به‌منظور دسته‌بندی کاربران استفاده می‌شود و براساس نتایج، عددی به علاقه‌مندی کاربران نسبت داده خواهد شد.

ازآنجاکه تعداد خوشه‌ها یا K در روش K-means می‌بایست از ابتدا مشخص باشد و ورودی الگوریتم است، و در پژوهش حاضر این تعداد مجهول می‌باشد، برای تعیین بهترین مقدار K از شاخص Dunn [۲۰] استفاده می‌شود. به این صورت که برای خوشه‌بندی کاربران در هر موضوع مقدار K از ۲ تا ۵۰ در نظر گرفته شده و خوشه‌بندی انجام شده است. در هر بار انجام عمل خوشه‌بندی مقدار شاخص Dunn توسط نرم‌افزار Matlab محاسبه شده است. بهترین مقدار K برای خوشه‌بندی کاربران در هر موضوع مقداری است که شاخص Dunn در آن حالت نسبت به سایر حالت‌ها بیشینه باشد. مقدار K بر این اساس برای موضوعات c_1 تا c_5 به ترتیب برابر با ۳، ۳، ۴، ۳ و ۵ به دست آمد که در مراحل بعدی روش مورد استفاده قرار می‌گیرد. ۵ موضوع موردنظر c_1 تا c_5 به ترتیب عبارت‌اند از اجتماعی، بازی، ورزشی، مذهبی و علوم.

۳-۲- اختصاص عدد به حدآستانه‌ی گره‌ها در موضوعات مختلف

با در نظر گرفتن مقادیر منتسب به علاقه‌مندی کاربران در مرحله قبل باید مقداری برای حد آستانه‌ی آن‌ها مفروض دانست. ازآنجاکه میزان علاقه-مندی و حد آستانه‌ی تحریک‌پذیری افراد برای انتشار محتوا رابطه‌ی معکوس با یکدیگر دارند، باید به نحوی به علاقه‌مندی بیش‌تر مقدار کم‌تری برای حد آستانه منتسب شود. بنابراین θ حد آستانه به صورت $\theta = 1-\alpha$ تعریف می‌شود، α مقدار علاقه‌مندی افراد می‌باشد. براساس مقادیر حدآستانه برای هر موضوع، بردار حد آستانه برای هر کاربر تعریف می‌شود.

www.SID.ir

Archive of SID

۳-۳- توسعه‌ی مدل انتشار حدآستانه بر مبنای الگوهای رفتاری گره‌ها

یکی از تفاوت‌های روش انتشار در این مقاله با روش‌های انتشار قبلی در نظر گرفتن موضوع محتوای انتشاریافته (در اینجا محتوا حداقل به یکی از ۵ موضوع c_1 تا c_5 تعلق دارد) و الگوهای رفتاری گره‌ها به عنوان ورودی‌های فرایند انتشار می‌باشد. به عنوان مثال یک محتوا می‌تواند هم خبری و هم ورزشی باشد. هر محتوا طبق رابطه (۲) تعریف می‌شود:

$$S = \{c_i \mid i \in N, c_i \in \{0,1\}\} \quad (۲)$$

در رابطه (۲)، S نام محتوا و C_i ها عنوان موضوعات مختلف هستند. صفر بودن C_i به این معناست که محتوای S دارای آن موضوع نیست و بلعکس.

از طرفی هر گره به ازای هر موضوع یک مقدار حدآستانه‌ی متفاوت خواهد داشت. چون علاقه‌مندی افراد نسبت به موضوعات مختلف متفاوت می‌باشد [۸]. مقادیر متفاوت حدآستانه‌ی هر گره در موضوعات مختلف به صورت یک بردار (t) نمایش داده شده و طبق رابطه (۳) تعریف می‌شود:

$$t_n = \{\theta_i, i \in N, \theta_i \in [0,1]\} \quad (۳)$$

در رابطه (۳)، t_n بردار حدآستانه‌ی گره n می‌باشد که مقادیر آن در بازه صفر و یک بوده و در هر موضوع i مقدار حدآستانه‌ی گره n مشخص می‌شود.

در مدل انتشار حد آستانه‌ی خطی فرایند انتشار مستقل از موضوع محتوا شبیه‌سازی شده است و موضوع محتوا و تناسب آن با علاقه‌مندی کاربران در نظر گرفته نمی‌شود؛ اما در واقعیت فرایند انتشار برای موضوعات مختلف در یک شبکه یکسان نیست. به همین دلیل در این مقاله پارامتر محتوا به همراه بردار حد آستانه‌ی افراد در مدل انتشار در نظر گرفته می‌شود. فرایند انتشار جدید به این صورت است که به منظور شروع انتشار از یک مجموعه گره اولیه، موضوع محتوا با مقدار مربوطه‌ی آن در بردار حد آستانه‌ی هر گره مطابقت داده شده و مقایسه بین جمع اثرات دریافتی از گره‌های همسایه و مقدار مربوط به موضوعات محتوا در بردار حد آستانه انجام می‌شود. اما این برای حالتی است که محتوا یک موضوعی باشد. اگر محتوا دارای چند موضوع باشد، حالت‌های دیگری می‌تواند اتفاق بیفتد. الگوهای رفتاری مختلف به ازای محتوای چندموضوعی بررسی می‌شود.

۴-۳- الگوهای رفتاری

رفتار بازنشر گره‌ها نسبت به هر محتوا بستگی به تمایلات و علاقه‌مندی آن‌ها دارد که در پارامتر θ ترجمه می‌شود و برخلاف کارهای پیشین، برای هر گره مقدار متفاوتی دارد. علاوه بر آن، مفهوم دیگری با نام الگوهای رفتاری تعریف می‌شود. الگوهای رفتاری مختلفی ممکن است به هنگام انتشار یک محتوا توسط گره‌های داخل شبکه انجام شود.

با فرض گره A که دارای بردار حدآستانه t می‌باشد و محتوای S چندموضوعی (مثلاً c_3 و c_5) که به A ارسال می‌شود، رفتار گره A در مقابل محتوای S می‌تواند یکی از سه حالت زیر باشد:

۱- **الگوی اول:** در این الگو، اگر اثر دریافت شده، از بیش‌ترین مقدار بردار حد آستانه‌ی مربوط به موضوعات محتوا بیش‌تر باشد، آن گره فعال شده و محتوا را انتشار می‌دهد؛ یعنی شرط فعال شدن برای الگوی رفتاری اول رابطه (۴) است.

$$\text{If } \sum_w b_{vw} \geq \max\{\theta_i \mid \forall i\ S_{c_i} \neq 0\} \text{ then node v is activated.} \quad (۴)$$

که در آن $\sum_w b_{vw}$ اثرات دریافت شده توسط گره v از تمامی همسایه‌های فعالش (w) می‌باشد و S_{c_i} هر مؤلفه از بردار موضوعی محتوا و θ_i هر مؤلفه از بردار حدآستانه‌ی افراد می‌باشد.

۲- **الگوی دوم**: در این الگو، اگر اثر دریافت شده، به میانگین حد آستانه‌ی مربوط به همه‌ی موضوعات محتوا برسد آن گره فعال شده و محتوا را انتشار می‌دهد؛ یعنی شرط فعال شدن برای الگوی رفتاری دوم رابطه (۵) است.

$$\text{If } \sum_w b_{vw} \geq avg\{\theta_i \mid \forall i\ S_{c_i} \neq 0\} \text{ then node v is activated.} \quad (۵)$$

۳- **الگوی سوم:** در این الگو، اگر اثر دریافت شده از کم‌ترین مقدار بردار حد آستانه‌ی مربوط به موضوعات محتوا بیش‌تر باشد، آن گره فعال شده و محتوا را انتشار می‌دهد (کمینه‌ی مقدار در بردار حد آستانه مدنظر است). در این حالت شرط فعال شدن رابطه (۶) است.

$$\text{If } \sum_w b_{vw} \geq \min\{\theta_i \mid \forall i\ S_{c_i} \neq 0\} \text{ then node v is activated.} \quad (۶)$$

افراد در یک شبکه‌ی اجتماعی ممکن است از الگوهای رفتاری متفاوتی از حیث انتشار یک محتوا تبعیت کنند و انتشار محتواهایی با موضوعات مختلف را تحت تأثیر قرار دهند [۸]. به عنوان مثال، دو فرد با علایق یکسان، که هر دو علاقه‌ی زیادی به محتوای ورزشی و علاقه‌ی کمی به محتوای اجتماعی دارند، به هنگام دریافت یک محتوا با دو موضوع اجتماعی و ورزشی رفتار متفاوتی خواهند داشت. با در نظر گرفتن میزان اثر دریافتی هر گره، ممکن است یکی از آن‌ها محتوای دریافتی را نشر دهد اما فرد دیگر محتوا را انتشار ندهد. به همین دلیل یکی از فرضیات در این مقاله، متفاوت بودن الگوهای رفتاری گره‌ها در شبکه است و به صورت تصادفی به هر گره الگویی نسبت داده می‌شود.

۵-۳- بهبود الگوریتم حریصانه

الگوریتم حریصانه با تقریب خوبی جواب بهینه یافتن k رأس اثرگذار را پیدا می‌کند؛ اما مشکل بزرگ این الگوریتم، زمان اجرای بالای آن است که مربوط به عنصر اصلی آن یعنی محاسبه‌ی انتشار اثر برای هر گره در تکرارهای مختلف الگوریتم می‌باشد [۲]. به همین دلیل در این مقاله، هدف کاهش میزان محاسبات الگوریتم حریصانه که مربوط به محاسبه‌ی انتشار اثر برای هر گره می‌باشد، است.

www.SID.ir

Archive of SID

ایده‌ی پیشنهادی، حذف گروهی از گره‌ها برای محاسبه‌ی انتشار اثر آن‌ها می‌باشد. به‌طوری‌که الگوریتم بهبودیافته‌ی جدید، با انجام محاسبات کم‌تر به جوابی نزدیک به جواب الگوریتم حریصانه‌ی پایه دست پیدا کند. از طرفی الگوریتم حریصانه بر اساس یک مدل انتشار، گره‌های تأثیرگذارتر را پیدا می‌کند. در الگوریتم جدید که الگوریتم بهبودیافته‌ی مبتنی بر محتوا نام دارد و به اختصار CBIG[1] نامیده می‌شود، از مدل انتشار حدآستانه‌ی توسعه‌یافته که پیش‌تر تشریح گردید، استفاده می‌شود. از طرفی در الگوریتم CBIG، تنها گره‌هایی مورد بررسی قرار می‌گیرند که پتانسیل بیش‌تری برای فعال کردن سایر گره‌ها داشته باشند. به این صورت که ابتدا با در نظر گرفتن الگوهای رفتاری گره‌ها و با توجه به موضوع محتوایی که در شبکه منتشر می‌شود، تعداد گره‌هایی که در اولین گام توسط هر گره v مستقل از فعال بودن سایر گره‌ها فعال می‌شوند، محاسبه شده و مقدار آن در پارامتر γ_v ذخیره می‌شود. با به دست آوردن پارامتر γ_v برای تمامی گره‌های شبکه، الگوریتم حریصانه به صورت شکل ۱ تغییر می‌یابد.

Initialize S=0
for i=1 to k do
  for each vertex v | $v \in V/S$
    if $\gamma_v > \rho$ do
      $s_v = \sigma(S \cup v)$
    end if
end for
$S = S \cup \{\arg\max_{v \in V/S}(s_v)\}$
end for
return S

شکل ۱: الگوریتم حریصانه توسعه‌یافته

طبق شکل ۱، در الگوریتم CBIG تنها گره‌هایی کاندید گره پرنفوذ هستند که مقدار پارامتر γ_v آن‌ها از مقدار ρ که یک عدد طبیعی است، بیش‌تر باشد و در اولین گام بیش از ρ گره را فعال کنند. تعیین مقدار ρ بستگی به شرایط دارد. درواقع مقدار ρ برحسب کاربرد باید به‌گونه‌ای انتخاب شود که تعادلی بین کاهش محاسبات انجام شده و میزان انتشار توسط گره‌های پرنفوذ انتخاب شده توسط الگوریتم بهبودیافته برقرار شود.

۶.۳- ارزیابی

ارزیابی روش پیشنهادی و مقایسه آن با الگوریتم پایه حریصانه (G) براساس معیارهای زیر انجام می‌شود:

- سرعت فعال شدن گره‌ها برحسب گره بر گام
- سرعت دریافت محتوا برحسب گره بر گام
- تعداد گره‌های فعال‌شده در پایان فرایند انتشار
- تعداد پردازش‌های انجام شده توسط الگوریتم G و CBIG

[1] Content-based Improved Greedy

۴- نتایج

ارزیابی روش پیشنهادی با هدف بیشینه‌سازی اثر تبلیغات با موضوعات مختلف در یک گروه ۳۰۵۴ نفری در شبکه اجتماعی پارسی یار [۲۱] انجام شد. همچنین به منظور حصول اطمینان، آزمایش‌هایی بر روی زیرشبکه‌های این مجموعه داده با ۵۰۰ گره و چگالی‌های متفاوت صورت گرفت. برخی از نمادهای به کار رفته در این بخش در جدول ۲ آمده است.

جدول ۲: نمادهای اختصاری استفاده شده در نمودارها

نماد	مشخصات
DS	شبکه با ۳۰۵۴ گره، ۱۹۶۰۲ یال و چگالی ۰,۰۰۴
DS1	شبکه با ۵۰۰ گره، ۱۱۳۲ یال و چگالی ۰,۰۱
DS2	شبکه با ۵۰۰ گره، ۶۶۲ یال و چگالی ۰,۰۰۵
C1	محتوای دو موضوعی با موضوعات c_1 و c_5
C2	محتوای دوموضوعی با موضوعات c_2 و c_3

در این پژوهش دو پارامتر جدید به نام‌های γ_v و ρ به الگوریتم حریصانه‌ی پایه (G) اضافه شدند که با اعمال شرطی در الگوریتم پایه، تعداد محاسبات آن را کاهش می‌دهند. به‌منظور اجرای الگوریتم CBIG با انجام آزمایش‌هایی بهترین مقدار برای ρ برابر با ۴ انتخاب شد. چون هم روند انتشار را بهبود می‌دهد و هم موجب کاهش تعداد محاسبات می‌شود. با توجه به این مقادیر اولیه، نتایج آزمایش‌ها در ادامه می‌آید.

شکل ۲ و جدول ۳ به اختصار نتایج اجرای روش پیشنهادی بر روی مجموعه داده DS1 و DS2 را نشان می‌دهد. دراین سری از آزمایش‌ها توزیع رفتار بین گره‌ها به صورت ۸۰-۱۰-۱۰ در نظر گرفته شده، یعنی ۱۰٪ گره‌ها رفتار نوع ۱، ۱۰٪ رفتار نوع ۲ و ۸۰٪ رفتار نوع ۳ را داشته‌اند. همچنین در شکل ۲ به منظور نمایش بهتر نمودارها مبدأ محورهای افقی به جای صفر مقدار کمینه‌ی آن‌ها در نظر گرفته شده است. مشاهده می‌شود که انتخاب گره‌های پرنفوذ توسط الگوریتم CBIG باعث افزایش سرعت متوسط دریافت محتوا و سرعت متوسط فعال شدن گره‌ها در بیش از ۷۵٪ موارد شده است. طبق جدول ۳، در این سری از آزمایش‌ها تعداد پردازش‌ها به طور متوسط ۸۲,۳۹٪ کاهش یافته است. تعداد گره‌های فعال‌شده در بدترین حالت ۸,۹۷٪ کاهش داشته است. بنابراین با توجه به DS1 و DS2 می‌توان نتیجه گرفت که تغییر الگوریتم حریصانه به میزان زیادی باعث کاهش تعداد محاسبات به گونه‌ای شده است که جواب حاصل از الگوریتم جدید CBIG به طور متوسط ۹۶,۴۲٪ پاسخ الگوریتم G را پوشش می‌دهد.

آن‌چنان‌که پیش‌تر اشاره شد مورد مطالعه این تحقیق یک گروه ۳۰۵۴ نفری با ۱۹۶۰۲رابطه دوستی بین اعضای آن در شبکه‌ی اجتماعی پارسی‌یار است. به همین دلیل آزمایش‌های مشابه روی این مجموعه داده‌ی با ۳۰۵۴ گره انجام شد.

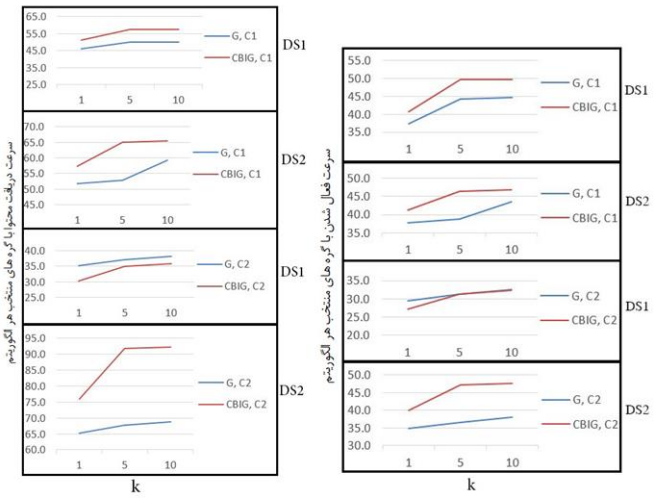
www.SID.ir

Archive of SID

نتایج آزمایش‌ها در شکل ۳ و جدول ۴ خلاصه شده است. ارزیابی نتایج حاصل از اجرای آزمایش‌های مذکور نشان می‌دهد که به طور کلی در ۸۳٪ موارد گره‌های منتخب الگوریتم CBIG باعث افزایش سرعت دریافت و سرعت فعال شدن گره‌ها شده‌اند. در شکل ۳ نیز همانند شکل ۲ به منظور نمایش بهتر و خلاصه‌تر نمودارها مبدأ محورهای افقی از مقادیر نزدیک به مقدار کمینه‌ی نمودارها در نظر گرفته شده است.

جدول ۳: نتایج تعداد پردازش‌ها و گره‌های فعال شده با دو روش CBIG و G در مجموعه داده DS1 و DS2

نوع محتوا و مجموعه داده	K	تعداد پردازش های CBIG	تعداد پردازش های G	درصد کاهش پردازش CBIG به G	تعداد گره‌های فعال شده توسط G	تعداد گره‌های فعال شده توسط CBIG	درصد پوشش تعداد گره های فعال CBIG نسبت به G
DS1, C1	1	14751	204139	92.77%	448	447	99.77%
	5	42912	205997	79.16%	486	447	91.97%
	10	69285	206057	66.37%	491	447	91.03%
DS2, C1	1	20966	194227	89.20%	413	413	100%
	5	20972	194808	89.23%	417	417	100%
	10	30165	195430	60.24%	422	422	100%
DS1, C2	1	24083	163514	99.50%	412	407	98.78%
	5	29376	164140	99.25%	439	408	92.93%
	10	30084	164528	81.71%	456	419	91.88%
DS2, C2	1	19951	82390	75.78%	279	279	100%
	5	19965	83635	79.00%	293	283	96.58%
	10	19973	85028	76.51%	304	286	94.07%



شکل ۲: الف سرعت فعال شدن شکل ۲: ب سرعت دریافت محتوا با دو روش CBIG و G در مجموعه داده DS1 و DS2

طبق جدول ۴، الگوریتم CBIG در شبکه DS باعث شده است تا محاسبات تا حد مطلوبی (به طور میانگین۶۲٫۵۳٪) کاهش یابد و در عین حال تعداد گره‌های فعال شده در پایان فرایند انتشار نیز در هر دو روش اختلاف بسیار ناچیزی داشته باشند. به طور متوسط میزان پوشش روش CBIG نسبت به G، ۹۹٫۲۷٪ است. لازم به ذکر است محاسبه‌ی میزان علاقه‌مندی گره‌ها در واقع جزء مراحل پیش‌پردازش داده‌های موردنیاز برای اجرای روش پیشنهادی و مقایسه‌ی آن با الگوریتم پایه می‌باشد و مدت زمان مذکور بسیار ناچیز و در مقایسه با سایر هزینه‌های زمانی قابل چشم-پوشی است.

در تمامی حالت‌ها روند انتشار توسط گره‌های انتخابی الگوریتم پایه و الگوریتم بهبودیافته نشان می‌دهد که الگوریتم G در مواردی که محدودیت گام روی انتشار اعمال شده باشد، الزاماً جواب بهینه را به عنوان خروجی برنمی‌گرداند و جواب الگوریتم CBIG در گام‌های پایین بهتر است.

۵- نتیجه‌گیری و کارهای آتی

در این مقاله سعی شد تا به مسئله‌ی انتشار در شبکه‌های اجتماعی از منظر جدیدی که تا پیش از این کم‌تر مورد توجه قرار گرفته بود پرداخته شود، و آن متفاوت بودن الگوهای رفتاری کاربران در شبکه‌های اجتماعی و متفاوت بودن علاقه‌مندی افراد به موضوعات مختلف است. بر اساس این دو شرایط جدید مسئله تعریف و مدل انتشار حدآستانه توسعه داده شد. پس از آن با استفاده از مدل انتشار توسعه‌یافته، الگوریتم حریصانه‌ی یافتن گره‌های پرنفوذ در شبکه‌های اجتماعی تغییر کرد. هدف از روش پیشنهادی کاهش محاسبات روش حریصانه و همچنین رسیدن به پاسخی بسیار نزدیک به پاسخ روش حریصانه بود. با در نظر گرفتن مقدار ۴ برای پارامتر ρ، طبق نتایج آزمایش‌ها بر روی مجموعه داده‌های DS1 و DS2 به طور متوسط محاسبات روش پیشنهادی جدید ۸۲٫۳۹ درصد نسبت به روش حریصانه پایه کاهش و ۹۶٫۴۲ درصد پوشش پاسخ حاصل شد. این ارقام برای مجموعه داده DS به ترتیب ۶۲٫۵۳ و ۹۹٫۲۷ درصد به دست آمد.

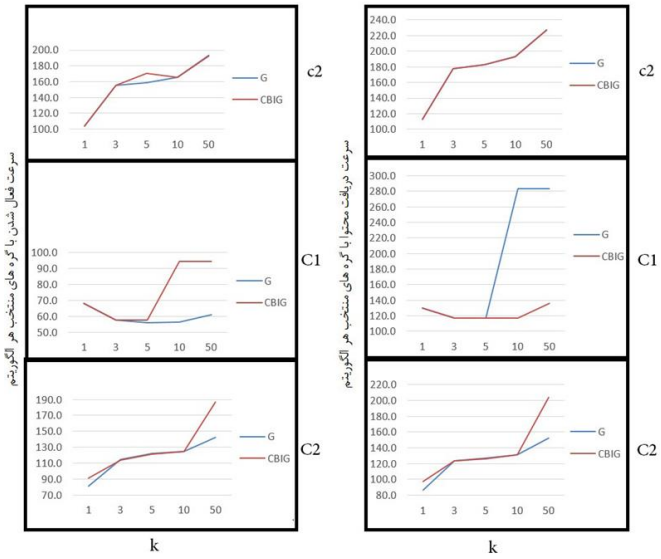
با رویکرد جدیدی که در این مقاله نسبت به مسئله‌ی بیشینه‌سازی اثر مطرح شد، می‌توان تحقیقات دیگری در همین راستا در جهت توسعه و بهبود روش پیشنهادی انجام داد، از جمله: ۱) تغییر تدریجی مقدار ρ به ازای هر k. عملکرد الگوریتم پیشنهادی وابسته به پارامتر ρ می‌باشد، برای بهبود عملکرد می‌توان الگوریتم پیشنهادی را طوری تغییر داد که برای مقادیر اولیه‌ی k پارامتر ρ عدد بزرگ‌تری باشد و به ازای مقادیر بعدی k مقدار پارامتر ρ کاهش یابد. ۲) اعمال محدودیت گام و یافتن گره‌های پرنفوذ به ازای هر گام. هر گام معادل یک واحد زمانی است و در بحث تبلیغات زمان بسیار حائز اهمیت است. ۳) توسعه روش مبتنی بر دیگر مدل‌های انتشار و ۴) مقایسه روش پیشنهادی با دیگر روش‌های پیدا کردن k گره پرنفوذ.

جدول ۴: نتایج تعداد پردازش‌ها و گره‌های فعال شده با دو روش CBIG و G در مجموعه داده DS

نوع محتوا	K	تعداد پردازش های CBIG	تعداد پردازش های G	درصد کاهش پردازش CBIG به G	تعداد گره‌های فعال شده توسط G	تعداد گره‌های فعال شده توسط CBIG	درصد پوشش تعداد گره های فعال CBIG نسبت به G
c2	1	1018600	2586262	60.61%	1451	1451	100%
	3	1217465	3082346	60.50%	2174	2172	99.90%
	5	1218489	3091701	60.58%	2217	2216	99.95%
	10	1600531	3111678	48.56%	2315	2315	100%
	50	1622771	3175585	48.89%	2706	2689	99.37%
C1	1	8573	62701	86.32%	408	408	100%
	3	26402	318597	91.71%	1507	1503	99.73%
	5	44794	321865	86.08%	1493	1493	100%
	10	69280	329833	78.99%	1522	1512	99.34%
	50	70376	391757	82.03%	1648	1512	91.74%
C2	1	1518110	3040005	50.60%	1544	1543	99.93%
	3	1896594	3751463	49.44%	2513	2512	99.96%
	5	1910239	3784415	49.52%	2695	2675	99.25%
	10	2137476	3803618	43.80%	2898	2898	100%
	50	2269787	3808199	40.39%	2986	2981	99.83%

www.SID.ir

Archive of SID



شکل ۳: الف سرعت دریافت محتوا شکل ۲: ب سرعت فعال شدن با دو روش CBIG و G در مجموعه داده DS

مراجع

[1] Narayanam, R. and Y. Narahari, *A shapley value-based approach to discover influential nodes in social networks.* Automation Science and Engineering, IEEE Transactions on, 2011. 8(1): p. 130-147.

[2] Kempe, D., J. Kleinberg, and É. Tardos. *Maximizing the spread of influence through a social network*. in *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*. 2003.

[3] Kiss, C. and M. Bichler, *Identification of influencers—measuring influence in customer networks.* Decision Support Systems, 2008. 46(1): p. 233-253.

[4] Chen, W., et al. *Influence Maximization in Social Networks When Negative Opinions May Emerge and Propagate*. in *SDM*. 2011.

[5] Zhou, J., Y. Zhang, and J. Cheng, *Preference-based mining of top-K influential nodes in social networks.* Future Generation Computer Systems, 2014. 31: p. 40-47.

[6] Chen, W., Y. Wang, and S. Yang. *Efficient influence maximization in social networks*. in *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*. 2009.

[7] Domingos, P. and M. Richardson. *Mining the network value of customers*. in *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*. 2001.

[8] Saito, K., et al ,.*Learning continuous-time information diffusion model for social behavioral data analysis*, in *Advances in Machine Learning*. 2009, Springer. p. 322-337.

[9] Estevez, P., P. Vera, and K. Saito. *Selecting the most influential nodes in social networks*. in *Neural Networks, 2007. IJCNN 2007. International Joint Conference on*. 2007.

[10] Ma, H., et al. *Mining social networks using heat diffusion processes for marketing candidates selection*. in *Proceedings of the 17th ACM conference on Information and knowledge management*. 2008.

[11] Chen, N., *On the approximability of influence in social networks.* SIAM Journal on Discrete Mathematics, 2009. 23(3): p. 1400-1415.

[12] Wang, Y., et al. *Community-based greedy algorithm for mining top-k influential nodes in mobile social networks*. in *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*. 2010.

[13] Narayanam, R. and Y. Narahari, *A shapley value-based approach to discover influential nodes in social networks.* IEEE Transactions on Automation Science and Engineering, 2010(99): p. 1-18.

[14] Zhang, Y., Z. Wang, and C. Xia. *Identifying key users for targeted marketing by mining online social network*. in *Advanced Information Networking and Applications Workshops (WAINA), 2010 IEEE 24th International Conference on*. 2010.

[15] Chen, Y.-C., W.-C. Peng, and S.-Y. Lee, *Efficient algorithms for influence maximization in social networks.* Knowledge and information systems, 2012. 33(3): p. 577-601.

[16] Li, Y.-M. and C.-Y .Lai. *A Diffusing Path Planning Mechanism for Marketing Information Propagation over Social Media*. in *System Sciences (HICSS), 2013 46th Hawaii International Conference on*. 2013.

[17] Mitra, P. and K. Baid. *Targeted advertising for online social networks*. in *Networked Digital Technologies, 2009. NDT'09. First International Conference on*. 2009.

[18] Yang, W.-S., et al. *Mining social networks for targeted advertising*. in *System Sciences, 2006. HICSS'06. Proceedings of the 39th Annual Hawaii International Conference on*. 2006.

[19] Frias-Martinez, E., S.Y. Chen, and X. Liu, *Survey of data mining approaches to user modeling for adaptive hypermedia.* Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on, 2006. 36 :(٦)p. 734-749.

[20] Halkidi, M., Y. Batistakis, and M. Vazirgiannis, *On clustering validation techniques.* Journal of intelligent information systems, 2001. 17(2-3): p. 107-145.

[21] Esmaeili, L., M. Nasiri, and B. Minaei-Bidgoli, *Analyzing Persian social networks: an empirical study.* IGI global, 2013: p. 167-172.

www.SID.ir

Archive of SID

www.SID.ir