

مروری بر معیارهای مختلف مرکزیت در شبکه‌های اجتماعی و چالش‌های موجود

کبری رحمتی^۱، حسن نادری^۲، آسیه قنبرپور^۳

^۱دانشجوی کارشناسی ارشد، دانشگاه علم و صنعت ایران، تهران،
K_Rahmati@comp.iust.ac.ir

^۲استادیار، دانشگاه علم و صنعت ایران، تهران،
naderi@iust.ac.ir

^۳دانشجوی دکتری، دانشگاه علم و صنعت ایران، تهران،
ghanbarpour@iust.ac.ir

چکیده

چگونگی محاسبه اهمیت گره به منظور پیدا کردن گره مهم در شبکه یک مساله اساسی است. در زمینه‌ی شبکه‌های اجتماعی، اهمیت گره‌ها به وسیله‌ی معیارهای مرکزیت محاسبه می‌شود. کلمه اهمیت، معانی گسترده‌ای دارد که منجر به تعاریف متعددی برای مرکزیت می‌شود و در محیط‌ها و کاربردهای متفاوت به روش‌های مختلفی محاسبه می‌شود. در واقع، مرکزیت یک معیار کمی با هدف آشکار-سازی اهمیت گره‌ها است. هر کدام از پارامترهای مرکزیت در شبکه‌های اجتماعی نقش خاصی دارند. در واقع، این معیارها مبتنی بر جنبه‌های متفاوتی از خصوصیات گره‌ها هستند و اغلب با یکدیگر در تضاد هستند. هر شبکه با توجه به روابط موجود در شبکه از خصوصیات خاصی برخوردار است. به دلیل این تفاوت‌ها در شبکه‌های اجتماعی، شاخص‌های ارائه شده و نتیجه‌گیری‌ها همواره برای همه‌ی انواع شبکه‌ها صادق نیستند. در نتیجه، با توجه به نوع مسئله، نحوه انتشار و همچنین احتمال تأثیرگذاری، بهترین معیار در هر محیط متفاوت است. داشتن دانش در مورد انواع مرکزیت برای انتخاب یک مرکزیت مناسب به منظور حل یک مسئله جاری نوعی از حداقل‌ها می‌باشد. به همین منظور در این مقاله انواع شاخص‌های مرکزیت مورد بررسی قرار خواهند گرفت.

کلمات کلیدی

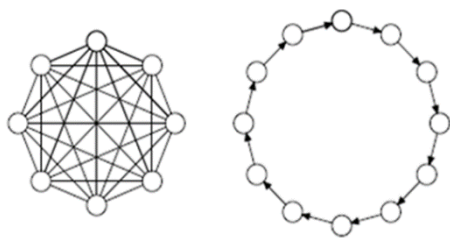
معیار مرکزیت، اهمیت گره، گلوگاه، گره، گراف.

به طور کلی، ارتباط اعضای مختلف یک شبکه‌ی اجتماعی به طور قابل توجهی متفاوت است. به منظور شناسایی کسانی که نقش محوری در یک شبکه اجتماعی بازی می‌کنند استفاده از معیارهای مرکزیت^۲ مناسب است. بنابراین در این مقاله وضعیت تحقیقات انجام شده تاکنون در مورد معیارهای مرکزیت در شبکه‌های پیچیده همچون شبکه‌های اجتماعی بررسی می‌شود. در بخش ۲ سه اصل مرکزیت بیان خواهند شد. در بخش ۳ یک دسته بندی سطح بالا برای معیارهای مرکزیت ارائه داده‌ایم. در بخش ۴ لیست تقریباً کاملی از حدود نوزده مهمترین معیار مرکزیت توضیح داده خواهند شد.

۲- اصول مرکزیت

از آنجا که هر کدام از شاخص‌ها اهداف متفاوتی را ارائه می‌دهند، تعریف مجموعه اصولی که تعیین کند مرکزیت چیست، معقول نیست. در واقع، معقول است که اصول ضروری که یک شاخص باید برای رفتار قابل پیش‌بینی رعایت کند را تعیین کنیم. تعریف چنین اصولی یک موضوع حساس و مهم است. در تعریف این اصول فرض می‌کنیم که معیارهای مرکزیت به برچسب خاص انتخاب شده برای هر گره وابسته نیستند بلکه فقط به ساختار گراف وابسته هستند **Error! Reference source not found.**

اگر یک گراف k -clique و یک گراف p -cycle را مانند شکل (۳) در نظر بگیرید، اصول مرکزیت به صورت زیر قابل تعریف است:



شکل (۲): نمونه گراف‌های clique و **Error! Reference source not found.**[3]

تعریف ۱ (اصل اندازه): گراف $S_{k,p}$ را در نظر بگیرید که از یک k -clique و یک p -cycle ساخته شده است. یک معیار مرکزیت، اصل اندازه را رعایت می‌کند اگر برای هر k یک P_k وجود داشته باشد به طوری که برای هر $p \geq P_k$ مرکزیت یک گره در p -cycle به شدت بزرگ‌تر از مرکزیت یک گره در k -clique است و اگر برای هر p یک k وجود داشته باشد به طوری که برای هر $k \geq K_p$ در $S_{k,p}$ مرکزیت یک گره در k -clique به شدت بزرگ‌تر از مرکزیت گره در p -cycle است. [3]

تعریف ۲ (اصل تراکم): گراف $D_{k,p}$ را که از یک k -clique و یک p -cycle ساخته شده و به وسیله‌ی یک پل دو طرفه $x \leftrightarrow y$ متصل شده‌اند را در نظر بگیرید، به طوری که x یک گره در دسته و y یک گره در چرخه است. یک معیار مرکزیت اصل تراکم را ارضا می‌کند اگر برای $k = p$ مرکزیت x به شدت بزرگ‌تر از مرکزیت y باشد.

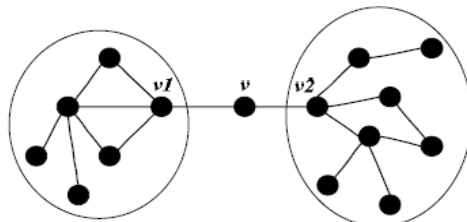
تعریف ۳ (اصل یکنواختی-امتیاز): یک معیار مرکزیت اصل یکنواختی امتیاز را رعایت می‌کند اگر برای هر گراف G و هر جفت گره x و y به طوری که $x \leftrightarrow y$ و وقتی که $x \rightarrow y$ را به G اضافه می‌کنیم، مرکزیت

۱- مقدمه

با مطالعه‌ی عمیق شبکه‌های پیچیده، بررسی قابلیت اطمینان و بقای شبکه‌های پیچیده به یک موضوع مهم پژوهشی تبدیل شده است. تحلیل شبکه‌های پیچیده یک کار بسیار مهم در سیستم شبکه برق، مخابرات، امنیت ملی، جامعه شناسی و غیره است. دو جنبه‌ی مهم این تحلیل‌ها عبارتند از چگونگی ایجاد و پایداری شبکه‌ها. تحقیقات نشان می‌دهد که وضعیت گره‌ها در شبکه‌ها یکسان نیست و تفاوت زیادی بین اهمیت آن‌ها وجود دارد. در عین حال، گره‌ها در وضعیت‌های مختلف، سطوح مختلفی از نفوذ در بقای شبکه دارند. به طور کلی، گره‌های مهم نقش کلیدی در عملکرد شبکه بازی می‌کنند. در نتیجه، بررسی تاثیر اهمیت گره در شبکه جهت تحلیل گره‌ها از دیدگاه‌های مختلف از جمله مراقبت و اهمیت به شدت مورد توجه است و به طور گسترده‌ای مورد بررسی قرار می‌گیرد، زیرا مربوط به موضوعات مهمی مثل تعیین گلوگاه‌ها، بررسی آسیب پذیری و یا انعطاف پذیری شبکه در برابر حملات، ایمن‌سازی در برابر بیماری‌های همه‌گیر، هماهنگ‌سازی شبکه، ناوبری ترافیک تشخیص مکان پل‌ها و غیره می‌باشد. در واقع، زمانی که گره (یا گره‌های) هسته مورد حمله قرار می‌گیرد، استحکام شبکه به طور جدی تحت تاثیر قرار خواهد گرفت. در نتیجه، با شناسایی گره‌های مهم و در نظر

گرفتن استراتژی‌های حفاظتی قبل از حمله، می‌توان به قابلیت اطمینان و بقای شبکه کمک کرد و ضریب فعال بودن شبکه را افزایش داد. بنابراین ارزیابی اهمیت گره در شبکه‌های پیچیده نه تنها مفید بلکه حیاتی است. در واقع، چالش اصلی برای تحلیل شبکه‌های اجتماعی، شناسایی افراد کلیدی در شبکه اجتماعی است.

در شبکه‌های پیچیده بزرگ، همه‌ی گره‌ها معادل نیستند. به عنوان مثال اگر گره در یک بن‌بست باشد، حذف آن هیچ تاثیری بر روی کل شبکه ندارد. در حالی که در حالت متضاد، با برش گره‌ی مانند گره v در شکل (۱)، این حذف ممکن است باعث قطع مؤلفه‌های شبکه شود. در زمینه‌ی شبکه‌های اجتماعی، اهمیت گره‌ها به «مرکزیت»^۱ وابسته است. با توجه به این که همه‌ی این تعریف‌ها بسیار متفاوت هستند، وقتی افراد با برخی کاربردهای خاص مواجه می‌شوند، ممکن است زیرمجموعه‌ای از این تعریف‌ها برای آن‌ها مفید باشد **Error! Reference source not found.**[1]



شکل (۱): اهمیت گره v در قطع گراف **Error! Reference source not found.**[1]

هر معیار در یک شبکه به منظور مقابله با یک دسته از مسائل، بر اساس یک معیار مجزا ارائه شده است. با این حال، گاهی اوقات، باید اهمیت گره را به طور کامل بر اساس معیارهای مختلف در نظر بگیریم **Error! Reference source not found.**[2][2]

$$C_D(v) = \text{deg}(v) \quad (1)$$

نقطه ضعف عمده DC این است که تماس‌های غیر مستقیم اصلاً در نظر گرفته نمی‌شوند. **Error! Reference source not found.**

۴-۲- معیار مرکزیت نزدیکی^۵

طبق این معیار، گرهی که مرکزیت نزدیکی بیشتری داشته باشد، گره مرکزی تر و مهم‌تری در شبکه خواهد بود. در واقع، اگر گره در مرکزی‌ترین مکان شبکه قرار گرفته باشد، اطلاعات از طریق آن در کوتاه‌ترین زمان در کل شبکه منتشر خواهند شد **Error! Reference source not found.** در واقع گرهی به‌عنوان گره مرکزی در نظر گرفته می‌شود که برای ارتباط با دیگر گره‌ها به تعداد کمی واسط نیاز داشته باشد **Error! Reference source not found.** فرض کنید $d(i,j)$ طول کوتاه‌ترین مسیر از گره v_i به گره v_j باشد. معادله مرکزیت نزدیکی به‌صورت زیر است:

$$C_C(v_i) = (N - 1) / \sum_{j=1}^N d(i, j) \quad (2)$$

بزرگ‌ترین مقدار $C_C(v_i)$ ، یعنی v_i نزدیک‌ترین گره به مرکز شبکه است، که نشان می‌دهد گره v_i یک موقعیت مهم در شبکه دارد. برای این معیار گراف باید قویاً متصل باشد. در غیر این صورت، برخی از مخرج‌ها ∞ خواهند بود و در نتیجه برای همه‌ی گره‌هایی که قابل دسترس نیستند امتیاز صفر به دست می‌آید [3]. روشی که برای استفاده از مرکزیت نزدیکی در گراف‌های غیر متصل پیشنهاد شده است، فقط در نظر گرفتن فاصله‌های کمتر از ∞ می‌باشد، به صورت زیر:

$$\frac{1}{\sum_{d(y,x) < \infty} d(y,x)} \quad (3)$$

در این روش مرکزیت گره‌هایی که هیچ گرهی به آن‌ها دسترسی ندارد برابر صفر در نظر گرفته می‌شود. ولی مشکلی که این روش دارد این است که به سمت گره‌هایی که تعداد گره‌های کمتری به آن‌ها دسترسی دارند، جهت‌گیری شدیدی دارد.

۴-۳- شاخص لین^۶

شاخص لین تعریف مرکزیت نزدیکی را برای گراف‌هایی با فاصله‌ی بینهایت اصلاح می‌کند. این شاخص با استفاده از مربع تعداد گره‌هایی که به گره دسترسی دارند، به گره‌هایی که تعداد بیشتری گره به آن‌ها دسترسی دارند اهمیت بیشتری می‌دهد، تعریف لین برای مرکزیت گره x عبارت است از:

$$\frac{| \{y | d(y,x) < \infty\} |^2}{\sum_{d(y,x) < \infty} d(y,x)} \quad (4)$$

منطق پشت این تعریف به این صورت است که ابتدا، فرض می‌کنیم مرکزیت نزدیکی معکوس متوسط فاصله‌ها است، که مستلزم یک ضرب در تعداد گره‌هایی است که به گره دسترسی دارند. این تغییر، مرکزیت نزدیکی را در گراف نرمال‌سازی می‌کند. در واقع، می‌خواهیم به گره‌هایی که گره‌های بیشتری به آن‌ها دسترسی دارند اهمیت بیشتری داده شود، در نتیجه دوباره در تعداد گره‌هایی که به گره دسترسی دارند ضرب می‌کنیم. همچنین، گره‌هایی که هیچ گره‌ای به آن‌ها دسترسی ندارد دارای مرکزیت ۱ هستند. شاخص لین

افزایش می‌یابد. این اصل اساساً توسط همه‌ی معیارهای مرکزیت بر روی گراف‌های متصل قوی رعایت می‌شود [3]. طبق بررسی‌های انجام شده بر روی معیارهای مرکزیت از نظر رعایت اصول مرکزیت، باید گفت فقط مرکزیت هارمونیک همه‌ی اصول فوق را رعایت می‌کند [3].

۳-۳- دسته‌بندی معیارهای مرکزیت

معیارهای مرکزیت را می‌توان به سه دسته‌ی کلی دسته‌بندی کرد، که در ادامه توضیح داده می‌شوند.

۳-۱- معیارهای هندسی

در معیارهای هندسی، اهمیت تابعی از فاصله است، به‌طور دقیق‌تر، اهمیت هندسی فقط به تعداد گره‌های موجود در هر فاصله بستگی دارد. برخی معیارهای هندسی عبارتند از: مرکزیت درجه، مرکزیت نزدیکی، مرکزیت هارمونیک، شاخص لین.

۳-۲- معیارهای طیفی

معیارهای طیفی، بردار ویژه مسلط چپ^۳ ماتریس به‌دست‌آمده از گراف را محاسبه می‌کنند و بر اساس اینکه ماتریس چگونه قبل از محاسبات اصلاح می‌شود، می‌توانیم تعدادی معیار مختلف به دست آوریم. برخی معیارهای

طیفی عبارتند از: بردار ویژه مسلط چپ، شاخص سلی، شاخص کاتز، PageRank، HITS، SALSA.

۳-۳- معیارهای مبتنی بر مسیر

معیارهای مبتنی بر مسیر در واقع همه‌ی مسیرها یا کوتاه‌ترین مسیرهای وارد شونده به یک گره را بررسی می‌کنند. درجه‌ی ورودی می‌تواند به عنوان یک معیار مبتنی بر مسیر نیز در نظر گرفته شود، به‌طوری که معادل با تعداد کوتاه‌ترین مسیرهای ورودی به طول یک است. با یک دید کلی همه‌ی معیارهای طیفی می‌توانند به‌عنوان معیارهای مبتنی بر مسیر در نظر گرفته شوند. معیار مرکزیت بینایی، یک نوع معیار مبتنی بر مسیر می‌باشد.

۴- انواع معیارهای مرکزیت

در این بخش انواع معیارهای مرکزیت و ویژگی‌های آن‌ها مورد بررسی قرار می‌گیرد.

۴-۱- معیار مرکزیت درجه^۴

اولین و ساده‌ترین مفهوم مرکزیت، درجه است، که به‌عنوان تعداد لینک‌های حادث بر یک گره تعریف می‌شود. در یک شبکه‌ی جهت‌دار، معمولاً دو معیار متمایز از مرکزیت درجه تعریف شده است، که عبارتند از درجه‌ی ورودی و درجه‌ی خروجی. مرکزیت درجه رأس v ، برای گراف $G=(V,E)$ با $|V|$ رأس و $|E|$ یال، به‌صورت زیر تعریف می‌شود:



برای ارائه‌ی یک راه‌حل برای مسائل ناشی از تعریف مرکزیت نزدیکی مناسب به نظر می‌رسد [7].

$$C_E(v_i) = v_x = \frac{1}{\lambda_{\max}(A)} \cdot \sum_{j=1}^N a_{jx} \cdot v_j \quad (4)$$

مرکزیت بردار ویژه رأس x به صورت زیر تعریف می‌شود:

$v = (v_1, v_2, \dots, v_n)^T$ به یک بردار ویژه برای مقدار ویژه
Error! ماکسیم $\lambda_{\max}(A)$ ماتریس همسایگی A اشاره می‌کند
Error! Reference source not found.
Error! [10]not found.

۴-۷- معیار مرکزیت گریز از مرکز^{۱۰}

مفهوم گریز از مرکز، مفهوم حداکثر فاصله بین جفت گره‌ها را در نظر می‌گیرد:

$$C_e(v_i) = \frac{1}{\max_{j \in V} d(i, j)} \quad (8)$$

یک گره مرکزی است اگر هیچ گره‌ای دور از آن نباشد. این معیار مرکزیت در واقع مختص گراف‌های متصل است و گره‌های مهم در انتشار اطلاعات را مشخص می‌کند [10][10].

۴-۸- معیار مرکزیت نفوذ^{۱۱}

مرکزیت نفوذ با این ایده ارائه شد که به طور خاص اهمیت گره از نظر کمک به نفوذ در شبکه را اندازه‌گیری کند. مرکزیت نفوذ برای یک گره در یک زمان، به عنوان احتمال مسیر نفوذ^{۱۲} از طریق گره تعریف می‌شود. یک مسیر نفوذ، کوتاه‌ترین مسیر بین یک جفت گره است، که به گره مبدأ نفوذ شده است. گره هدف می‌تواند در وضعیت نفوذ شده، نفوذ نشده و نفوذ جزئی باشد.

$$C_P^t(v) = \frac{1}{N-2} \sum_{s \neq v \neq r} \frac{\sigma_{sr}(v) \cdot x_s^t}{\sigma_{sr} \cdot \sum [x_i^t] - x_v^t} \quad (9)$$

به طوری که σ_{sr} تعداد کل کوتاه‌ترین مسیرها از گره s به گره r است و $\sigma_{sr}(v)$ تعدادی از این مسیرها است که از v می‌گذرد. وضعیت نفوذ گره i در زمان t با x_i^t نشان داده می‌شود و دو حالت خاص دارد، $x_i^t = 0$ که نشان دهنده‌ی وضعیت نفوذ نشده در زمان t است و $x_i^t = 1$ که نشان دهنده

ی وضعیت نفوذ کامل در زمان t است. مقادیر بین صفر و یک وضعیت نفوذ جزئی را نشان می‌دهد 00.

۴-۹- معیار مرکزیت cross-clique

مرکزیت cross-clique یک گره، در یک گراف پیچیده اتصال گره به دسته‌های مختلف را مشخص می‌کند. یک گره با اتصال cross-clique بالا انتشار اطلاعات یا بیماری در یک گراف را تسهیل می‌کند. دسته‌ها زیر گراف‌هایی هستند که در آن‌ها، هر گره به همه‌ی گره‌های دیگر موجود در آن متصل است. اتصال cross-clique گره v در گراف (V, E) ، به صورت $X(v)$ تعریف می‌شود به طوری که $X(v)$ تعداد دسته‌هایی است که رأس v به آن‌ها متعلق است [12][12].

۴-۱۰- معیار مرکزیت تجزیه K-shell

تجزیه k-shell به طور بازگشتی گره‌هایی که درجه کمتر یا مساوی k دارند را هرس می‌کند. اگر فرض کنیم $S = (G, E|G)$ ، زیر شبکه‌ی حاصل از تجزیه

۴-۸- معیار مرکزیت هارمونیک^۷

مرکزیت هارمونیک تصحیح شده‌ی مرکزیت نزدیکی سنتی است، که برای دسترسی به گره‌های غیرقابل دسترس طراحی شده است. علاوه بر این، مرکزیت هارمونیک به عنوان یک شاخص مرکزیت همه‌منظوره‌ی عالی برای گراف‌های جهت‌دار دلخواه استفاده می‌شود. در واقع این معیار برای مقابله با مشکل مرکزیت نزدیکی، برای یک شبکه جهت‌دار عمومی میانگین فاصله‌ها را با میانگین هارمونیک همه‌ی فاصله‌ها جایگزین می‌کند. بنابراین مرکزیت هارمونیک x را با این فرض اینکه $0 = \infty^{-1}$ به صورت زیر تعریف می‌کنیم:

$$\sum_{y \neq x} \frac{1}{d(y, x)} \quad (5)$$

ممکن است تفاوت این فرمول با فرمول (۲) جزئی به نظر برسد، اما در واقع این یک تغییر اساسی است. مرکزیت هارمونیک همبستگی قوی با مرکزیت نزدیکی در شبکه‌های ساده دارد. اما، می‌تواند برای اعمال روی گراف‌هایی که متصل قوی نیستند سودمند باشد [8][8][8].

۴-۵- معیار مرکزیت بینابینی^۸

معیار مرکزیت بینابینی تعداد دفعاتی که یک گره به عنوان پل در طول کوتاه‌ترین مسیر بین دو گره دیگر عمل می‌کند را مشخص می‌کند و به عنوان معیاری برای تعیین کمیت کنترل یک فرد روی ارتباط موجود بین دیگر افراد در یک شبکه معرفی شده است. در BC یک گره که متصل‌تر باشد در بسیاری از کوتاه‌ترین مسیرهای ممکن بین جفت گره‌های دیگر وجود خواهد داشت.

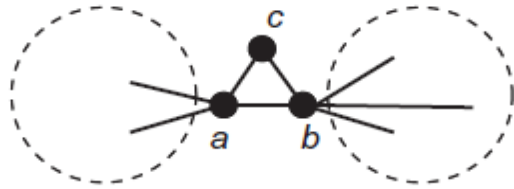
$$C_B(v_i) = \sum_{i \neq j \neq r \in V} \frac{k_{jr}(v_i)}{k_{jr}} \quad (6)$$

k_{jr} تعداد کوتاه‌ترین مسیرها از گره j به گره r در شبکه است، $k_{jr}(v_i)$ تعداد کوتاه‌ترین مسیرها از طریق گره v_i است. مرکزیت بینابینی یک گره، ظرفیت بار آن گره بین گره‌های دیگر است و مقدار بینابینی بیشتر، تأثیرگذاری بیشتر گره v_i در شبکه را نشان می‌دهد **Error! Reference source not found.** [13]، **Error! Reference source not found.** [12]. BC یکی از معیارهای استاندارد مرکزیت گره است، که در اصل برای تعیین اهمیت یک گره (فرد) در یک شبکه (اجتماعی) معرفی شده است. گره‌هایی با BC بالا در اتصال و انتشار اطلاعات مهم هستند [9]. BC مناسب‌ترین مرکزیت برای تعیین اهمیت گره و تأثیر آن بر روی اتصال کلی گراف است [10][10]. یکی از معایب BC این است که فقط گره‌هایی را در نظر می‌گیرد که درگیر کوتاه‌ترین مسیرها هستند [10][10].

۴-۶- معیار مرکزیت بردار ویژه^۹

ایده اصلی مرکزیت بردار ویژه این است که اهمیت یک گره نه تنها به وسیله خودش تعیین می‌شود (خصوصیت خودش است)، بلکه اغلب تحت تأثیر خصوصیت اهمیت همسایگانش است. یک گره که به گره‌های مهم متصل شده باشد، بر اهمیت خودش نیز تحت تأثیر این ارتباط افزوده خواهد شد. **Error! Reference source not found.** [5]. با این ایده،

از هر گره سه هر گره دیگر است. random walk betweenness. برابر است با تعداد دفعاتی که، به طور متوسط با تمامی جفت‌های ممکن (s,t)، یک پیاده‌روی تصادفی شروع شده از s و پایان یافته در طول راه از t عبور می‌کند. مشکل این روش این است که نیاز به دانش عمومی دارد و این مانع استفاده از آن در یک محیط توزیع شده می‌شود [14][14][10].



شکل (۴): یک نمونه که در آن betweenness centrality علامت‌دهی اهمیت گره c به آن امتیاز کمی می‌دهد [10][10]

۴-۱۳ - معیار مرکزیت محلی

معیار مرکزیت محلی، هم نزدیک‌ترین همسایه و هم نزدیک‌ترین همسایه‌های بعدی را در نظر می‌گیرد. مرکزیت محلی گره v به صورت زیر تعریف می‌شود، $C_L(v)$:

$$Q(u) = \sum_{w \in \Gamma(u)} N(w) \quad (10)$$

$$C_L(v) = \sum_{u \in \Gamma(v)} Q(u) \quad (11)$$

به طوری که $\Gamma(u)$ مجموعه‌ی نزدیک‌ترین همسایه‌های گره u و N(w) تعداد نزدیک‌ترین و نزدیک‌ترین همسایه‌های بعدی گره w است.

به طور کلی، در مقایسه با سه معیار مرکزیت شناخته شده درجه، نزدیکی و بینایی، معیار مرکزیت محلی عملکرد خیلی بهتری نسبت به مرکزیت‌های درجه و بینایی دارد و تقریباً به خوبی معیار مرکزیت نزدیکی است در حالی که پیچیدگی محاسباتی بسیار کمتری دارد [13].

۴-۱۴ - معیار مرکزیت Second order

مرکزیت second order به روش توزیع شده قابل محاسبه است. این روش

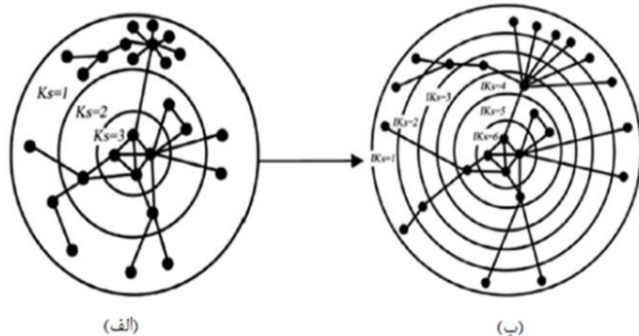
متکی بر ملاقات با پیاده‌روی تصادفی بی طرفانه است و به صورت محلی برای هر گره مقداری را فراهم می‌کند که منعکس کننده اهمیت آن گره است. در واقع هر گره باید از محله‌ی مستقیم خود آگاه باشد و نیاز به هرگونه اطلاعات عمومی ندارد. برای این منظور، هر گره زمان سپری شده بین ملاقات‌های پیاده‌روی تصادفی که زمان بازگشت نامیده می‌شود و انحراف استاندارد زمان‌های بازگشت را ثبت می‌کند.

نکته‌ی کلیدی این است که در پیاده روی تصادفی گره‌های مرکزی در مقایسه با دیگر گره‌ها به طور منظم ملاقات می‌شوند. در واقع، برای تعیین دقیق گره‌های مهم از انحراف استاندارد زمان‌های بازگشت استفاده می‌شود [10][14].

۴-۱۵ - معیار مرکزیت بردار ویژه مسلط چپ

اولین و بارزترین معیار طیفی، بردار ویژه‌ی مسلط چپ ماتریس مجاورت است. در واقع، بردار ویژه‌ی مسلط می‌تواند به عنوان نقطه ثابت یک محاسبه تکرار شونده باشد که در آن هر گره با امتیاز مشابهی شروع می‌شود و سپس آن امتیاز را با مجموع امتیازات گره‌های مقابل خود جایگزین می‌کند. سپس

k-shell روی گراف G است. زیر گراف s، یک k-core از گراف نامیده می‌شود، اگر و فقط اگر درجه‌ی تمامی گره‌های باقیمانده در آن بزرگ‌تر از k باشد و بزرگ‌ترین زیر شبکه باشد. زیر شبکه‌ای که مقدار Ks گره‌های آن برابر با k است، k-shell نامیده می‌شود. یک نمودار از تجزیه-k shell در شکل (۴) (الف) نشان داده شده است، شبکه به ۳ تا shell متفاوت تقسیم شده است، با $Ks=1$ و $Ks=2$ و $Ks=3$. هر گره یک مقدار Ks می‌گیرد. [13][13] Error! Reference source not found.



شکل (۳): مثالی از تجزیه k-shell و تجزیه k-shell

بهبود یافته [13][13]

۴-۱۱ - تجزیه k-shell بهبود یافته

فرض کنید شبکه اولیه G_0 است، ابتدا گره‌هایی را پیدا می‌کنیم که درجه آن‌ها در G_0 مینیمم است و مقدار آن‌ها را به عنوان $Ks_1=1$ تعریف می‌کنیم، سپس زیر شبکه‌ی جدید G_1 را با حذف این گره‌ها به دست می‌آوریم. به طور مشابه، گره‌هایی که درجه آن‌ها در G_1 مینیمم است را پیدا کرده و مقدار آن‌ها را به صورت $Ks_2=Ks_1+1$ تعریف می‌کنیم و به طور مداوم و بازگشتی تا زمانی که تمامی گره‌ها حذف شوند ادامه می‌دهیم. اگر درجه‌ی یک گره در k امین حذف صفر شود، مقدار آن Ks به صورت $(K+1)$ تعریف می‌شود. هر گره می‌تواند با استفاده از تجزیه k-shell یک مقدار Ks منحصر به فرد داشته باشد، همان طور که در شکل (۴) (ب) نشان داده شده

است. در اینجا تمایز دقیق تری از خصوصیات محلی متفاوت بین گره‌ها ارائه می‌شود [13].

۴-۱۲ - معیار مرکزیت Random Walk Betweenness

یکی از معایب BC این است که فقط گره‌هایی را در نظر می‌گیرد که درگیر کوتاه‌ترین مسیرها هستند. در شکل (۵)، گره c، که خارج از کوتاه‌ترین مسیرهای متصل کننده خوشه‌های چپ و راست است، با وجود اهمیت واضح آن برای مسیرهای جایگزین، امتیاز مرکزیت بینایی بسیار کمی می‌گیرد. این گونه گره‌های کنار گذاشته شده می‌توانند اهمیت حیاتی برای انعطاف پذیری شبکه در موازنه بار یا در آستانه شکست گره‌های کوتاه‌ترین مسیر داشته باشند. random walk betweenness برای حل این مشکل معرفی شد، ایده این است که یک فرآیند تصادفی اغلب مسیرهای غیر بهینه را محاسبه می‌کند. با فرض داشتن دانش در مورد کل گراف، این روش از ماتریس همسایگی استفاده می‌کند و شامل راه اندازی یک پیاده‌روی تصادفی

Error! Reference Error! Reference source not found.
[18][18]source not found.

۱۹-۴ - SALSا

این معیار از استعاره تقویت متقابل بین hubbiness و authoritativeness استفاده می کند، اما L1-normalizing ماتریس های A و A^T را در نظر می گیرد. با $\alpha_0=1$ شروع می کنیم و به صورت زیر ادامه می دهیم **Error!**
Reference source not found.

$$h_{i+1} = a_i \bar{A}^T \quad (۱۶)$$

$$a_{i+1} = h_{i+1} \bar{A} \quad (۱۷)$$

۵ - نتیجه

هر شبکه با توجه به روابط موجود در شبکه از خصوصیات خاصی برخوردار است. به همین دلیل در شبکه های اجتماعی، شاخص های ارائه شده و نتیجه گیری ها همواره برای همه ی انواع شبکه ها صادق نیستند. در نتیجه برای هر محیط، با توجه به نوع مسئله، نحوه انتشار و همچنین احتمال تاثیر گذاری، بهترین معیار متفاوت است. اهمیت گره یک مفهوم مهم است، اگر چه بسیاری از مقالات علمی برای توصیف خواص شبکه ها از آن استفاده می کنند. برخی مقالات ایده های جدیدی را برای تعریف اهمیت گره ارائه می دهند، به منظور ارزیابی اهمیت جامع گره، گاهی از معیارهایی استفاده می شود که از ترکیبی از معیارهای مرکزیت برای ارزیابی اهمیت گره استفاده می کنند. به طور کلی، در هر محیط با توجه به نوع مسئله و شاخص های اهمیت مورد نظر، معیار مرکزیت مورد استفاده و روش محاسبه ی اهمیت گره های متفاوت است.

ضمیمه

در جدول (۱) فرمول (در صورت وجود)، مزیت، عیب و کاربردهای معیارهای ارائه شده در این مقاله به صورت فشرده در یک جدول ارائه شده اند. توضیحات مختصری نیز در مورد هر کدام از معیارها به همراه مراجع آن ها نیز آورده شده است.

مراجع

- [1] Bojin Zheng, Deyi Li, Guisheng Chen, Wenhua Du, Jianmin Wang, "Ranking the Importance of Nodes of Complex Networks by the Equivalence Classes Approach", Elsevier, (2012)
- [2] Jianwei Wang, Lili Rong, Tianzhu Guo, "A new measure of node importance in complex networks with tunable parametrs", 4th International Conference on IEEE, Wireless Communications, Networking and Mobile Computing, pp-1-4, (2008)
- [3] Duanbing Chen, Linyuan Lu, Ming-Sheng Shang, Yi-Cheng Zhang, Tao Zhou, "Identifying influential nodes in complex networks", Physica A: Statistical Mechanics and its Applications, Vol. 391, pp-1777-1787, (2011)
- [4] Andrea Landherr, Bettina Friedl, Dr. Julia Heidemann, "A Critical Review of Centrality Measures in Social Networks", Business & Information Systems Engineering, Vol. 2, pp-371-385, (2010)
- [5] Yi Yu, Suohai Fan, "Node Importance Measurement Based on the Degree and Closeness Centrality", Journal of Information & Commputational Science, pp-1281-1291, (2015)

بردار نرمال سازی شده و فرایند تا زمان همگرایی تکرار می شود. بردارهای ویژه ی مسلط در عمل روی گراف هایی که متصل قوی نیستند شکست می خورند. **Error! Reference source not found.**[15][15]

۱۶-۴ - شاخص کاتز^{۱۵}

شاخص کاتز نفوذ نسبی یک گره را با اندازه گیری تعداد همسایه های مستقیم و غیرمستقیم گره محاسبه می کند. اتصالات غیر مستقیم با ضریب کاهش α جریمه می شوند. در واقع، به هر اتصال یک وزن α اختصاص داده می شود. شاخص کاتز به صورت زیر بیان می شود:

$$C_{Katz}(i) = \sum_{k=1}^{\infty} \sum_{j=1}^n \alpha^k (A^k)_{ji} \quad (۱۲)$$

با توجه به فعل و انفعالات بین قدرت ماتریس همسایگی و تعداد مسیرهای اتصال بین دو گره، برای اینکه حاصل جمع بالا محدود شود، ضریب کاهش α باید کمتر از $1/\lambda$ باشد، به طوری که λ مقدار ویژه مسلط ماتریس A است **Error! Unknown switch argument.**
Error! Unknown switch argument.
Error! Unknown switch argument.

۱۷-۴ - PageRank

PageRank یکی از شاخص های طیفی است که امروزه عمدتاً به دلیل استفاده در الگوریتم رتبه بندی گوگل بیشتر بحث و نقل می شود. این شاخص به صورت زیر محاسبه می شود:

$$p = \alpha p \bar{A} + (1 - \alpha)v \quad (۱۳)$$

به طوری که \bar{A} ماتریس مجاورت li-normalized شده گراف است، $\alpha \in [0, .1]$ یک عامل تعدیل است و v یک بردار اولویت است. ادعا می شود که امتیاز PageRank یک احتمال توزیع صفحات وب است، که دارای II norm است، اما اگر A دارای سطرهای خالی باشد، ضرورتاً درست نیست. مقالات علمی چندین راه برای این مشکل پیشنهاد می کنند تا ماتریس \bar{A} را تصادفی ۱۶ کنند. یک راه حل مشترک جایگزین کردن سطرهای خالی با بردار اولویت v خودش است، اما راه حل های دیگری نیز ارائه شده است. **Error! Error! Reference source not found.**
Error! Error! Reference source not found.[17][16]

۱۸-۴ - HITS

الگوریتم HITS از استعاره وب از "تقویت متقابل" استفاده می کند. یک صفحه وب معتبر است اگر به وسیله ی هاب های بسیار خوبی به آن اشاره شود (صفحاتی که شامل لیست خوبی از صفحات معتبر هستند) و یک هاب خوب است اگر به صفحات معتبری اشاره کند. این یک فرایند تکراری را نشان می دهد که در یک زمان امتیاز معتبر بودن a_i و امتیاز هاب بودن b_i را با شروع از $a_0=1$ محاسبه می کند و سپس قوانین به روزرسانی را اعمال می کند.

$$h_{i+1} = a_i A^T \quad (۱۴)$$

$$a_{i+1} = h_{i+1} A \quad (۱۵)$$

این فرایند به بردار ویژه سمت چپ ماتریس ATA همگرا می شود، و امتیاز Authority نهایی را می دهد. به طور عکس، با در نظر گرفتن بردار ویژه مسلط سمت چپ AAT، امتیاز hubbiness نهایی را می دهد

- [9] Paolo Boldi, Sebastiano Vigna, "Axioms for Centrality", Social and Information Networks, (2013)
- [10] Anne-Marie Kermarrec, Erwan le Merrer, Bruno Sericola, Gilles Tredan, "Second order centrality: Distributed assessment of nodes criticality in complex networks", Computer Communications, Vol. 34, pp-619-628, (2011)
- Mahendra piraveenan, Mikhail Prokopenko, Liaquat Hossain, "percolation centrality: Quantifying Graph-
- [6] G. Sabidussi, The centrality index of a graph, Psychometrika, Vol (31), pp-581-603, (1966)
- [7] Nan Lin. Foundations of Social Research. McGraw-Hill, New York, (1976)
- [8] Massimo Marchiori and Vito Latora. Harmony in the small-world. Physica A: Statistical Mechanics and its Applications, pp-539-546, (2000)

جدول (۱): مرور کلی معیارهای مرکزیت

معیار مرکزیت	فرمول	مزیت-عیب	کاربرد	توضیحات
درجه (Degree)	$C_D(v) = \text{deg}(v)$	مزیت: پیچیدگی محاسباتی کمی نسبت به بقیه معیارها دارد Error! Reference source not found. عیب: اتصال‌های غیرمستقیم در نظر گرفته نمی‌شوند Error! Reference source [4]not found. عیب: مبتنی بر خصوصیت محلی گره است و اطلاعات کمی از ساختار شبکه به ما می‌دهد Error! Reference source [13][13]not found.	برای اندازه‌گیری فعالیت‌های ارتباطی در شبکه استفاده می‌شود Error! Reference source [2][1]not found.	این معیار اهمیت محلی گره را در نظر می‌گیرد و موقعیت گره در شبکه را در نظر نمی‌گیرد [5] گره‌هایی با درجه بالا تأثیری در اتصال گراف ندارد [10]
نزدیکی (Closeness)	$C_C(v_i) = (N - 1) / \sum_{j=1}^N l_{ij}$	عیب: فقط مختص گراف‌های متصل است Error! Reference source [3]not found. عیب: پیچیدگی محاسباتی بالا برای شبکه‌های اجتماعی آنلاین [10]	برای تعیین گره‌هایی که در انتشار اطلاعات مؤثر هستند استفاده می‌شود [5] برای نمایش استقلال یا بهره‌وری ارتباطات استفاده می‌شود Error! Reference source not found. نفوذ روی همه‌ی گره‌های شبکه را در نظر می‌گیرد [5]	در شناسایی نفوذ گره نسبت به مرکزیت درجه بهتر عمل می‌کند اما پیچیدگی محاسباتی بالاتری دارد [9] تأثیری در اتصال گراف ندارد [10]
شاخص لین (Lin's Index)	$\frac{ \{y d(y, x) < \infty\} ^2}{\sum_{d(y,x) < \infty} d(y, x)}$	مزیت: اصلاح‌شده‌ی مرکزیت نزدیکی برای گراف‌های غیر متصل است Error! Reference source [7]not found.	قابل استفاده برای گراف‌های غیر متصل Error! Reference source not found.	مانند مرکزیت نزدیکی است ولی در آن گره‌هایی که تعداد گره‌های بیشتری به آن‌ها دسترسی دارند مهم‌تر می‌باشند Error! Reference source not found.
هارمونیک (Harmonic)	$\sum_{y \neq x} \frac{1}{d(y, x)}, (\infty^{-1} = 0)$	مزیت: اصلاح‌شده‌ی مرکزیت نزدیکی برای گراف‌های غیر متصل است Error!	شاخص مرکزیت همه‌منظوره‌ی عالی برای گراف‌های جهت‌دار دلخواه Error!	در گراف‌های با فاصله‌ی بینهایت میانگین هارمونیک بهتر از میانگین حسابی عمل می‌کند Error! Reference source not found.

found.	Reference source not found. مناسب برای استفاده در بازیابی اطلاعات Reference source not found.	Reference source [8]not found.		
اولین و بارزترین معیار طیفی است [15]		عیب: در عمل روی گرافهایی که متصل قوی نیستند شکست می-خورد [15][15]	توضیح در (۵-۱۵)	بردار ویژه ی مسلط چپ (Left Dominant Eigenvector)
یک تعمیم از مرکزیت درجه است [16] مرکزیت کاتز یک نوع مرکزیت بردار ویژه است [16][16]	برای محاسبه ی مرکزیت در گرافهای بدون چرخه ی جهت دار مناسب است [16]	مزیت: اتصالات مستقیم و غیرمستقیم را در نظر می گیرد، برخلاف مرکزیت درجه که فقط اتصالات مستقیم را در نظر می-گیرد Error! Reference source [16]not found.	$C_{Katz}(i) = \sum_{k=1}^{\infty} \sum_{j=1}^n \alpha^k (A^k)_{ji}$	شاخص کاتز (Katz's Index)

نسبت به HITS سرعت بالاتری دارد [16]	بر روی گرافهای جهت دار تمرکز دارد Error! Reference source [16]not found. به طور خاص در جهت گرافهای وب ایجاد شده است Error! Reference source [17]not found. با هدف پیاده سازی در موتورهای جستجو ایجاد شده است Error! Reference source [16]not found.	در شناسایی گره های بانفوذ مؤثر است ولی پیچیدگی بالایی دارد [17]	$p = \alpha p \bar{A} + (1 - \alpha)v$	رتبه بندی صفحه (PageRank)
نسبت به PageRank به query کاربر حساس است و دقت بالاتری دارد [18] برخلاف PageRank که اهمیت صفحات را فقط از یک بعد در نظر می گیرد، این روش برای صفحات اهمیت را از دو بعد در نظر	بر روی گرافهای جهت دار تمرکز دارد به منظور رتبه بندی داده های بازیابی شده استفاده می شود [18]	مزیت: قابلیت ترکیب با دیگر رتبه بندی های مبتنی بر بازیابی اطلاعات [18][18] عیب: زمان پرس و جوی بالا [18] عیب: به علت سرعت کم	$h_{i+1} = a_i A^T$ $a_{i+1} = h_{i+1} A$	HITS

می گیرد [18]	محاسبه HITS امکان پذیری کمی دارد [18][18]			
بهبود دهنده روش HITS با نرمال سازی ماتریس [19]	در تجزیه و تحلیل ساختار لینک کاربرد دارد	مزیت: قابلیت ترکیب با دیگر رتبه بندی های مبتنی بر بازیابی اطلاعات را دارد	$h_{i+1} = a_i \bar{A}^T$ $a_{i+1} = h_{i+1} \bar{A}$	سالسا (SALSA)
در شناسایی نفوذ گره نسبت به مرکزیت درجه بهتر عمل می کند اما پیچیدگی بالاتری دارد [9] قابل محاسبه به صورت توزیع شده نیست [10] مناسب ترین معیار برای بررسی اتصال کلی گراف است [10]	برای تعیین نفوذ و نقش واسطه گری گره ها در تعامل بین گره های دیگر استفاده می شود Error! Reference source not found. [2] جهت تعیین گره هایی که حذف آن ها باعث اختلال سریع در شبکه می شوند و در اتصال و انتشار اطلاعات مهم هستند کاربرد دارد [9][3] برای نمایش پتانسیل کنترل ارتباطات استفاده می شود Error! Reference source not found.	عیب: پیچیدگی محاسباتی بالا برای شبکه های اجتماعی آنلاین [9] عیب: فقط گره هایی را در نظر می گیرد که درگیر کوتاه ترین مسیرها می باشند [10] عیب: نیاز به دانش عمومی از کل گراف دارد [10]	$C_B(v_i) = \sum_{i \neq j \in V} \frac{k_{jr}(v_i)}{k_{jr}}$	بینابینی (Betweenness)
اصلاح شده ی مرکزیت بینابینی است و برخلاف آن اهمیت کلیه گره ها را بررسی می کند [14]	برای تعیین گره های بانفوذ در انتشار اطلاعات استفاده می شود [14]	نیاز به دانش عمومی از کل گراف دارد [14] قابل محاسبه به صورت توزیع شده نیست [14]	توضیح در (۵-۱۲)	پیاده روی تصادفی بینابینی (Random Walk Betweenness)
تأثیری در اتصال گراف ندارد [10] مرکزیت بردار ویژه یک گره تابعی از مرکزیت همسایه های آن گره است [10]	تمایل به شناسایی مراکز clique های بزرگ دارد [10]	مزیت: اطلاعات محلی گره را در نظر می گیرد [5]	$C_E(v_i) = v_x$ $= \frac{1}{\lambda_{\max}(A)} \cdot \sum_{j=1}^N a_{jx} \cdot v_j$	بردار ویژه (Eigenvector)

	در پیدا کردن گره هایی که تأثیر بیشتری در انتشار اطلاعات دارند کاربرد دارد [12]	مزیت: کارایی خوب [12][12]	$X(v) = \text{number of cliques to which vertex } v \text{ belongs.}$	Cross-clique
به خوبی مرکزیت نزدیکی عمل می کند ولی پیچیدگی زمانی بهتری دارد [9]	پیدا کردن گره هایی که تأثیر بیشتری در انتشار اطلاعات دارند [9] برای اعمال در شبکه های ناهمگن مناسب ترین است [9]	مزیت: پیچیدگی زمانی پایین Error! Reference source not found.	$Q(u) = \sum_{w \in \Gamma(u)} N(w)$ $C_L(v) = \sum_{u \in \Gamma(v)} Q(u)$	محلی (Local)

<p>به روش توزیع شده محاسبه می شود [[10]] برای قطع مؤلفه های گراف قابل استفاده نیست [[10]]</p>	<p>برای مقابله با توزیع ناهمگن درجه از Second order توزیع شده استفاده می شود [[10]] برای تعیین میزان تأثیر گره ها در کوتاه کردن طول مسیرها استفاده می شود [[10]]</p>	<p>مزیت: قدرت این الگوریتم در سادگی آن نهفته است که متکی بر ملاقات با random walk ساده است [[10]] مزیت: نیاز به اطلاعات عمومی از گراف ندارد [[10]]</p>	<p>توضیح در (۵-۱۴)</p>	<p>مرتبه دوم (Second order)</p>
<p>نشان دهنده ی ویژگی های عمومی شبکه است Error! Reference source not found. [13]</p>	<p>اهمیت گره در کمک به نفوذ در شبکه را تعیین می کند Error! Reference source not found.</p>	<p>مزیت: پیچیدگی محاسباتی کم 00</p>	$C_p^t(v) = \frac{1}{N-2} \times \sum_{s \neq v \neq r} \frac{\sigma_{sr}(v)}{\sigma_{sr}} \frac{x_s^t}{\sum [x_i^t] - x_v^t}$	<p>نفوذ (Percolation)</p>
<p>نشان دهنده ی ویژگی های عمومی شبکه است Error! Reference source not found. [13]</p>	<p>برای شبکه های بزرگ مناسب است [5]</p>	<p>مزیت: پیچیدگی زمانی پایینی دارد Error! Reference source not found. [10] عیب: با گره های موجود در یک لایه یکسان به طور برابر و بدون تمایز برخورد می کند Error! Reference source not found.</p>	<p>توضیح در (۵-۱۱)</p>	<p>K-shell</p>
<p>تأثیری در اتصال گراف ندارد [[10]]</p>	<p>گره های مهم در انتشار اطلاعات را مشخص می کند [10]</p>	<p>عیب: فقط مختص گراف های متصل است [10]</p>	$C_e(v_i) = \frac{1}{\max_{j \in V} d(i, j)}$	<p>گریز از مرکز (Eccentricity)</p>

[19] Ronny Lempel and Shlomo Moran, SALSAs: the stochastic approach for link-structure analysis, ACM Trans. Inf. Syst. , Vol (2), pp-131-160, (2001)

[11] Theoretic Impact of Nodes during Percolation in Networks”, Vol. 8, (2013)
[12] Mohammad Reza Faghani, “A Study of XSS Worm propagation and Detection Mechanisms in online Social Networks”, IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY, Vol. 8, pp-1815-1826, (2013)
[13] Zhonghua Liu, Cheng Jiang, Juyun Wang, Hua Yu, “The node importance in actual complex networks based on a multi-attribute ranking method”, Knowledge-Based Systems, Vol. 84, pp-56-66, (2015)
[14] Newman, A measure of betweenness centrality based on random walks, Social Netw. pp-39-54, (2005)
[15] Abraham Berman and Robert J. Plemmons. Nonnegative Matrices in the Mathematical Sciences. Classics in Applied Mathematics. SIAM, (1994)
[16] Leo Katz, A new status index derived from sociometric analysis, Psychometrika, pp-39-43, (1953)
[17] Nick Craswell, David Hawking, and Trystan Upstill, Predicting fame and fortune: Pagerank or indegree. In Proceedings of the Australasian Document Computing Symposium, ADCS2003, pp-31-40, (2003)
[18] Jon M. Kleinberg, Authoritative sources in a hyperlinked environment. Journal of the ACM, Vol (5), pp-604-632, (1999)

- 1 centrality
- 2 centrality measures
- 3 left dominant eigenvector
- 4 degree
- 5 closeness
- 6 Lin's Index
- 7 harmonic
- 8 betweenness
- 9 dominant eigenvector
- 10 eccentricity
- 11 percolation centrality
- 12 percolation path
- 13 local centrality
- 14 left dominant eigenvector
- 15 katz index
- 16 stochastic