Archive of SID

چارچوبی برای استخراج شبکه و تحلیل شبکه‌های اجتماعی

فاطمه واشقانی فراهانی[۱]، لیلا اسماعیلی[۲]، سید علیرضا هاشمی گلپایگانی[۳*]، زهرا عبادی ابوذر[۴]

[۱] کارشناسی ارشد مهندسی فناوری اطلاعات، دانشگاه صنعتی امیرکبیر، تهران، ایران، f.far@aut.ac.ir

[۲] دانشجوی دکتری مهندسی فناوری اطلاعات، دانشگاه صنعتی امیرکبیر، تهران، ایران، leila.esmaeili@aut.ac.ir

[۳] عضو هیات علمی و استادیار دانشکده‌ی مهندسی کامپیوتر و فناوری اطلاعات، دانشگاه صنعتی امیرکبیر، تهران، ایران، sa.hashemi@aut.ac.ir

[۴] کارشناس ارشد مهندسی فناوری اطلاعات، دانشگاه صنعتی امیرکبیر، تهران، ایران، ebadi@aut.ac.ir

چکیده

تحلیل شبکه‌های اجتماعی یکی از حوزه‌های تحقیقاتی پرمخاطب برای محققان، جامعه‌شناسان و دانشگاهیان در دهه اخیر شناخته شده است. با توجه به کاربرد این روش تحلیل در علوم مختلف، تاکنون چارچوبی مشخص، ساده و قابل درک برای محققان با پیش‌زمینه‌های علمی متفاوت و بعضا غیرفنی جهت تعیین فرایند استخراج و تحلیل شبکه‌ای ارائه نشده است. از این رو، در این مقاله یک چارچوب برای استخراج شبکه و تحلیل شبکه‌های اجتماعی ارائه شده و چارچوب پیشنهادی بر روی یکی از شبکه‌های اجتماعی ملّیِ فارسی زبان اجرا شده است. چارچوب ارائه شده در این مقاله می‌تواند برای تحلیل دیگر ساختارهای قابل نگاشت به شبکه در حوزه‌های مختلف از جمله تجارت، روانشناسی، جامعه‌شناسی، بیوتکنولوژی و غیره دارد به کار گرفته شود.

واژه‌های کلیدی

شبکه‌های اجتماعی برخط، تحلیل شبکه اجتماعی، استخراج شبکه، معیارهای مرکزیت

۱- مقدمه

دهه اخیر به دلیل گسترش کاربردهای تعاملی مبتنی بر اینترنت از جمله شبکه‌های اجتماعی برخط و شبکه‌های اجتماعی پیام‌رسان با قابلیت استفاده در تلفن‌های همراه و تبلت‌ها می‌تواند به دهه «انفجار شبکه‌های اجتماعی» نام‌گذاری شود. نخستین بار مبحثی با عنوان شبکه‌های اجتماعی در سال ۱۹۶۰ در دانشگاه ایلی نیوز در ایالت متحده امریکا مطرح شد و بعد از آن در سال ۱۹۹۷، نخستین سایت شبکه اجتماعی به نشانی اینترنتی SixDegrees.com شروع به کار کرد [۳]. بر اساس آمار سال ۲۰۱۴، تعداد شبکه‌های اجتماعی در جهان بالغ بر چند صد میلیون شده است که از این بین facebook با ۱,۳۵ بیلیون کاربر، پرجمعیت‌ترین شبکه اجتماعی به شمار می‌رود. Qzone، Google+، LinkedIn، Instagram، Twitter و Tumblr در رتبه‌های بعدی ایستاده‌اند [۴]. طبق آمار پیش‌بینی می‌شود کاربران شبکه‌های اجتماعی در سال ۲۰۱۸، بالغ بر ۲,۵۵ بیلیون نفر باشند [۵].

در ایران نیز، اولین شبکه اجتماعی ملی تقریبا ۱۲ سال پیش با نام Cloob افتتاح شد. به طور رسمی از رشد سایت‌های شبکه اجتماعی فارسی‌زبان، کاربران، ویژگی‌ها و فعالیت‌های آن‌ها آماری در دست نیست. اما در سال ۲۰۰۹، درحالی‌که از میان ۲۰ وبگاه برتر و پربازدید جهانی، ۸ وبگاه جز شبکه‌های اجتماعی محسوب می‌شدند، در ایران این آمار، ۱۳ وبگاه شبکه اجتماعی در بین ۲۰ وبگاه پربازدید بوده است که شماری از این ۱۳ وبگاه فارسی‌زبان و ملی بوده‌اند [۱]. همچنین طبق آمار سال ۲۰۱۵ از [۶] پرمخاطب‌ترین شبکه‌های اجتماعی در ایران به ترتیب Facebook، Google+، Twitter و Tumblr بوده است که البته مخاطبین فیسبوک قابل مقایسه با سه شبکه دیگر نیست.

تحلیل شبکه‌های اجتماعی و کاوش در داده‌های ذخیره‌شده در پایگاه‌داده‌های این شبکه‌ها با مقاصد علمی، اجتماعی، سیاسی، تجاری، بازاریابی، کسب و کار و ... در دو سطح ملی و بین‌المللی از سال ۱۹۹۸ مطرح شد و در دهه اخیر رو به گسترش است؛ کشورها، دانشگاه‌ها و موسسات علمی-پژوهشی مختلف با انگیزه‌های متفاوت در این مقوله سرمایه‌گذاری کرده‌اند. تحلیل شبکه‌های اجتماعی، روابط اجتماعی را از نگاه نظریه شبکه و گراف می‌بینند و در آن، هر رابطه متشکل از گره و یال می‌باشد. گره‌ها بازیگران منفرد در داخل شبکه و یال‌ها رابطه بین گره‌ها هستند. انواع ارتباطات بین گره‌ها وجود دارد. به بیان دیگر، تحلیل شبکه اجتماعی عبارت است از نگاشت و اندازه‌گیری روابط و همکاری‌ها در بین افراد، گروه‌ها، سازمان‌ها و هر موجودیتی که قابلیت پردازش اطلاعات و دانش را داشته باشد [۷].

تاکنون، تحلیل شبکه‌های اجتماعی فارسی‌زبان عمدتا بر روی شبکه‌های بلاگی مانند «پرشین‌بلاگ» و «پارسی‌بلاگ» متمرکز بوده است.

* نویسنده مسئول

www.SID.ir

Archive of SID

این پژوهش‌ها بر اساس مجموعه‌داده‌های از پیش آماده‌شده توسط محققین دیگر، داده‌های ساختاری حاصل از اجرای خزش‌گرها و مجموعه‌داده حاصل از داده‌های ذخیره‌شده در پایگاه‌داده بوده‌اند؛ همچنین به دلیل محدودیت‌های پردازشی، حجم کوچکی از داده‌ها تحلیل شده است [۸] [۹] [۱۰]. در پژوهشی دیگر، داده‌های ذخیره‌شده در پایگاه داده شبکه اجتماعی فارسی‌زبان «پارسی‌یار[۱]» برای شخصی‌سازی گروه‌های پیشنهادی به کاربران شبکه اجتماعی مورد استفاده قرار گرفته است [۱۱]. به دلیل استفاده از داده‌های خام در این پژوهش، تجربیات مرتبط با پیش‌پردازش و آماده‌سازی این مجموعه داده در [۱۲] آمده است. این در حالی است که بخش اعظمی از تحقیق‌ها و پروژه‌هایی که در حوزه تحلیل شبکه‌های اجتماعی در ایران انجام می‌شود بر روی شبکه‌های اجتماعی بین‌المللی بوده و یا در بعضی موارد با توجه به هدف محقق، روش‌های ساخت تصادفی گراف و یا DOE استفاده شده‌اند [۱۳].

طبق مطالعات انجام شده توسط نویسندگان، تحقیقات حوزه استخراج و تحلیل شبکه‌های اجتماعی (SNA) تاکنون فاقد یک چارچوب ساختارمند که فرایند استخراج و تحلیل شبکه‌ای از انواع داده‌های مختلف را فراهم نماید، می‌باشد. با توجه به کاربرد SNA در علوم مختلف، چارچوبی که برای استفاده محققان در علوم مختلف قابل درک باشد، ضروری است. همچنین، از آن جا که شبکه‌های اجتماعی برخط نمودی از جوامع واقعی هستند و با توجه به اختلاف‌های فرهنگی، مذهبی، اجتماعی و غیره، تعمیم برخی از نتایج حاصل از تحلیل شبکه‌های اجتماعی بین‌المللی در سطح ملی، تصمیم‌گیری و سیاست‌گذاری‌ها بر اساس آن، منطقی و صحیح نیست. چرا که ملیت‌های مختلف علاوه بر داشتن زبان متفاوت، فرهنگ و در نتیجه رفتار متفاوتی نسبت به یکدیگر دارا هستند که همین رفتار در نحوه شکل‌گیری تعاملات و روابط آن‌ها تاثیرگذار است [۲]. از این رو، اهداف اصلی این مقاله عبارتند از:

- ارائه چارچوبی جامع به منظور تحلیل شبکه‌های اجتماعی
- تحلیل یک شبکه اجتماعی ملی فارسی زبان بر اساس چارچوب پیشنهادی ارائه شده

برای تحقق هدف دوم از شبکه اجتماعی «پارسی‌یار» استفاده می‌شود و ابزار Gephi به منظور انجام تحلیل‌ها مورد استفاده قرار می‌گیرد.

در ادامه، بخش‌های مختلف مقاله به صورت زیر سازماندهی شده است: مفاهیم پایه در تحلیل شبکه‌های اجتماعی در بخش ۲ آمده است، بخش ۳ به ارائه چارچوب تحلیل پیشنهادی می‌پردازد. اجرای چارچوب تحلیل و نتایج حاصل از آن در بخش ۴ آمده است. بخش پایانی به نتیجه‌گیری و بیان کارهای آتی اختصاص یافته است.

۲- مفاهیم پایه

تحلیل شبکه‌های اجتماعی از سال ۱۹۳۰ توسط گروه‌های متعددی در زمینه‌های مختلف از جمله روان‌شناسی، مردم‌شناسی و ریاضیات به طور مستقل، توسعه یافته است [۱۴]. بعد از آن در دهه ۱۹۶۰، «جهان کوچک» توسط میلگرام که بعدها تحت عنوان «شش درجه جدایی» محبوبیت یافت، مطرح شد. این عبارت تعداد روابط (پیوند) مورد نیاز برای ارتباط هر دو نفر را بیان می‌کند [۱۵]. این حوزه نسبتا تا سال ۱۹۹۸ ناشناخته باقی ماند. در سال ۱۹۹۸، وات و استروگاتز پدیده جهان کوچک را مجددا احیا کردند و از آن پس، مقالات بسیاری در حوزه شبکه‌های اجتماعی نوشته شد و این زمینه دوباره متولد شد [۱۵]. امروزه، تحلیل شبکه‌های اجتماعی کاربردهای بسیاری در حوزه‌های مختلف علمی و کاربردی از جمله مقابله با پول‌شویی و تروریسم پیدا کرده است.

تحلیل شبکه‌های اجتماعی، روابط اجتماعی و ساختارهای شبکه‌ای را از نگاه نظریه گراف می‌بینند. از این منظر، گره‌ها بازنمای بازیگران فردی در داخل شبکه و یال‌ها بازنمای ارتباطاتی چون دوستی، سازمانی، خویشاوندی و ... بین افراد هستند. در هر زمینه کاربردی، اصطلاحات معادلی به جای گره و یال استفاده می‌شود که هر یک درست بوده و به جای یکدیگر قابل استفاده هستند. به عنوان مثال در زمینه علم کامپیوتر، یک شبکه مجموعه‌ای از اشیاء است به نام گره‌ها که جفت‌های مشخصی از این اشیاء توسط اتصلاتی با نام لبه‌ها متصل شده‌اند (فصل ۲ از [۱۶]). بنابراین چنانچه یک شبکه اجتماعی یک ساختار اجتماعی بین بازیگران باشد، می‌توان تحلیل شبکه‌های اجتماعی را، نگاشت واقعیت و اندازه‌گیری روابط و جریان‌ها در بین افراد، گروه‌ها، سازمان‌ها، کامپیوترها یا هر موجودیتی که قابلیت پردازش اطلاعات/دانش را داشته باشد، تعریف کرد [۱۷].

۲-۱- انواع شبکه

ساده‌ترین ساختار شبکه متشکل از گره‌ها برای نمایش یک نوع گره و متشکل از یال‌ها برای نمایش یک نوع از رابطه یا اتصال است که می‌توان آن را به صورت یک گراف یا ماتریس نمایش داد؛ مثلا شبکه حاصل از روابط دوستی بین افراد شامل گره‌های فردی و روابط دوستی می‌باشد. اما در شبکه‌های واقعی ممکن است تنها یک نوع رابطه بین بازیگران وجود نداشته باشد و حتی روابط از یک درجه اهمیت یا وزن برخوردار نباشند. همچنین ممکن است تمام گره‌ها، از یک نوع مانند فرد، کامپیوتر، سازمان و ... نباشند. بنابراین می‌توان شبکه را از دو منظر، مبتنی بر ویژگی‌های گره و مبتنی بر ویژگی‌های رابطه تقسیم‌بندی نمود. انواع شبکه در جدول ۱ و ۲ خلاصه شده است که بر اساس تقاطع آن‌ها انواع شبکه قابل تعریف است.

۲-۲- ابعاد تحلیل شبکه‌های اجتماعی

با توجه به طیف مسائل مختلف در تحلیل شبکه‌های اجتماعی می‌توان سه بعد اصلی برای تحلیل شبکه‌ای تعریف کرد؛ که هر مسئله در تحلیل شبکه‌های اجتماعی بر اساس یک یا چند بعد قابل تعریف است [۱۸]. بعد ۱ یا بعد ساختار لایه‌ای همه روابط بین بازیگران در شبکه را توضیح می‌دهد. روابط ممکن است بر اساس ارتباط مستقیم بین افراد از طریق تلفن یا پیامک باشد، یا به صورت ضمنی از فعالیت‌های افراد در سیستم‌های مختلف حاصل شود مانند به اشتراک‌گذاری یا انجام همکاری در یک پروژه.

[۱] در سال‌های اخیر این شبکه اجتماعی به طور مستقل در دسترس نبوده و اطلاعات آن به سرویس وبلاگ «پارسی‌بلاگ» انتقال یافته و با آن ادغام شده است.

www.SID.ir

Archive of SID

بعد ۲ یا بعد زمان، تحلیل زمانی شبکه‌های اجتماعی را ممکن می‌سازد. در واقع به دلیل وجود یک پنجره زمانی که یک مدت زمان با اندازه مشخصی است این امکان فراهم می‌شود. تحلیل ممکن است بر اساس یک تصویر لحظه‌ای در یک زمان مفروض انجام شود که در نتیجه روابط در آن لحظه وجود دارند؛ یا امکان دارد بر اساس روابط ایجاد شده مبتنی بر فعالیت‌های افراد در مدت زمان مشخص انجام شود که در نتیجه متناسب با اندازه مدت زمان، تحلیل شبکه‌ای می‌تواند محدود شود. بعد ۳ بعد اجتماع یا گروه است. یک گروه، یک اجتماع است که افراد در داخل آن اغلب همکاری (روابط) بیشتری نسبت به خارج از آن دارند. بنابراین گروه‌ها از داخل به طور قوی‌تری به یکدیگر متصل هستند و ارتباطات ضعیف‌تری در خارج با دیگر گروه‌ها دارند.

جدول ۱: انواع شبکه مبتنی بر ویژگی‌های گره

نام شبکه	توضیح
شبکه‌های یک حالته	شبکه‌هایی که یک مجموعه از گره‌ها در آن وجود داشته و گره‌ها به یکدیگر شبیه هستند. برای مثال همه آن‌ها کارمندان یک سازمان یا شرکت‌کنندگان در یک همایش هستند [۱۹].
شبکه‌های دو/چند حالته	روابط بین دو مجموعه از گره‌ها تعریف می‌شود. به عنوان مثال دانشجویان و دانشگاهی که در آن تحصیل می‌کنند. اغلب چنین داده‌ای به داده «وابستگی» یا «actor-by-event» مصطلح می‌باشد [۱۹] و فصل ۱۷ از [۲۰].

جدول ۲: انواع شبکه مبتنی بر ویژگی‌های رابطه

نام شبکه	توضیح
شبکه‌های ساده / ترکیبی	شبکه‌هایی که دارای یک نوع رابطه (مانند دوستی، خویشاوندی، مبادلات مالی و ...) میان بازیگران هستند، شبکه‌های ساده نامیده می‌شود؛ و شبکه‌هایی که به طور همزمان، بیشتر از یک نوع ارتباط بین بازیگران‌شان وجود دارد شبکه‌های ترکیبی یا شبکه‌های با روابط نامتجانس خوانده می‌شوند.
شبکه‌های جهت‌دار / بدون جهت	هر رابطه ممکن است جهت‌دار باشد (با یک بازیگر مبدا آغاز می‌شود و به یک بازیگر هدف می‌رسد)، یا ممکن است رابطه نشان‌دهنده یک هم‌رخداد، حضور مشترک یا یک رابطه قراردادی (عهد و پیمان) بین جفت بازیگران باشد که چنین رابطه‌ای بدون جهت است. روابط جهت‌دار، با فلش در گراف نشان داده می‌شوند و ممکن است متقابل باشند (A، B را انتخاب کند و B نیز A را انتخاب نماید)، چنین روابطی به صورت فلش دو طرفه مشخص می‌شوند (فصل ۳ از [۲۰]).
شبکه‌های وزن‌دار / بدون وزن	روابط ممکن است دارای وزن و قدرت متفاوت باشند. قدرت و وزن رابطه در یک گراف می‌تواند اسمی یا دودویی (نشان‌دهنده وجود یا عدم وجود رابطه) باشد یا به صورت یک عدد علامت‌دار (نشان‌دهنده یک رابطه منفی، یک رابطه مثبت یا بدون رابطه) نشان داده شود. همچنین می‌تواند با یک عدد ترتیبی (نشان‌دهنده میزان قدرت)، یا به صورت ارزش‌گذاری (میزان فاصله یا سطح نسبی) مشخص شود (فصل ۳ از [۲۰]).

۲-۳- ساختار شبکه اجتماعی

ساختار یک شبکه اجتماعی با یک گراف یا شبکه ساده به عنوان یک چندتایی <V,E> یا یک جفت $SN := (V, E)$ تعریف می‌شود که:

- V: یک مجموعه ناتهی از بازیگرانی است که موجودیت‌ها را بازنمایی می‌کند، مانند انسان‌ها، سازمان‌ها و نهادها و ...؛ همچنین به آن گره، راس، عضو و عامل نیز گفته می‌شود.
- E: یک مجموعه از یال‌ها است (روابط بین بازیگران که به آن‌ها کمان، اتصال یا لبه نیز گفته می‌شود). یک یال توسط یک چندتایی <x,y> یا $e=(x,y)$ بازنمایی می‌شود که $x, y \in V$ و $x \neq y$ است.
- اگر یال‌ها جهت‌دار باشند برای هر دو یال مانند <x2,y2> و <x1,y1> اگر $< x1, y1 > \neq < x2, y2 >$ و $x1 \neq x2$ باشد، آن‌گاه $y1 \neq y2$ است.

۲-۴- شاخص‌های تحلیل شبکه اجتماعی

میشل شبکه‌های اجتماعی را ارتباط میان افراد در یک گروه، شامل ویژگی‌های افراد و روابط مابین آن‌ها تعریف می‌نماید [۷]. تحلیل شبکه اجتماعی می‌تواند تاثیر روابط بین افراد و سازمان‌ها را جستجو و بهره‌برداری کند و تکنیکی مفید برای تحلیل اطلاعات یا منابع تبادل شده بین افراد، گروه‌ها و سازمان‌ها باشد. زیرا تمرکز آن به جای تحلیل خصوصیات قطعی افراد عضو یک گروه، بر روی معیارهای ساختاری است.

معیارها و شاخص‌های تحلیل شبکه‌های اجتماعی متناسب با نوع شبکه بسیار متنوع هستند. در این مقاله برای بررسی و تحلیل روابط میان کاربران و میزان شدت آن‌ها از شاخص‌های مرکزیت درجه، درجه ورودی، بینابینی، Embeddedness، n-clique و وزن یال‌ها استفاده می‌شود. در ادامه تعریف مختصری از این شاخص‌ها در شبکه‌های ساده، یک حالته، بدون جهت و بدون وزن ارائه می‌شود.

اندازه شبکه: شبکه را با گراف G=<V,E> در نظر گرفته که در آن V مجموعه گره‌ها و E مجموعه یال‌ها می‌باشد. اندازه شبکه شامل تعداد گره‌ها |V| و تعداد روابط |E| موجود بین آنها است (البته تعداد راس‌ها به تنهایی اندازه شبکه را نشان نمی‌دهد) [۷].

مرکزیت[1]: مجموعه‌ای از شاخص‌ها است که اهمیت گره‌ها در ساختار شبکه را مشخص می‌کنند. برخی از مهمترین آن‌ها عبارتند از:

مرکزیت درجه[2]: مرکزیت درجه یک گره به تعداد همسایه‌های آن گره وابسته است و به بقیه گره‌ها در شبکه ارتباطی ندارد. هر چه میزان درجه یک گره بیش‌تر باشد، دسترسی آن به منابع بیش‌تر بوده و مرکزی‌تر محسوب می‌شود [۷]. در گراف غیر جهت‌دار G:=(V, E) با |V| راس و |E| یال، مرکزیت درجه راس v_i مطابق فرمول (۱) تعریف می‌شود [۲۱]:

$$C'_D(v_i) = \frac{d(v_i)}{|V|-1} \qquad (1)$$

[1] Centrality
[2] Degree Centrality

www.SID.ir

Archive of SID

در این فرمول اندیس D به معنی درجه است و $d(v_i)$ تعداد راس‌هایی را نشان می‌دهد که به راس v_i متصل هستند.

درجه ورودی[1] یا $\deg^-_{(v)}$: در شبکه‌های جهت‌دار، درجه ورودی تعداد یال‌های جهت‌دار است که به گره وارد می‌شوند [۷].

درجه خروجی[2] یا $\deg^+_{(v)}$: در شبکه‌های جهت‌دار، برابر با تعداد یال‌های جهت‌دار است که از گره خارج می‌شوند [۷].

مرکزیت بینابینی[3]: این مرکزیت تعداد دفعاتی را که یک گره بین کوتاه‌ترین مسیر بین دو گره دیگر قرار دارد، نشان می‌دهد. همچنین معیاری است برای تعیین میزان کنترل یک فرد در ارتباطات بین افراد دیگر در شبکه اجتماعی [۷]. بینابینی راس v در فرمول (۲) نشان داده شده است [۲۲]:

$$C_B(v) = \sum_{s\neq v\neq t\epsilon V} \frac{\sigma_{st}(v)}{\sigma_{st}} \qquad (۲)$$

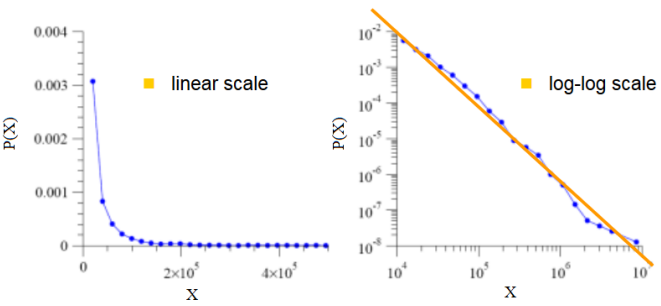
در این فرمول σ_{st} تعداد کل کوتاه‌ترین مسیرها از نود s تا نود t است و $\sigma_{st}(v)$ تعداد مسیرهایی است که از v می‌گذرد.

Embeddedness: تعداد همسایگان مشترک دو راس را بوسیله‌ی این شاخص می‌توان یافت. این شاخص از حاصل تقسیم اندازه مجموعه حاصل از اشتراک همسایگان دو راس A و B بر اندازه مجموعه حاصل از اجتماع همسایگان دو راس A و B بدست می‌آید [۷].

n-clique: یک n-clique گراف کاملی است که شامل n راس است [۲۳].

۲-۵- شبکه‌های scale-free

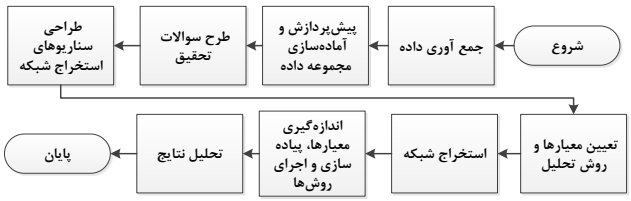
شبکه‌هایی که توزیع درجات آن‌ها از توزیع power-law تبعیت می‌کنند را شبکه‌های scale-free می‌نامند [۸]. در توزیع power-law چولگی به سمت راست است و توزیع نامتقارن است. همچنین نمودار log-log این توزیع به صورت یک خط راست نمایش داده می‌شود. شکل ۱ نمودار توزیع Power-Law را نشان می‌دهد.



شکل ۱: نمودار توزیع Power-Law

۳- چارچوب پیشنهادی تحلیل شبکه‌های اجتماعی

چارچوب پیشنهادی تحلیل شبکه‌های اجتماعی طبق شکل ۲ دارای هشت مرحله اصلی است.



شکل ۲: چارچوب پیشنهادی برای تحلیل شبکه‌های اجتماعی

ابتدا مجموعه داده مورد نظر به یکی از روش‌های مرسوم جمع‌آوری شده [۲] و سپس به منظور کیفیت بخشیدن به داده‌ها و پاکسازی آن، پیش پردازش انجام می‌شود. پس از آماده شدن مجموعه داده، سوالات تحقیق طراحی و با استفاده از سناریوهای تعریف شده به استخراج شبکه‌های مرتبط با هر سناریو پرداخته می‌شود. هر سناریو پاسخگوی یک یا چند سوال است. پس از استخراج شبکه بر اساس سناریوی تعریف شده، با اندازه‌گیری معیارها و شاخص‌ها، پاسخ سوالات تهیه می‌شود.

استفاده از کامپیوترها در زمینه تحلیل شبکه‌های اجتماعی یکی از مزایای مهم به شمار می‌رود. امروزه ابزارهای کامپیوتری بسیاری به منظور بصری‌سازی، شبیه‌سازی و تحلیل ساختارهای شبکه‌ای توسعه یافته‌اند (فصل ۲ از [۲۰]). از جمله می‌توان به Gephi، NodeXL، NetLogo، Pajek، iGraph، R، UCINET و ... اشاره کرد.

۴- نتایج اجرای چارچوب پیشنهادی

۴-۱- جمع‌آوری، پیش‌پردازش و آماده‌سازی مجموعه داده

مجموعه داده این تحقیق، شبکه اجتماعی پارسی یار، برگرفته از کار پژوهشی [۱۲] است، از این رو، مجموعه داده از قبل جمع‌آوری، پیش پردازش و آماده شده است. شبکه اجتماعی پارسی‌یار یک شبکه اجتماعی ملی و متعلق به سرویس وبلاگ‌نویسی پارسی‌بلاگ می‌باشد که در حال حاضر غیرفعال شده است و بر خلاف اکثریت شبکه‌های اجتماعی، بیش‌تر کاربران آن جوان، دانشجو و مرد می‌باشند. این شبکه شامل ۷۸۴۶۷ کاربر و ۳۳۵۹ گروه است. بیشتر کاربران این شبکه (بجز کاربرانی که مکان خود را مشخص نکرده‌اند) اهل استان‌های تهران، اصفهان و خراسان هستند. در این مقاله، تمرکز بر روی تحلیل روابط میان کاربران است؛ هر کاربر در این مجموعه داده دارای کد شناسایی منحصربفردی است.

۴-۲- طرح سوالات تحقیق

مجموعه داده، متشکل از کاربران شبکه اجتماعی و کلیه روابط مربوط به آنها است. بنابراین سوالات تحقیق با هدف تحلیل روابط بین کاربران شبکه اجتماعی عبارتند از:

- Q1: کدام کاربران در شبکه بیشترین تبادل پیام را دارند؟
- Q2: کدام کاربر بیش‌ترین تعداد درخواست دوستی را دریافت کرده است؟
- Q3: کدام کاربران نقش کلیدی‌تری در برقراری ارتباط با بقیه کاربران دارند؟
- Q4: کدام کاربران، دوستان مشترک بیشتری دارند؟

[1] Indegree
[2] Outdegree
[3] Betweenness Centrality

www.SID.ir

Archive of SID

- Q5: بیشترین تعداد افرادی که در شبکه دوستی با هم دوست هستند، چقدر است؟
- Q6: کدام کاربران تمایلات محتوایی شبیه‌تری دارند، به عبارت دیگر کدام کاربران در پست‌های مشترک بیشتری نظر ثبت کرده‌اند؟

۴-۳- طراحی سناریوهای استخراج شبکه

استخراج یک شبکه اجتماعی به معنای تشخیص و شناسایی روابط میان موجودیت‌ها است. یک شبکه اجتماعی در دو مرحله استخراج می‌شود. ابتدا باید مجموعه گره‌های شبکه شناسایی شوند و سپس یال‌های شبکه یا طریقه اتصال گره‌ها تعریف شوند.

در این مقاله، کاربرانی که در شبکه عضو هستند بعنوان گره در نظر می‌شوند و با توجه به سوالات تحقیق، یال‌های شبکه مشخص خواهند شد. از این رو، سناریوهای استخراج شبکه در این مقاله برای مجموعه داده پارسی‌یار طبق جدول ۳ مشخص می‌شود.

جدول ۳: سناریوهای استخراج شبکه و ویژگی‌های آنها

کد	نام	گره	یال	وزن	جهت	سوال
S1	شبکه تبادل پیام	فرد	تبادل پیام میان افراد	دارد	دارد	Q1
S2	شبکه تبادل درخواست دوستی	فرد	تبادل درخواست دوستی میان افراد	دارد	دارد	Q2
S3	شبکه روابط دوستی	فرد	وجود رابطه دوستی میان افراد	دارد	ندارد	Q3 Q4 Q5
S4	شبکه نظرهای افراد	فرد	نظردهی بر روی پست مشترک	دارد	ندارد	Q6

۴-۴- تعیین معیارها و روش تحلیل

برای پاسخ به سوالات تحقیق در بخش ۴-۲ و بر اساس سناریوهای مشخص شده در بخش ۴-۳، یک یا مجموعه‌ای از معیارها و شاخص‌های تحلیل شبکه‌های اجتماعی انتخاب می‌شود. برای پاسخگویی به سوالات از ترکیبی از شاخص‌های تعریف شده در بخش ۴-۲ استفاده شده است که جزئیات در جدول ۴ آمده است.

۴-۵- استخراج شبکه

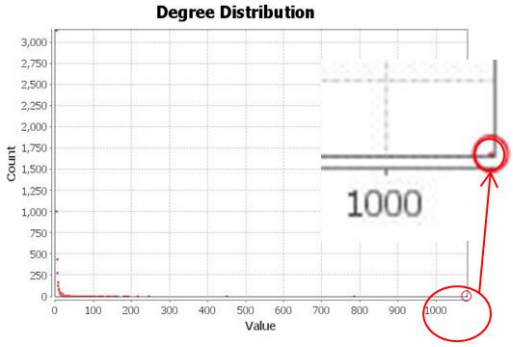
در این مرحله چهار شبکه‌ای که سناریوی استخراج آنها در بخش ۴-۳ در جدول ۳ بیان شد بر اساس مجموعه‌ای از کوئری‌ها از مجموعه داده اصلی استخراج می‌شوند. خروجی کوئری‌ها به طور حداقل شامل رئوس هر شبکه و یال‌ها (دو راس ابتدا و انتهای هر یال) می‌باشد.

۴-۶- اندازه‌گیری معیارها، پیاده‌سازی، اجرای روش‌ها و تحلیل نتایج

برای درک بهتر، دو مرحله آخر چارچوب پیشنهادی در اینجا ادغام شده است تا متناسب با هر سوال، تحلیل نتایج بعد از محاسبات و اندازه‌گیری‌ها توسط ابزار گفی ارائه شود.

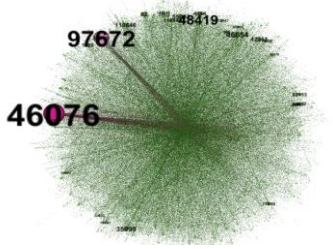
۴-۶-۱- Q1: کدام کاربران در شبکه بیشترین تبادل پیام را دارند؟

پاسخ اول: افرادی که نسبت به سایرین تبادل بیش‌تری داشته‌اند (در اینجا تعداد پیام‌های تبادل شده مهم نیست و میزان مشارکت افراد در تبادل مهم است). برای این سوال از شاخص درجه استفاده می‌نماییم و گره‌هایی که بیش‌ترین درجه (مجموع درجه ورودی و خروجی) را در شبکه S1 دارا هستند، پاسخ سوال می‌باشند. شکل ۳ توزیع درجات رئوس در پاسخ اول به سوال Q1 را نشان می‌دهد.



شکل ۳: توزیع درجات رئوس در پاسخ اول به سوال Q1 در شبکه S1 (میانگین درجات: ۲,۷۲۲)

همان‌طور که شکل ۳ نشان می‌دهد، یک گره دارای بیش‌ترین درجه است، در نتیجه این کاربر با بیش‌ترین تعداد کاربران دیگر به تبادل پیام پرداخته است. شکل ۴ نشان می‌دهد که گره ۴۶۰۷۶، با بیش‌ترین درجه، بیشترین مشارکت را در تبادل پیام دارد.



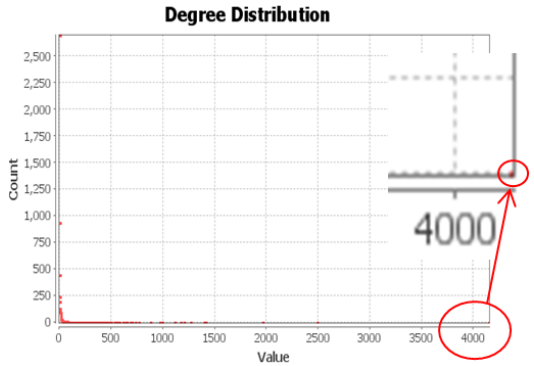
شکل ۴: گراف پاسخ اول به سوال Q1 در شبکه S1

پاسخ دوم: افرادی که پیام‌های بیشتری را در شبکه تبادل کرده‌اند. برای این بخش باید درجه‌های وزن‌دار[1] را محاسبه نماییم. این شاخص وزن‌های درجه هر یک از رئوس را با هم جمع می‌نماید و این برابر است با تعداد پیام‌هایی که هر گره در شبکه ارسال و دریافت کرده است. شکل ۵ توزیع درجات وزن‌دار در پاسخ دوم به سوال Q1 را نشان می‌دهد.

[1] Weighted Degree

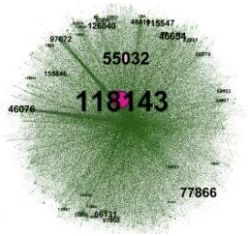
www.SID.ir

Archive of SID



شکل ۵: توزیع درجات وزن‌دار در پاسخ دوم به سوال Q1 در شبکه S1
(میانگین درجات وزن‌دار: ۸،۰۳۵)

بر اساس شکل ۵ بیش‌ترین مقدار این شاخص بیشتر از ۴۰۰۰ می‌باشد و این بدین معنی است که این کاربر بیش‌ترین تعداد پیام را در کل شبکه تبادل کرده است. شکل ۶ نشان می‌دهد که گره ۱۱۸۱۴۳ پیام‌های بیشتری را در شبکه مبادله می‌کند.



شکل ۶: گراف پاسخ دوم به سوال Q1 در شبکه S1

همان‌طور که در شکل‌های ۳ و ۵ مشخص است توزیع درجات رئوس این گراف‌ها، از Power-Law تبعیت می‌کند.

Q2-۲-۶-۴: کدام کاربر بیشترین تعداد درخواست دوستی را دریافت کرده است؟

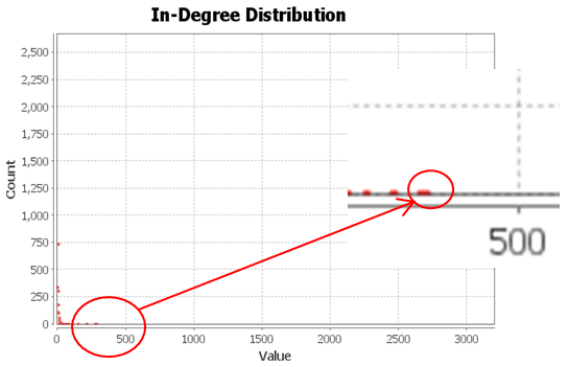
برای یافتن کاربری که بیش‌ترین تعداد درخواست را از کاربران دیگر دریافت نموده است باید شاخص درجه ورودی گره‌ها را محاسبه نمود و گره‌ای که بیش‌ترین درجه ورودی را دارد به عنوان پاسخ این سوال درنظر گرفت. شکل ۷ توزیع درجات ورودی شبکه S2 را نشان می‌دهد.

همان‌طور که در شکل ۷ مشاهده می‌شود بیش‌ترین درجه گره کمتر از ۵۰۰ می‌باشد. در شکل ۸ گره با بیش‌ترین درجه ورودی (سمت چپ‌ترین گره، گره ۱۱۳) به وضوح قابل تشخیص است. این کاربر بیش‌ترین درخواست دوستی را در شبکه دریافت نموده است.

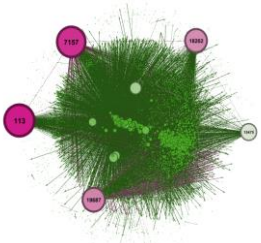
شکل ۹، نمودار توزیع درجات شبکه S2 را نشان می‌دهد. این نمودار نیز از توزیع Power-Law پیروی می‌نماید.

Q3 -۳-۶-۴: کدام کاربران نقش کلیدی‌تری در برقراری ارتباط با بقیه دارند؟

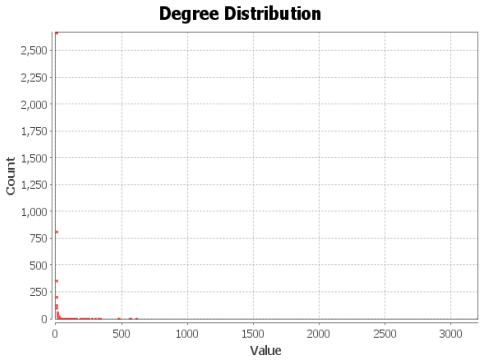
برای پاسخ به این سوال باید کاربر یا کاربرانی که نقش واسط میان کاربران دیگر را دارا هستند و مانند پلی افراد را به هم متصل می‌نمایند در شبکه شناسایی شوند؛ در این صورت مشخص می‌شود که کدام کاربران نقش کلیدی در برقراری ارتباط میان افراد دارا می‌باشند. بدین منظور از شاخص مرکزیت بینابینی استفاده می‌شود. شکل ۱۰ نمودار بینابینی رئوس در شبکه S3 را نشان می‌دهد.



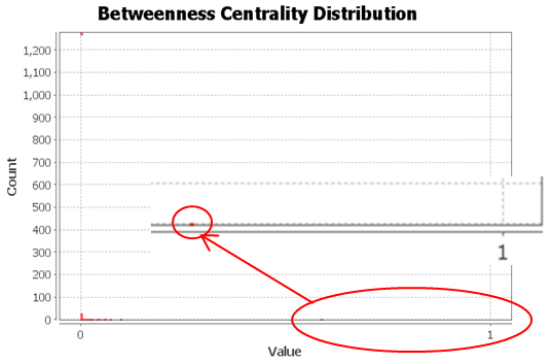
شکل ۷: توزیع درجات ورودی شبکه S2



شکل ۸: گراف شبکه S2



شکل ۹: نمودار توزیع درجات در شبکه S2

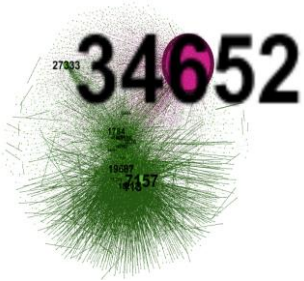


شکل ۱۰: نمودار بینابینی رئوس در شبکه S3

در شکل ۱۰ مشاهده می‌شود که بینابینی تعداد زیادی از گره‌های شبکه S3 صفر می‌باشد و تعداد محدودی از گره‌ها دارای بینابینی بین صفر و یک هستند. در شکل ۱۱ به وضوح مشاهده می‌شود که گره ۳۴۶۵۲ دارای بیش‌ترین مقدار بینابینی در شبکه S3 است.

www.SID.ir

Archive of SID

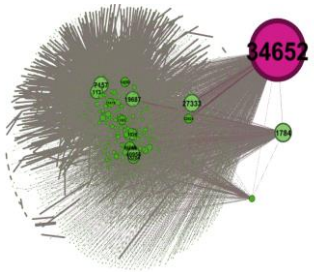


شکل ۱۱: گراف شبکه S3

۴-۶-۴- Q4: کدام اعضای شبکه، دوستان مشترک بیشتری دارند؟

برای بدست آوردن همسایگان مشترک یک گره از شاخص Embeddedness در شبکه‌های جهان کوچک استفاده می‌شود. بر اساس نتایج حاصل از محاسبه این شاخص، متوسط همسایگان مشترک برابر ۰،۰۵۶ و متوسط Embeddedness برابر ۸،۵۰۱ است.

شکل ۱۲ گره‌های دارای بیشترین تعداد همسایه‌های مشترک را نشان می‌دهد. در این شکل یال‌هایی که صورتی هستند دارای بیشترین مقدار شاخص Embeddedness می‌باشند. بر اساس این شکل، گره‌های ۳۴۶۵۲ و ۱۷۸۴ بیش‌ترین تعداد دوست مشترک را دارند.



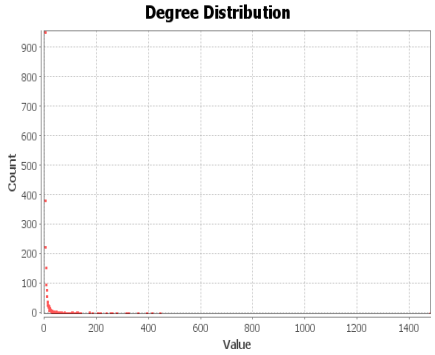
شکل ۱۲: گره های دارای بیش‌ترین تعداد همسایه‌های مشترک شبکه S3

۴-۶-۵- Q5: بیشترین تعداد افرادی که در شبکه دوستی با هم دوست هستند، چقدر است؟

هدف از این سوال بدست آوردن افرادی است که همگی با یکدیگر رابطه دارند. از شاخص n-clique برای پاسخ به این سوال استفاده می‌شود.

با استفاده از این شاخص مشخص شد که بیشترین اندازه هر کلیک چهار (4-clique) می‌باشد و این بدین معنا است که بیش‌ترین تعداد افرادی که در شبکه، همگی با یکدیگر دوست هستند چهار می‌باشد. به عنوان نمونه یکی از 4-cliqueها در این شبکه به شکل یک چهارتایی از گره‌ها با شماره‌های ۳۴۶۵۲، ۲۲۱۴۰، ۱۵۵۲۹ و ۲۹۱۲۵ است.

در شکل ۱۳ نمودار توزیع درجات رئوس شبکه S3 نشان داده شده است، این شبکه نیز از توزیع Power-Law تبعیت می‌کند.

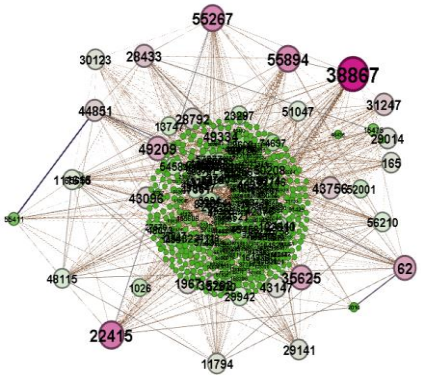


شکل ۱۳: نمودار توزیع درجات رئوس در شبکه S3

۴-۶-۶- Q6: کدام کاربران در بیشترین پست‌های مشترک نظر گذاشته‌اند؟

برای یافتن کاربرانی که بر روی بیشترین تعداد پست مشترک نظر داده‌اند از وزن یال‌ها استفاده می‌شود؛ گره‌هایی پاسخ این سوال می‌باشند که یال متصل به آن‌ها بیشترین وزن را داراست. یال متصل به گره‌های ۴۴۸۵۱ و ۵۵۴۱۱ که ضخیم‌تر از بقیه یال‌ها هستند دارای وزن بیشتری بوده و این گره‌ها پاسخ سوال می‌باشند. شکل ۱۴ گراف شبکه S4 را نشان می‌دهد. در این گراف رنگ‌آمیزی بر اساس درجه گره‌ها و وزن یال‌ها انجام شده است.

شکل ۱۵ نمودار توزیع درجات رئوس شبکه S4 را نشان می‌دهد. شبکه S4 نیز از توزیع Power-Law پیروی می‌کند. همچنین جدول ۴ به اختصار نتایج تحلیل شبکه اجتماعی پارسی‌یار را بیان می‌کند. برای خلاصه‌سازی کد سوالات درج شده است.



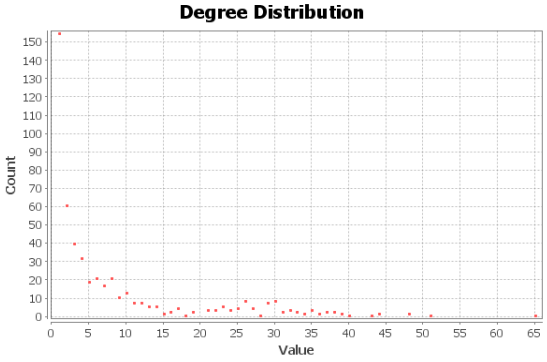
شکل ۱۴: گراف شبکه S4

جدول ۴: نتایج تحلیل شبکه اجتماعی پارسی‌یار

سوال	سناریو	معیار تحلیل
Q1	S1: شبکه تبادل پیام	- درجه راس‌ها - جمع وزن‌های درجه هر راس
Q2	S2: شبکه درخواست دوستی	درجه ورودی گره‌ها
Q3	S3: شبکه روابط دوستی	مرکزیت بینابینی
Q4		Embeddedness
Q5		n-clique
Q6	S4: شبکه نظرهای افراد	وزن یال‌ها

www.SID.ir

Archive of SID



شکل ۱۵: نمودار توزیع درجات رئوس در شبکه S4

بعد از مشخص شدن پاسخ همه سوالات، نتایج بدست آمده با مدیر پارسی‌یار جهت ارزیابی به اشتراک گذاشته شد. همچنین ویژگی‌های شخصی کاربرانِ پاسخ نیز بدست آمد. به عنوان مثال گره ۱۱۸۱۴۳ در سوال Q1 مدیر یکی از گروه‌های پرمخاطب در شبکه پارسی‌یار است.

۵- نتیجه‌گیری و کارهای آتی

تحلیل شبکه‌های اجتماعی یکی از روش‌های تحلیل مبتنی بر ساختار گراف است که به واسطه مقبولیت انواع شبکه‌های اجتماعی و اتصال موجودیت های مختلف (از اتصال افراد تا اتصال اشیاء) بسیار کاربرد دارد. در این مقاله، یک چارچوب پایه برای استفاده از این روش تحلیل، فارغ از مجموعه داده و حوزه کاربرد ارائه شد که می‌تواند راهگشای محققان در علوم مختلف باشد.

در این مقاله از داده‌های شبکه اجتماعی پارسی‌یار به منظور اجرای گام به گام چارچوب پیشنهادی استفاده شد. در نتیجه ۶ سوال طرح و متناسب با آنها ۴ شبکه متفاوت استخراج و بر اساس معیارهای مختلفِ تحلیل شبکه‌های اجتماعی، پاسخ سوالات مشخص شد. بر اساس چارچوب پیشنهادی و نتایج اجرای آن، استفاده از تمام معیارها، شاخص‌ها و روش های تحلیل شبکه‌های اجتماعی برای تحلیل یک مجموعه داده توصیه نمی‌شود. بلکه متناسب با سوالات تحقیق می‌بایست آن‌ها را انتخاب کرد.

در مطالعات بعدی قصد داریم چارچوب پیشنهادی را در حوزه‌های سازمانی و تجاری به کار بریم و همچنین ابعاد دیگر مجموعه داده پارسی‌یار را مورد تحلیل قرار دهیم.

مراجع

[۱] هانی زواره‌ای، سید حسن؛ اسماعیلی، لیلا؛ پیرمحمدیانی، روجیار؛ منتی، صبا؛ چالش‌ها و راهبردهای اخلاقی در شبکه‌های اجتماعی، اولین کنگره فناوری اطلاعات و ارتباطات ایران، تهران، خرداد ۸۹.

[۲] لیلا اسماعیلی، سید علیرضا هاشمی گلپایگانی، "چالش های موجود در کاوش و تحلیل شبکه های اجتماعی فارسی زبان"، همایش تخصصی بررسی ابعاد شبکه های اجتماعی (سلسله همایش های دفاع سایبری ۳)، صفحه ۱۹۱-۲۰۳، تهران، ایران، مهر ۱۳۹۲.

[3] Social Network, www.en.wikipedia.org/wiki/Social_network, Last Access: 6/29/2016.

[4] The 10 Biggest Social Networks Worldwide, http://www.adweek.com/socialtimes/largest-social-networks-worldwide/504044, Last Access: 6/29/2016.

[5] Number of social network users worldwide from 2010 to 2019, http://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/, Last Access: 6/29/2016.

[6] Iran Social Media Usage Statistics Using Mobile, https://www.statsmonkey.com/sunburst/21374-iran-mobile-social-media-usage-statistics-2015.php, Last Access: 6/29/2016.

[7] K. Faust and S. Fitzhugh, Social Network Analysis: An Introduction, The Odum Institute, University of North Carolina, Chapel Hill, July 16 to 20, 2012.

[8] K. Sheykh Esmaili, M. Jamali, M. Neshati, H. Abolhassani and Y. Soltan-Zadeh, Experiments on Persian Weblogs, In Proceedings of the WWW06 Workshop on Web Intelligence, 2006.

[9] Sh. Sahebi, F. Oroumchian and R. Khosravi, An Enhanced Similarity Measure for Utilizing Site Structure in Web Personalization System, 2008 IEEEWICACM International Conference on Web Intelligence and Intelligent Agent Technology, pp. 82-85, 2008.

[10] Z. Borhani-fard, B. Minaei-Bidgoli and H. Alinejad, Applying Clustering Approach in Blog Recommendation, Journal of Emerging Technologies in Web Intelligence, Vol. 5, No. 3, pp. 296-301, 2013.

[11] L. Esmaeili, M. Nasiri and B. Minaei-Bidgoli, Personalizing Group Recommendation to Social Network Users, in Web Information Systems and Mining LNCS, Springer Berlin Heidelberg, Vol. 6987, pp. 124-133, 2011.

[12] L. Esmaeili, M. Nasiri and B. Minaei-Bidgoli, Analyzing Persian Social Networks: An Empirical Study, IJVCSN, 3(3), pp. 46-65, 2011.

[13] G. A. Afzali, B. and S. A. Hashemi G., Design Context aware recommender system model in m-commerce platform using collaborative filtering approach in social networks, CJASR, 2(7), pp. 1-10, 2013.

[14] S. P. Borgatti, "Notes on the History of Social Network Analysis," 1997. http://www.analytictech.com/networks/history.htm, Last Access: 6/29/2016.

[15] J. Scott, What is Social Network Analysis?. Bloomsbury Academic, 2012.

[16] D. Easley and J. Kleinberg, Networks, Crowds, and Markets. 2010.

[17] M. Jamali and H. Abdolhassani, "Different Aspects of Social Network Analysis," in IEEWICIACM International Conference on Web Intelligence, 2006.

[18] P. Kazienko, K. Musial, E. Kukla, T. Kajdanowicz, and P. Bródka, "Multidimensional Social Network: Model and Analysis," in Lecture Notes in Computer Science, Lecture Notes in Artificial Intelligence, Springer, 2011, pp. 378–387.

[19] T. Opsahl, "Triadic closure in two-mode networks: Redefining the global and local clustering coefficients," Soc. Netw., vol. 35, pp. 159–167, 2013.

[20] R. A. Hanneman and M. Riddle, Introduction to social network methods. 2005.

[21] S. Uddin and M. J. Jacobson, "Dynamics of email communications among university students throughout a semester", Elsevier, Journal of Computer & Education 64, pp. 95-103, 2013.

[22] Centrality, https://en.wikipedia.org/wiki/Centrality, Last Access: 6/29/2016.

[23] J. Wang, T. Kraska, M. J. Franklin and J. Feng, "CrowdER: Crowdsourcing Entity Resolution", arXiv:1208.1927v1, Proceedings of the VLDB Endowment (PVLDB), Vol. 5, No. 11, pp. 1483-1494, Aug 2012.

www.SID.ir

Archive of SID

www.SID.ir