

یافتن افراد تأثیرگذار در گراف شبکه‌های اجتماعی براساس الگوریتم CSCS و مقدار شاپلی در نظریه بازی

مریم خادمی*، نیما شیخ‌خانی، پونه خدابخش

دانشگاه آزاد اسلامی واحد تهران جنوب

پذیرش ۹۷/۱۱/۱۹

دریافت ۹۷/۰۳/۲۸

چکیده

پیشرفت‌های اخیر شبکه‌های اجتماعی آنلاین به ویژه کاربردهای آن در دنیای فناوری و اطلاعات مدرن، موجب گسترش چشمگیر نظریه‌های گراف و بازی شده است و توجه بسیاری از محققین ریاضی، متخصصین علوم کامپیوتر و تحلیل‌گران آماری را به خود جلب کرده است. یکی از ویژگی‌های مهم و کلیدی شبکه‌های اجتماعی این است که گسترش روابط بین افراد می‌تواند در تصمیم‌گیری آنها، تأثیر به‌سزای داشته باشد. لذا یکی از مباحث مطرح و کاربردی در شبکه‌های اجتماعی، یافتن تأثیرگذارترین و بانفوذترین افراد در راستای بیشینه‌سازی تأثیر فعالیت‌های آنها در ایجاد تبلیغات و بررسی انتشار در خرید کالا، پخش شایعات مخرب، انتشار اخبار کاذب، مهندسی انتخابات و ... است. در این مقاله، ابتدا به بررسی انتشار میان گره‌ها با استفاده از مرکزیت مقدار شاپلی، تقسیم یک شبکه به جوامع کوچک‌تر و مدل آشناری در نظریه‌بازی‌ها می‌پردازیم. سپس برای یافتن تأثیرگذارترین و بانفوذترین افراد در گراف شبکه‌های اجتماعی الگوریتم CSCS پیشنهاد گردیده که روی مجموعه داده‌های مختلفی پیاده‌سازی شده است. در نهایت، نتایج الگوریتم پیشنهادی با نتایج سایر الگوریتم‌های موجود مقایسه شده است.

واژه‌های کلیدی: گراف شبکه‌های اجتماعی، نظریه بازی، بیشینه‌سازی نفوذ، مقدار شاپلی، جوامع، الگوریتم CSCS.

مقدمه

در دهه‌های اخیر، رشد و گسترش اینترنت موجب شده است تا مردم به‌روشنی دیگری دانش، عقاید و نظریه‌های خود را به اشتراک بگذارند. شبکه‌های اجتماعی آنلاینی مانند فیسبوک، توئیتر، تلگرام، اینستاگرام، لینکدین و گوگل پلاس و غیره، کشف و انتشار اطلاعات را بسیار آسان کرده‌اند. این ساختار از مجموعه‌ای از گره‌ها تشکیل شده‌اند که به‌عنوان افراد یا سازمان‌هایی در نظر گرفته شده‌اند که به‌دلیل اهداف مشخصی هم‌چون دوستی یا همکاری باهم در ارتباط هستند. بدین ترتیب می‌توان در هر شبکه اجتماعی، کاربران را به منزلهٔ رئوس و ارتباط (یک‌طرفه و یا دوطرفه) را به‌عنوان یال و میزان احتمال ارتباط یا حجم تبادل اطلاعات آنها را به‌عنوان وزن در آن گراف وزن‌دار و یا جهت‌دار در نظر گرفت. بدیهی است می‌توان با استفاده از ماتریس‌های مجاورت و وقوع میزان درجه ارتباطی افراد را با سایر اعضای شبکه را یافت که هر چقدر درجه بالاتر باشد نشان‌دهندهٔ میزان ارتباط بیشتر است ولی این به‌تنهایی کافی نیست [۱] زیرا ممکن است آن فرد صرفاً به سایر اعضا در یک گروه و یا یک جامعه متصل شود ولی هیچ فعالیت خاص یا اثربخشی که مورد توجه و یا استقبال سایر کاربران و سیاست‌مداران آن شبکه باشد، را انجام ندهد. پس نمی‌توان او را فردی با نفوذ و تأثیرگذار در آن شبکه قلمداد کرد. از این‌رو، بررسی ویژگی‌های گراف، تجزیه و تحلیل دقیق ریاضی و آماری شبکه‌های

اجتماعی تحقیقات مختلفی را تحت پوشش قرار می‌دهد. یکی از مهم‌ترین موضوعات، که توجه بسیاری از بازرگانان و تجار را در شبکه‌های اجتماعی به خود جلب کرده است، بازرگانی ویروسی برای تبلیغ کالاهای جدید است [۲]، که در آن روابط و تأثیر مشتریان بر یکدیگر برای تبلیغ کالا به کار می‌رود. به‌عنوان مثال، شرکتی را در نظر بگیرید که قصد تبلیغ محصولات خود را دارد. شرکت به کمک بازرگانی ویروسی قادر است که کالاهای خود را به برخی از افراد تأثیرگذار در شبکه اجتماعی ارائه دهد و سپس این افراد این کالاها را با داشتن کالا و همین‌طور آگهی‌های کلامی تبلیغ کند. فرایند انتخاب k گره، که در یک شبکه تأثیرگذار است و می‌تواند شمار زیادی از گره‌ها را تحت تأثیر قرار دهد، حداکثرسازی نفوذ و تأثیرگذار گفته می‌شود. یافتن افراد بانفوذ و مؤثر در گراف شبکه‌های اجتماعی نقش کلیدی در پیشینه کردن تأثیر نقش آنها دارد [۳]. هم‌چنین با توجه به عملکرد و کارکردهای متفاوت کاربران در شبکه‌های اجتماعی برای به اشتراک‌گذاری دانش و تجارب خود در زمینه‌های مختلف زندگی، نمی‌توان از تأثیر فعالیت‌های سودآور یا مخرب افراد با نفوذ در این شبکه‌ها غافل ماند. از این رو می‌توان ضرورت و اهمیت شناسایی این گونه افراد در شبکه‌های اجتماعی را در راستای سالم‌سازی و بهره‌وری هر چه بیش‌تر آن با عنایت به گسترش دانش تخصصی در علوم کامپیوتر، نظریه گراف و نظریه بازی‌ها دانست. ضمن این که نباید فراموش کرد که درصد عظیمی از مخاطبان این نوع رسانه‌ها را نوجوانان و جوانان تشکیل می‌دهند که از نظر توسعه، مهم‌ترین سرمایه انسانی هر کشور محسوب می‌شوند. از این‌رو، نتایج این تحقیق با بهره‌گیری از دانش نوین ریاضی در مقایسه با واکاوی سایر نتایج موجود صحت و دقت خوبی داشته و نقش به‌سزایی در تعامل افراد، تأمین نیازهای کاربران، تعمیق و نهادینه کردن مشارکت مردمی در چشم‌انداز این رسانه اجتماعی مجازی در جامعه مدنی دارد. بدیهی است گسترش کاربردهای ریاضی می‌تواند انگیزه‌های بیش‌تری در دانش‌آموختگان آن برای توسعه مفاهیم پایه و بنیادی این علوم را فراهم آورد.

مفاهیم پژوهش

انتشار

یکی از ویژگی‌های شبکه‌های اجتماعی انتشار اطلاعات است. مردم اغلب از رفتارهای یکدیگر یاد می‌گیرند و تصمیم دوستان یک فرد روی تصمیم او اثرگذار است. چارچوب مفهومی انتشار اطلاعات به گسترش ایده‌ها یا اطلاعات فنی درون یک سیستم اجتماعی اشاره دارد، که این گسترش جریان یا انتقال از یک منبع به منبع دیگر را نشان می‌دهد. این ارتباط می‌تواند میزان رضایت‌مندی از یک محصول به‌خصوص را به‌وسیلهٔ گیرندگان آن تغییر دهد [۴]. در مسائل پیشینه‌سازی نفوذ، حداقل دو ورودی نیاز است: یک گراف از شبکه اجتماعی و یک مدل انتشار، که نحوهٔ انتشار اطلاعات در شبکه را مشخص می‌کند. نحوهٔ انتشار اطلاعات در شبکه‌ها به درستی شناخته نشده است و هنوز محققان علوم مختلف، مانند علوم کامپیوتر و اپیدمیولوژی، به بررسی روش‌های مختلف می‌پردازند. مدل‌های انتشار را می‌توان به دو دسته تقسیم‌بندی کرد: مدل آستانه خطی [۵] و مدل آبخاری مستقل. مدل آبخاری مستقل در هر گام با مجموعه‌ای از رؤس (یا گره‌ها) شروع می‌شود و بعد از آن فعال‌سازی آبخاری صورت می‌پذیرد. بدین ترتیب که، زمانی که یک رأس فعال شد، می‌تواند رؤس (یا گره‌های) همسایه غیرفعال را با احتمال موفقیت بیش‌تر، فعال کند. از سوی دیگر، مدل آستانه خطی براساس استفاده از مقدار آستانه عمل می‌کند. در این مدل، یک گراف جهت‌دار تعریف می‌شود که یال‌های آن وزن‌دار شده است. w_{ab} وزن یال بین دو رأس a و b ($w_{ab} \geq 0$) در نظر گرفته می‌شود و بیان‌گر میزان تأثیرگذاری رأس a روی b است [۶]. زمانی رأس a فعال می‌شود که تعداد همسایگان فعال a بیش‌تر از مقدار آستانه a

باشد [۷] به‌طور خلاصه، دو تفاوت اصلی بین روش‌های آستانه خطی و روش آبخاری مستقل، وجود دارد، در درجه اول گروه‌های فعال در روش آستانه خطی بیش از یک شانس دارند تا بر همسایه‌های غیرفعال خود تأثیر بگذارند. و درجه دوم در مدل آستانه خطی، همسایه‌های فعال رأس v ، همگی بر رأس v نیز تأثیر می‌گذارند، در حالی که در مدل آبخاری مستقل، همسایگان فعال تنها یک شانس دارند تا به‌طور مستقل بر رأس v تأثیر بگذارند.

شناسایی جوامع در شبکه‌های اجتماعی

به‌طور کلی، افراد به دلایل مختلفی مانند شغل، دوستی، علاقه، موقعیت و تحصیلات در جوامع گروه‌بندی می‌شوند. در جوامع، تعداد روابط درون گروهی از تعداد روابط بین‌گروهی بیشتر است [۸]، [۹]. از نقطه نظر ریاضی، یک جامعه به‌عنوان زیرگراف G' از گراف G است که مجموع درجات درون گراف G' بیش‌تر از مجموع تمامی درجه‌های آن به زیرگراف‌های دیگر است [۱۰].

بازی‌های همکارانه و مقدار شاپلی

نظریه بازی، شاخه‌ای از ریاضیات کاربردی است که با استفاده از مدل‌های ریاضی به تحلیل روش‌های همکاری یا رقابت موجودات منطقی و هوشمند می‌پردازد [۱۱]. این نظریه در مهندسی کامپیوتر، اقتصاد، زیست‌شناسی، علوم سیاسی، علوم اجتماعی، روابط بین‌الملل، فلسفه و غیره کاربرد فراوان دارد. نظریه بازی در تلاش است تا به‌وسیله ریاضیات، رفتار را در شرایط راهبردی یا در یک بازی که در آن موفقیت فرد در انتخاب کردن وابسته به انتخاب دیگران است، برآورد کند [۱۲]. در روابط اجتماعی، افراد و گروه‌ها هر یک منافع خاص خود را داشته و علت اصلی وارد شدن آنها در این روابط در نهایت کسب آن منافع خاص است. خلاصه کلام این که در جوامع مدنی مسئله تعیین تکلیف منافع افراد و گروه‌ها و به‌طور کلی مسئله قدرت تبدیل به یک بازی استراتژی شده است که اگر فرض را بر این قرار بدهیم که شرکت‌کنندگان رفتاری خردگرایانه داشته و هدفشان تأمین حداکثر منافع خویش است، در این صورت این بازی‌ها تن به تحلیل منطقی ریاضی می‌دهند. ابزاری ریاضی که چنین تحلیل‌هایی را ممکن می‌سازد نظریه بازی است [۱۳]. تمرکز این نظریه در سال‌های اول شکل‌گیری آن بر بازی‌های دارای تعارض خالص (بازی‌های با حاصل جمع صفر) بود. سایر بازی‌ها همکارانه در نظر گرفته می‌شدند، به این معنا که شرکت‌کنندگان اعمال خود را به اتفاق یک‌دیگر انتخاب کرده و انجام می‌دهند. اما در تحقیقات اخیر بر بازی‌هایی متمرکز شده که نه دارای حاصل جمع صفر هستند و نه به‌طور خالص همکارانه هستند. در این بازی‌ها بازیکنان اقدامات خود را به‌صورت جداگانه انتخاب می‌کنند، اما روابط آنها با یک‌دیگر حاوی رقابت و همکاری است [۱۴].

یک بازی همکارانه را می‌توان با استفاده از راه‌حلی‌هایی تجزیه و تحلیل کرد که مقدار کلی بازی را میان بازیکنان تقسیم می‌کند. یکی از این راه‌حل‌ها مقدار شاپلی است. این روش سعی دارد به مفهوم عدالت در تابع مشخصه بازی بپردازد. عدالت به این معنی که نتیجه نهایی هر بازیکن در برابر مشارکتش چقدر تناسب دارد. این مفهوم پس از مبدع آن لوید اس شاپلی به نام مقدار شاپلی نام‌گذاری شد [۱۵]. در تابع مشخصه، $G = (N, v)$ ، فرض کنید که \prod_N مجموعه‌ای از جایگشت‌های N باشد. با توجه به جایگشت $\pi \in \prod_N$ ، مجموعه‌ی تمامی افراد است، قبل از این که بازیکن i در π قرار گیرد. به عبارت دیگر، $S_\pi(i) = \{j \in N \mid \pi(j) < \pi(i)\}$

مشارکت حاشیه‌ای یک بازیکن براساس جایگشت π در بازی $G = (N, v)$ با استفاده از $\Delta_{\pi}^G(i)$ تعیین می‌شود، که بدین صورت محاسبه می‌شود:

$$\Delta_{\pi}^G(i) = v(S_{\pi}(i) \cup \{i\}) - v(S_{\pi}(i)) \quad (1)$$

باتوجه به $G = (N, v)$ و $|N| = n$ ، مقدار شاپلی بازیکن $i \in N$ می‌تواند به کمک رابطه (۲) محاسبه شود [۱۶]:

$$\varphi_i(G) = \frac{1}{n!} \sum_{\pi \in \Pi_N} \Delta_{\pi}^G(i) \quad (2)$$

مقدار شاپلی خصوصیات منحصر به فردی دارد، که عبارتند از:

- کارایی: $\sum_{i \in N} \varphi_i(G) = v(N)$
 - بازیکن ساختگی: اگر i بازیکن ساختگی باشد، آن گاه $\varphi_i(G) = 0$.
 - شباهت: اگر i و j در G مشابه باشند، آن گاه $\varphi_i(G) = \varphi_j(G)$.
 - جمعیت بودن: برای هر $i \in N$ داریم: $\varphi_i(G^1 + G^2) = \varphi_i(G^1) + \varphi_i(G^2)$ که $G^1 = (N, v^1)$ و $G^2 = (N, v^2)$ دو تابع مشخصه بازی روی مجموعه بازیکن یکسان هستند.
- با وجود تمامی مزایا و ویژگی‌های مقدار شاپلی، این روش از معایبی، هم چون زمان طولانی اجرا، رنج می‌برد [۱۷].

محاسبه مرکزیت براساس مقدار شاپلی

به‌طور کلی، مرکزیت، به یک رتبه‌بندی سازگار از گره‌ها (یا رئوس) بر اساس یک ویژگی سازگار در یک شبکه اجتماعی اشاره دارد. معیار مرکزیت، عددی را به هر گره (یا رأس) اختصاص می‌دهد که میزان اهمیت آنها را براساس ویژگی سازگار مشخص می‌کند. مرکزیت‌های مختلفی معرفی شده‌است، که سه تا از معروف‌ترین آن عبارتند از: مرکزیت درجه، مرکزیت نزدیکی و مرکزیت بینابینی.

ویژگی مشترک میان بسیاری از معیارهای مرکزیت این است که آنها می‌توانند میزان اهمیت یک گره بر اساس نقش گره تخمین بزنند. با این وجود، گاهی بهتر است که نقش گره در حالت جمعیت بررسی شود. چارچوبی که با این مسائل سروکار دارد مرکزیت شبکه مبتنی بر نظریه بازی است. این معیار به کمک مشارکتی که ممکن است بین گره‌ها رخ دهد، رتبه‌بندی سازگاری را برای هر گره ارائه می‌دهد. پژوهشی که در سال ۲۰۱۳ انجام شد [۱۷]، روش‌های کارایی را برای محاسبه مقدار شاپلی برای انواع مختلفی از بازی‌ها معرفی کرد.

الگوریتم (۱)، پیاده‌سازی شده در محیط متلب چگونگی محاسبه مقدار شاپلی را نشان می‌دهد. در این جا فرض شده که یک گره در صورتی مؤثر است که بتواند n گرهی همسایه خود را تحت تأثیر قرار دهد. زمان پیچیدگی این الگوریتم $O(|V| + |E|)$ است، که برای شبکه‌های اجتماعی بسیار بزرگ میزان و مقدار مناسبی است.

کارهای مرتبط

ریچاردسون و دمینگز، اولین کسانی بودند که روی پیشینه‌سازی تأثیر در شبکه‌های اجتماعی مطالعه کردند [۱۵] و [۱۸]. در سال ۲۰۰۱ آنها بازاریابی به صورت یک شبکه اجتماعی مدل‌سازی کردند که مشتریان را گره‌های شبکه در نظر گرفتند و برای مدل‌سازی تأثیرات آنها از مدل احتمالاتی مبتنی بر مارکوف استفاده کردند. در سال ۲۰۰۳، مسئله پیشینه‌سازی نفوذ به صورت مسئله بهینه‌سازی گسسته شد. کمپ و همکارانش، مسئله ساده حریصانه‌ای با ضریب

تخمین $1 - 1/e$ را مطرح کردند که به شدت مشکل مقیاس‌پذیری دارد. هم‌چنین، مسئله NP-hard هم محسوب می‌شود [۱۹]. بعد از آن، چندین روش مختلف برای حل و بهبود الگوریتم‌های حریصانه مطرح شد.

Algorithm 1: Computing the Shapley value

1. *Input: Graph $G(V;E)$, positive integer n*
 2. *Output: Shapley value of all nodes in V (G)*
 3. *foreach $v \in V$ (G) do*
 4. $SV[v] = \min\left(1; \frac{n}{1+degG(v)}\right);$
 5. *foreach $u \in NG(v)$ do*
 6. $SV[v] += \max\left(0; \frac{degG(u)-n+1}{degG(u)(1+degG(u))}\right)$
 7. *end*
 8. *end*
 9. *return SV*
-

در سال ۲۰۰۷، روش نوینی برای انتخاب گره‌ها معرفی شد که برای نمایش چگونگی توزیع آنها در یک گراف به کار می‌رفت. هم‌چنین، آنها از تابع هدفی، مانند زمان جستجو، استفاده کردند تا الگوریتم حریصانه تنبلی برای یافتن گره نزدیک به بهینه را توسعه دهند، که الگوریتم CELF نام دارد. سرعت شناسایی انتشار این روش تا ۷۰۰ بار سریع‌تر از الگوریتم حریصانه بود [۲۰].

هم‌چنین، پندیت و همکارانش [۲۱] الگوریتمی برای یافتن K گره با بیش‌ترین نفوذ در شبکه معرفی کردند. الگوریتم آنها در دو مرحله، یادگیری و جستجو، اجرا می‌شد. در مرحله یادگیری، اطلاعات در هر گره قرار می‌گیرد. در نتیجه، ماتریسی به دست می‌آید که به‌عنوان ورودی به مرحله بعدی وارد می‌شود تا k عضو تأثیرگذار را جستجو و انتخاب کند. الگوریتم LNG در سال ۲۰۱۴ معرفی شد که گره‌های مؤثر را برای بیشینه‌سازی نفوذ تحت مدل آستانه خطی انتخاب کند. الگوریتم LNG نه‌تنها بهتر از الگوریتم حریصانه کلاسیک عمل می‌کرد، بلکه زمان اجرا را نیز کاهش داد [۲۲]. در همین دوره، روش‌های جدید دیگری ظهور کردند که الگوریتم‌های فراابتکاری و ژنتیک به کار بردند [۲۳]، [۲۴]، [۲۵] (Xie, 2011 & Jiang, Song, Cong, Wang, Si). ژیانگ و همکارانش روش متفاوتی را معرفی کردند که برای اولین بار از الگوریتم شبیه‌سازی تبرید برای انتخاب k گره که بیش‌ترین تأثیر را داشتند، بهره می‌برد. آنها دو روش اکتشافی و یک روش محاسباتی نوین را معرفی کردند که به ترتیب به فرایند همگرایی SA و روش پیشنهادی سرعت بخشید [۲۱]. با گذشت زمان به‌کارگیری الگوریتم‌های فراابتکاری متداول‌تر شد، به‌طوری‌که، در سال ۲۰۱۶، الگوریتم‌های فراابتکاری برای حل مسئله بیشینه‌سازی نفوذ و ارائه مدل انتشار، مانند مدل انتشار جستجوی کوکو^۱ [۲۳] و الگوریتم بهینه‌سازی انبوه ذرات گسسته [۲۵] ارائه شدند.

به‌کارگیری الگوریتم‌های تکاملی و بهینه‌سازی تا سال ۲۰۱۸ ادامه پیدا کرد، لیژانگ و همکارانش در سال ۲۰۱۸ روش نوینی برای بیشینه‌سازی نفوذ ارائه دادند. آنها استراتژی جستجو نزولی درجه (DDS) را ارائه دادند. DDS قادر

1. Cuckoo Search Diffusion Model (CSDM)
2. Degree-descending Search Strategy (DDS)

است گروهی از گره‌ها که انتشار نفوذ بین آنها متناظر با مرکزیت درجه آنها است، تولید کند. بر اساس DDS، آنها الگوریتم تکاملی توسعه دادند که می‌تواند با حذف شبیه‌سازی‌های زمان‌بر الگوریتم‌های حریمانه، به‌طور قابل ملاحظه‌ای کارایی را بهبود بخشد [۲۷].

باتوجه به روش‌هایی که در این بخش توصیف شد، و دیگر روش‌های موجود در مسئله بیشینه‌سازی نفوذ، می‌توان این روش‌ها را به دو دسته مجزای حریمانه و اکتشافی مطابق جدول (۱) تقسیم کرد.

جدول ۱. روش‌های مختلف ارائه شده برای حل مسئله بیشینه‌سازی نفوذ

شرح	سال	روش	
به‌کارگیری مدل احتمالاتی مارکوف	۲۰۰۱	Richardson and Domingos	الگوریتم حریمانه
ارائه یک الگوریتم حریمانه ساده با استفاده از الگوریتم بهینه‌سازی گسسته	۲۰۰۳	Kempe et al.	
یک الگوریتم حریمانه تئوری برای یافتن گره تقریباً بهینه	۲۰۰۷	Cost-Effective Lazy Forward algorithm (CELF)	
یافتن مجموعه‌هایی غیرهم‌پوشان از گره‌ها که نفوذ و تأثیر آنها حداکثر تأثیر در تمام شبکه را داشته باشند.	۲۰۰۷	Set Covering Greedy (SCG) algorithm	
بهبودیافته الگوریتم CELF با استفاده از	۲۰۱۱	CELF++	
به‌کارگیری مدل آستانه خطی	۲۰۱۴	LNG	الگوریتم اکتشافی
رفع مشکل مقیاس‌پذیری و امکان اعمال بر روی شبکه‌های بسیار بزرگ با میلیون‌ها گره و یال	۲۰۱۰	Chen et al.	
تقسیم شبکه اجتماعی به جوامع کوچک‌تر و انتخاب گره‌های مؤثر بر اساس سه معیار مرکزیت درجه، مرکزیت نزدیکی و تعداد ارتباطات.	۲۰۱۱	Anjerani and Moeini	
یافتن k گره که بیش‌ترین تأثیر را دارند و استفاده از دو روش اکتشافی و یک روش محاسباتی نوین برای سرعت بخشیدن به فرایند همگرایی در شبیه‌سازی تیرید.	۲۰۱۱	Simulated Annealing Based method	
تبدیل مسئله بیشینه‌سازی تأثیر به مسئله بهینه‌سازی و به‌کارگیری DPSO	۲۰۱۶	DPSO Based Method	
اعمال جستجوی انتشار تأثیر تنها بر مسیرها با طول ۱	۲۰۱۶	Gonzalez et al.	
میتنی بر الگوریتم جستجوی کوکو	۲۰۱۶	Cuckoo Search Diffusion Model (CSDM)	
به‌کارگیری استراتژی جستجو نزولی درجه	۲۰۱۸	Laizh on g Cui et al.	

روش پیشنهادی

بر اساس بازی همکارانه G_2 ، ارزش هر ائتلاف برابر است با گره‌هایی که در آن قرار دارند یا حداقل مجاور n گره داخل ائتلاف هستند. در روش پیشنهادی، بعد از این‌که مقدار شاپلی برای هر گره در شبکه اجتماعی محاسبه شد، به‌دست آوردن مجموع مقادیر شاپلی گره‌های موجود در هر اجتماع، مقدار شاپلی هر جامعه به‌دست می‌آید. این مقدار به‌عنوان معیاری برای ارزیابی اجتماع به‌کار می‌رود. الگوریتم (۲) مراحل اجرای انتخاب k گره تأثیرگذار را نشان می‌دهد، که n میانگین درجات گره‌ها است. در الگوریتم (۲)، به ازای هر گره، با استفاده از الگوریتم (۱)، مقدار شاپلی آن محاسبه می‌شود (خط ۵). سپس جهت داشتن معیاری برای انتخاب عادلانه، مقادیر شاپلی هر گره درون جامعه با هم جمع شده تا مقدار شاپلی جامعه به‌دست آید (خط ۶). در نهایت، گره‌های درون جوامع براساس مقدار شاپلی آنها به‌صورت نزولی مرتب می‌شود و انتخاب عادلانه اعمال می‌شود (خط ۱۲).

Algorithm 2: Community based Shapley value Centrality Selection(CSCS)

```

1. Input: Graph(V, E), Graph communities  $C_0, C_1, \dots, C_n$ , positive integer  $k$ ;
2. foreach  $c$  in communities
3.   var CommunityValue = 0;
4.   foreach  $n$  in  $c$ .nodes
5.      $n$ .ShapleyValue = Compute Shapley value of  $n$  using Algorithm 1
6.     CommunityValue +=  $n$ .ShapleyValue;
7.   end
8.    $c$ .Value = CommunityValue;
9. end
10. Compute fair selection of each community based on it's value
11. foreach  $c$  in communities
12.   Sort nodes within community in decreasing order and take nodes base on fair selection
13. end

```

Algorithm 3: Computing Communities Selection Count

```

1. Input: Graph  $G(V, E)$  Communities,  $K$ : Total Selection Number
2. var TotalSum = 0;
3. foreach (var item in Communities)
4.   TotalSum += item.Value;
5. end
6. var CommunitiesSelectionSum = 0;
7. foreach (var item in Communities)
8.   var Count = (item.Value / TotalSum) * K;
9.   item.SelectionCount = Round(Count);
10.  item.RoundingDifferences = Count - Round(Count);
11.  CommunitiesSelectionSum += Round(Count);
12. end
13. While (true)
14.  if (CommunitiesSelectionSum == K)
15.    break;
16.  if (CommunitiesSelectionSum > K)
17.    C = Find Community with SelectionCount > 0 and minimum RoundingDifferences
18.    C.SelectionCount = C.SelectionCount - 1;
19.    C.RoundingDifferences = 0;
20.    CommunitiesSelectionSum --;
21.  end
22.  if (CommunitiesSelectionSum < K)
23.    C = Find Community with SelectionCount > 0 and maximum RoundingDifferences
24.    C.SelectionCount = C.SelectionCount + 1;
25.    C.RoundingDifferences = 0;
26.  end
27. end

```

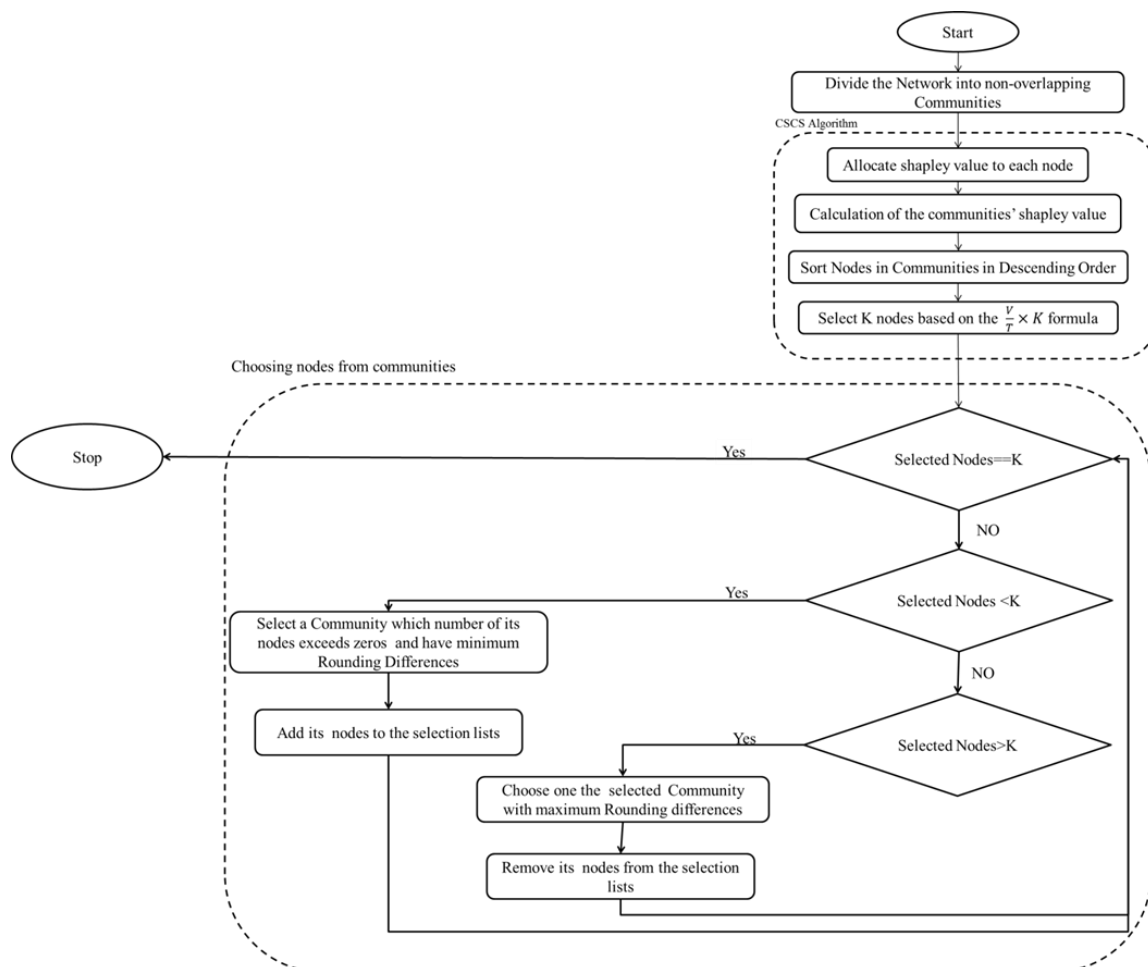
شکل (۱)، مراحل اجرای الگوریتم پیشنهادی را نشان می‌دهد که به دو مرحله تقسیم شده است. در ابتدا مقدار شاپلی به هر گره در شبکه اجتماعی اختصاص می‌یابد و سپس شبکه به جوامع مجزایی تقسیم می‌شود. در ادامه الگوریتم انتخاب منصفانه روی شبکه اعمال می‌شود. در نهایت گره‌های انتخاب شده برای یافتن K گره تأثیرگذار ارزیابی می‌شوند.

در روش پیشنهادی، یک شبکه اجتماعی به جوامع غیرهمپوشانی تقسیم می‌شود. به عبارت دیگر، هر گره باید تنها در یک گروه قرار داشته باشد. از آنجایی که به هر جامعه مقدار شاپلی منحصر به فردی اختصاص می‌یابد، از این رو، الگوریتم انتخاب منصفانه‌ای را برای انتخاب گره‌ها از جوامع به کار می‌گیریم.

انتخاب منصفانه گره‌ها از جوامع با استفاده از رابطه (۳) به دست می‌آید، که در آن V ، T و K به ترتیب مقدار شاپلی جامعه، مجموع مقادیر شاپلی تمامی جوامع و تعداد گره‌های انتخابی است.

$$\frac{V}{T} \times K \quad (3)$$

با توجه به گرد کردن مقدار عددی به دست آمده از رابطه (۳)، مقدار کلی تمامی گره‌ها برابر K نمی‌شود. از این رو، برای حل این مشکل، از الگوریتم انتخاب جوامع (۳) استفاده کردیم.



شکل ۱. فلوجارت الگوریتم پیشنهادی

آزمون الگوریتم پیشنهادی

در جدول (۱)، شش شبکه اجتماعی برای آزمون الگوریتم پیشنهادی به کار گرفته شده است، که حاوی اطلاعاتی مانند تعداد گره‌ها، یال‌ها و زمان اجرا است. آزمون‌ها با استفاده از سیستم‌های مختلفی انجام شده است. ولی نتایج به دست آمده در جدول (۲) با استفاده از سیستمی با مشخصات پردازنده با سرعت ۲٫۲ گیگاهرتز و حافظه اصلی ۴ گیگابایت در محیط متلب انجام شده است. نتایج حاصل نشان می‌دهد که الگوریتم پیشنهادی سطح کارایی را به طور قابل ملاحظه‌ای افزایش داده است. همچنین، به دلیل به کارگیری خصوصیات شبکه، این روش برای شبکه‌های اجتماعی بزرگ‌تر بسیار مناسب است.

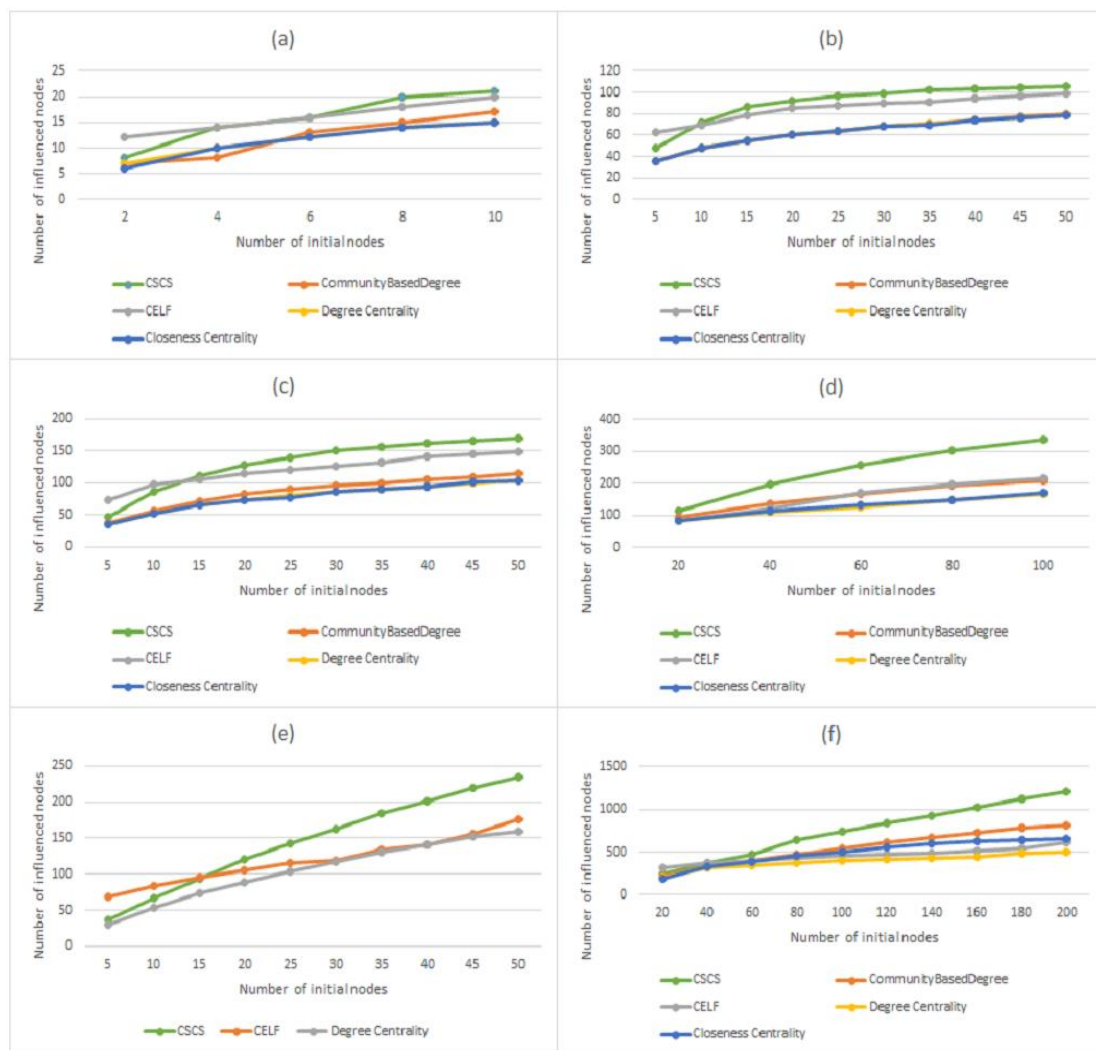
برای مقایسه نتایج و ارزیابی الگوریتم پیشنهادی از چندین الگوریتم و معیار استفاده شده است، که عبارتند از:

- **مرکزیت درجه:** انتخاب گره‌هایی با درجه‌های بالاتر (مرکزیت درجه بالاتر)
- **مرکزیت نزدیکی استاندارد:** مرکزیت نزدیکی نشان‌دهنده نزدیکی یک گره به دیگر گره‌ها است. مرکزیت نزدیکی استاندارد براساس همبندی گراف شبکه اجتماعی محاسبه می‌شود، برای گراف‌های تصادفی تُنک شبکه‌های اجتماعی کاربردی ندارد.
- **درجه مبتنی بر اجتماع^۱ [۲۸]:** این الگوریتم در سال ۲۰۱۱ معرفی شد. از معایب این الگوریتم این است که برای گراف‌های ناهمبند مناسب نیست.
- **الگوریتم CELF [۲۰]:** این الگوریتم در سال ۲۰۰۷ ارائه شد. بررسی‌ها نشان می‌دهد که این روش بسیار زمان‌بر است.

جدول ۲. شرح مجموعه داده‌های مختلف به کار گرفته شده در الگوریتم پیشنهادی

دفعات اجرا توسط الگوریتم‌های مختلف					تعداد گره‌ها	تعداد یال‌ها	شبکه‌های اجتماعی
Degree Centrality	Closeness Centrality	CSCS	Community Based-Degree	CELF			
1000	1000	1000	1000	100	34	79	Zachary's Karate Club [29]
1000	1000	1000	1000	100	113	2163	Hypertext 2009 dynamic contact network [30]
1000	1000	1000	1000	50	412	9428	Sample Facebook dataset 1 (Largest Connected Component) ang, Song, Cong, Wang, Si, & Xie, 2011)
1000	-	1000	-	50	500	1257	Sparse Random Graph (p=0.005)
100	100	100	100	10	4039	88234	Sample Facebook Dataset 2 [31]

آزمون‌ها با استفاده از دو مدل، آستانه خطی و روش آبخاری مستقل، انجام شده است. چنان‌که در بخش‌های قبل گفته شد، برای به دست آوردن وزن یال‌ها از مدل آستانه خطی استفاده کردیم. با توجه به نتایج الگوریتم اکتشافی CSCS فقط با الگوریتم درجه مبتنی بر اجتماع مقایسه شده و از بقیه الگوریتم‌ها صرف نظر شده است. برای تقسیم شبکه به جوامع مجزا از الگوریتم CSCS به کار گرفته شد، با این وجود هیچ محدودیتی در به کارگیری دیگر الگوریتم‌های شناسایی جوامع وجود ندارد. یکی از مزایای این الگوریتم عدم وجود هیچ وابستگی به نوع الگوریتم یافتن جامعه است، که می‌تواند گره‌های مؤثر را براساس معیار خود شناسایی کند. برای ارزیابی الگوریتم CSCS، دو مدل انتشار متفاوت، مدل آبخاری مستقل و مدل آستانه خطی، به کار گرفته شده است. شکل (۲) و (۳) نتایج آزمون‌های انجام شده روی شبکه‌های اجتماعی مختلف را نشان می‌دهد. براساس نتایج نشان داده شده در شکل (۲)، تعداد گره‌های اولیه روی تعداد گره‌های مؤثر در الگوریتم‌های مختلف تأثیر می‌گذارد. مدل انتشار آنها مدل آستانه خطی است. واضح است که در بین الگوریتم‌های مختلف که روی شبکه اعمال شده است، الگوریتم CSCS با افزایش گره‌های اولیه نتیجه بهتری داشته است و می‌تواند که روی افراد بیش‌تری تأثیر بگذارد. در شکل (۳)، نتایج الگوریتم CSCS با الگوریتم community Degree-based مقایسه شده است. الگوریتم community Degree-based از مدل آبخاری مستقل استفاده می‌کند. با توجه به نمودارهای زیر، نتایج تقریباً یکسان است و هر دو الگوریتم کاملاً شبیه هم عمل می‌کنند.

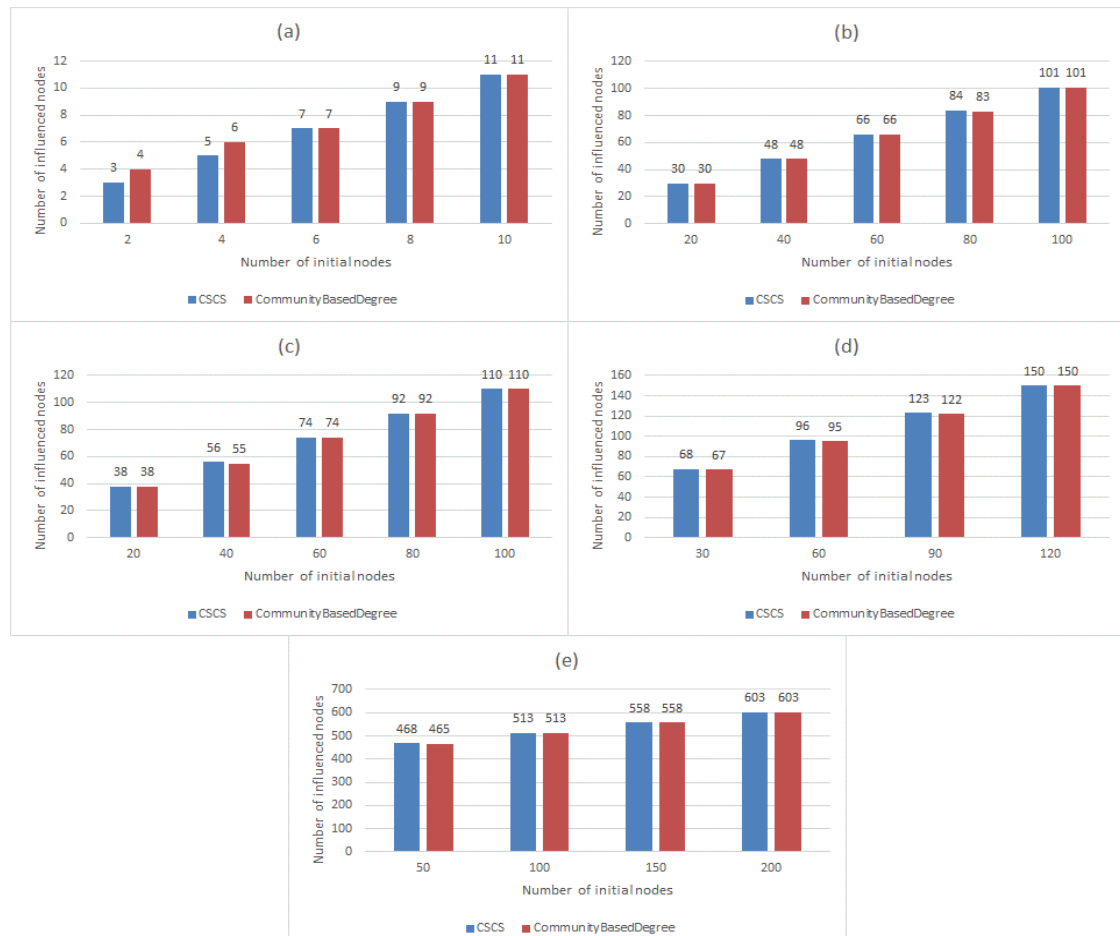


شکل ۲. آزمایش‌های انجام شده با استفاده از مدل آستانه خطی روی شبکه‌های اجتماعی
 a) Zachary's Karate Club, (b) Hypertext, (c) Jazz musicians network, (d) Sample Facebook dataset 1, (e) Random Sparse Graph, (f) Sample Facebook dataset 2

مقایسهٔ زمان اجرای الگوریتم‌ها به‌گونه‌ای انجام شده است که ابتدا ۱۰ گره از شبکه اجتماعی Jazz به‌عنوان گره‌های مؤثر انتخاب می‌شود. در حین انتخاب این ۱۰ گره، برخی محاسبات مانند مقدار شاپلی، مرکزیت درجه و مرکزیت نزدیکی، نیز اجرا می‌شود. سپس این محاسبات را می‌توان برای اجرای بعدی ذخیره کرد. در اجرای بعدی، ۲۰ گره بانفوذ و تأثیرگذار از همین شبکه انتخاب، و از محاسبات قبلی استفاده می‌شود. جدول ۳، زمان میانگین اجراهای الگوریتم را نشان می‌دهد.

جدول ۳. مقایسهٔ زمان اجرای الگوریتم در شبکه اجتماعی Jazz

Running time of algorithms (Millisecond)						
CSCS	CELF	Degree Centrality	Community Based-Degree	Closeness Centrality	select k-nodes	
16	83725	1	16	3214	K=10	
1	84007	1	16	1	K=20	



شکل ۳. آزمایش‌های انجام شده با استفاده از مدل مستقل آشنایی روی شبکه‌های اجتماعی با $P=0.01$: a) Zachary's Karate Club, b) Hypertext, c) Jazz musicians network, d) Sample Facebook dataset 1, e) Sample Facebook dataset 2

نتیجه‌گیری

در این پژوهش، الگوریتم اکتشافی نوینی تحت عنوان CSCS برای بررسی و حل مسئله پیشینه‌سازی نفوذ کاربران در شبکه‌های اجتماعی، ارائه شده است. این مسئله بر تأثیرگذاری فعالیت مجموعه بزرگی از افراد درون شبکه اجتماعی تمرکز دارد و با استفاده جوامع و مرکزیت مقدار شاپلی حل می‌شود.

این الگوریتم در یک شبکه، افراد با نفوذ و تأثیرگذار را با به‌کارگیری ویژگی‌های آنها شناسایی خواهد کرد. با مقایسه نتایج به‌دست آمده با دیگر الگوریتم‌هایی مانند، CELF و الگوریتم‌های حریصانه، که سطح انتشاری برای هر گره در نظر می‌گیرند، و زمان اجرای بسیار طولانی دارند، دریافتیم که نمی‌توان این الگوریتم‌ها را برای شبکه‌های بزرگ به‌کار برد، ولی الگوریتم CSCS زمان اجرا و کارایی را به‌طور چشم‌گیری بهبود بخشیده است. بنابراین می‌توان این الگوریتم را بر روی شبکه‌ها بزرگ‌تر اعمال کرد.

علاوه بر این، برای ارزیابی بیش‌تر الگوریتم CSCS، این الگوریتم را روی شش شبکه اجتماعی مختلف اعمال کردیم و نتایج آن را با چهار الگوریتم معیار دیگر مقایسه کردیم. نتایج حاصل از آزمون‌ها نشان می‌دهد که در مدل آستانه خطی، با افزایش تعداد نقاط اولیه، الگوریتم CSCS به توان عملیاتی قابل قبولی دست پیدا می‌کند و گره‌های

بیش‌تری را تحت تأثیر قرار می‌دهد. همچنین، در مدل آبخاری مستقل، توان عملیاتی الگوریتم CSCS اغلب به الگوریتم درجه مبتنی بر اجتماع شباهت دارد، و حتی در برخی مواقع بهتر نیز عمل می‌کند.

منابع

1. Wallis W., "Graph Theory with Applications (JA Bondy and USR Murty)", Society for Industrial and Applied Mathematics (1979).
2. Fu P., Zhu A., Ni H., Zhao X., Li X., "Threshold behaviors of social dynamics and financial outcomes of Ponzi scheme diffusion in complex networks", *Physica A: Statistical Mechanics and its Applications*, Vol. 490 (2018) 632-642.
3. Yang Y., Xu Y., Wang E., Lou K., Luan D., "Exploring influence maximization in online and offline double-layer propagation scheme", *Information Sciences*, Vol. 450 (2018) 182-199.
4. Narayanam R., Narahari Y., "A shapley value-based approach to discover influential nodes in social networks", *IEEE Transactions on Automation Science and Engineering*, Vol. 8, No. 1, (2011) 130-147.
5. Granovetter M., "Threshold models of collective behavior", *American journal of sociology*, Vol. 83, No. 6 (1978) 1420-1443.
6. Riquelme F., Gonzalez-Cantergiani P., Molinero X., Serna M., "Centrality measure in social networks based on linear threshold model", *Knowledge-Based Systems*, vol. 140 (2018) 92-102.
7. Zhang H., Mishra S., Thai M. T., Wu J., Wang Y., "Recent advances in information diffusion and influence maximization in complex social networks", *Opportunistic Mobile Social Networks*, Vol. 37, No. 1.1 (2014).
8. Scott J., Carrington P. J., "The SAGE handbook of social network analysis", SAGE publications (2011).
9. Wasserman S., Faust K., "Social network analysis: Methods and applications", Cambridge university press (1994).
10. Radicchi F., Castellano C., Cecconi F., Loreto V., Parisi D., "Defining and identifying communities in networks", *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 101, No. 9 (2004) 2658-2663.
11. Gibbons R., "A primer in game theory", Harvester Wheatsheaf (1992).
12. Schmidt C., "Game theory and economic analysis: a quiet revolution in economics", Routledge (2003).
13. Osborne M. J., Rubinstein A., "A course in game theory", MIT press (1994)

14. Morton D. D., "Game theory, A nontechnical introduction", Dover publications, Inc., Mineola, NewYork (1983).
15. Shapley L., "A Value for N Person Games", RAND P 295, Santa Monica, CA: Rand Corporation (1952).
16. Chalkiadakis G., Elkind E., Wooldridge M., "Computational aspects of cooperative game theory," *Synthesis Lectures on Artificial Intelligence and Machine Learning*, Vol. 5, No. 6 (2011) 1-168,.
17. Michalak T. P., Aadithya K. V., Szczepanski P. L., Ravindran B., Jennings N. R., "Efficient computation of the Shapley value for game-theoretic network centrality", *Journal of Artificial Intelligence Research*, Vol. 46 (2013) 607-650.
18. Domingos P., Richardson M., "Mining the network value of customers", in *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining* (2001) 57-66.
19. Kempe D., Kleinberg J., Tardos E., "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) 137-146.
20. Leskovec J., et al., "Cost-effective outbreak detection in networks", in *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*, (2007) 420-429.
21. Pandit S., Yang Y., Chawla N. V., "Maximizing information spread through influence structures in social networks," in *2012 IEEE 12th International Conference on Data Mining Workshops* (2012) 258-265.
22. Zhou S., Yue K., Fang Q., Zhu Y., Liu W., "An efficient algorithm for influence maximization under linear threshold model", in *The 26th Chinese Control and Decision Conference (2014 CCDC)* (2014) 5352-5357.
23. Sinha N., Annappa B., "Cuckoo Search for Influence Maximization in Social Networks", in *Proceedings of 3rd International Conference on Advanced Computing, Networking and Informatics*(2016) 51-61.
24. Yang Y.-C., "A NewGreedy Genetic Algorithm for Influence Maximization in Social Network" (2016).
25. Gong M., Yan J., Shen B., Ma L., Cai Q., "Influence maximization in social networks based on discrete particle swarm optimization", *Information Sciences*, Vol. 367 (2016) 600-614.
26. Jiang Q., et al., "Simulated Annealing Based Influence Maximization in Social Networks", in *AAAI*, Vol. 11 (2011) 127-132.

27. Cui L., et al., "DDSE: A novel evolutionary algorithm based on degree-descending search strategy for influence maximization in social networks", *Journal of Network and Computer Applications*, Vol. 130 (2018) 119-130.
28. Anjerani M., Moeini A., "Selecting influential nodes for detected communities in real-world social networks", in *Electrical Engineering (ICEE), 2011 19th Iranian Conference on*. IEEE, (2011) 1-6.
29. Zachary W. W., "An information flow model for conflict and fission in small groups", *Journal of anthropological research* (1977) 452-473.
30. Isella L., et al., "What's in a crowd? Analysis of face-to-face behavioral networks", *Journal of theoretical biology*, Vol. 271, No. 1 (2011) 166-180.
31. Leskovec J., McAuley J., "Learning to discover social circles in ego networks", in *Advances in neural information processing systems* (2012) 539-547.