



Identifying Influential People in Large Scale Social Network Using Global and Local Structural Information

Nooshin Sharifi^{1*}, Mehdi Bateni²

¹ Masters student of Software Engineering Sheikh Bahaee University, Esfahan, Iran
nooshin.sharifi@shbu.ac.ir

² P.H.D of Computer Engineering Sheikh Bahaee University, Esfahan, Iran
bateni@shbu.ac.ir

Abstract

Today, the importance of publication and the strength of architecture in the network, has attracted the attention of many researchers to identify influential people in the social networks. Therefore, many methods have been proposed to discover the effective nodes. However, each of the proposed methods has its own limitations. The GLS algorithm is one of the methods for detecting and ranking effective nodes. The GLS is more accurate than similar classical and modern methods by combine two dimensions of global and local node information. The GLS has a time complexity of exponential order. Therefore, using this method for large networks is very expensive and in some cases impossible. In this article, we have modified the structure of GLS to reduce its execution time. This change is made by considering the common neighbors of each node with its neighbors only up to a radius of two in the global influence calculation section. The algorithm presented in this article (NGLS) in addition to maintaining the accuracy of GLS, has a time complexity of linear order. NGLS runtime is reduced by at least 50% on 15 real-world networks and by at least 80% on 23 simulated networks compared to the GLS algorithm. NGLS can be used on a variety of networks with different scales.

Keywords: Influential nodes, Global and local information, Large networks, Centrality Measures

شناسایی افراد تاثیرگذار در شبکه‌های اجتماعی بزرگ با استفاده از اطلاعات ساختاری سراسری و محلی

نوشین شریفی^{۱*}، مهدی باطنی^۲

*دانشجوی کارشناسی ارشد مهندسی نرم‌افزار، گروه کامپیوتر، دانشگاه شیخ‌بهائی، اصفهان،
*nooshin.sharifi@shbu.ac.ir

دکتری مهندسی کامپیوتر، گروه کامپیوتر، دانشگاه شیخ‌بهائی، اصفهان
bateni@shbu.ac.ir

چکیده

امروزه اهمیت مبحث انتشار و استحکام معماری در شبکه، توجه محققین زیادی را به سمت شناسایی افراد تاثیرگذار در شبکه‌های اجتماعی جلب کرده است. بنابراین روش‌های زیادی برای کشف گره‌ی تاثیرگذار ارائه شده است. با این وجود هر کدام از معیارهای ارائه شده دارای محدودیت‌های خاص خود هستند. الگوریتم GLS یکی از متدهای کشف و رتبه‌بندی گره‌های تاثیرگذار است. GLS با ادغام دو بعد اطلاعات سراسری و محلی گره‌ها، از دقت بیشتری نسبت به متدهای کلاسیک و نوین مشابه برخوردار است. GLS دارای پیچیدگی زمانی از مرتبه‌ی نمایی است. بنابراین استفاده از این متد برای شبکه‌های بزرگ هزینه‌ی زیادی را در بر دارد و در برخی موارد غیرممکن است. در این مقاله ساختار GLS را به منظور کاهش زمان اجرای آن تغییر داده‌ایم. این تغییر با در نظر گرفتن همسایه‌های مشترک هر گره با همسایه‌هایش تنها تا شعاع دو در بخش محاسبه‌ی تاثیر سراسری انجام شده است. الگوریتم ارائه شده در این پژوهش (NGLS) ضمن حفظ دقت GLS، دارای پیچیدگی زمانی از مرتبه‌ی خطی است. زمان اجرای NGLS روی ۱۵ شبکه‌ی دنیای واقعی حداقل ۵۰٪ و روی ۱۳ شبکه‌ی شبیه‌سازی شده حداقل ۸۰٪ نسبت به الگوریتم GLS کاهش یافته است. NGLS قابل استفاده در انواع شبکه‌ها با مقیاس‌های متفاوت است.

کلمات کلیدی

گره‌های تاثیرگذار، اطلاعات سراسری و محلی، شبکه‌های بزرگ، معیارهای مرکزیت

این روند به شدت با توپولوژی شبکه‌ها ارتباط دارد [۲]. با توجه به اهمیت کشف گره‌های تاثیرگذار تاکنون روش‌ها و متدهای بسیاری با دیدگاه‌های مختلف ارائه شده است که هر کدام از این متدها دارای محدودیت‌های خاصی هستند. این مسئله باعث شده است که کشف گره‌های تاثیرگذار یک مسئله‌ی باز باقی بماند. در میان روش‌های ارائه شده در این زمینه، به این علت که گره‌های تاثیرگذار اطلاعات سراسری یا موقعیتی بیشتری را در مقایسه با سایر گره‌ها در اختیار دارند [۳] الگوریتم GLS [۱۱] از ترکیب بعد سراسری و بعد

۱- مقدمه

در سال‌های اخیر با ظهور شبکه‌های اجتماعی آنلاین (OSN) علاقه به تجزیه و تحلیل شبکه‌های اجتماعی (SNA) بسیار زیاد شده است. در این میان، دانش در مورد نقش گره‌های شبکه در فرآیند انتشار برای ایجاد روش‌های کارآمد برای جلوگیری یا تسریع این فرآیند، در مورد بیماری‌ها یا اطلاعات، بسیار مهم است [۱]. بسیاری از مطالعات به وضوح نشان داده‌اند که

احتمال پایه، این محدودیت رفع گشته و متد NEC ارائه شده است اما دارای پیچیدگی زمانی از مرتبه‌ی نمایی است. در [۸] بر اساس مدل گسترش SI، مرکزیت‌های سنتی در یک شاخص مرکزیت جدید (SIRC) ترکیب شده‌اند. اما این مدل از نظر زمانی وابسته به مسئله‌ی بهینه‌سازی، زمان محاسبه‌ی شاخص‌های مرکزیت و مدل گسترش نفوذ است. در رده‌بندی گره‌ها، بر اساس معیارهای مختلف، ممکن است که یک گره موجود در یک لیست تولید شده در دیگری وجود نداشته باشد. در [۹] برای غلبه بر این چالش، جمع‌بندی رتبه مبتنی بر اجماع (ACRA) پیشنهاد شده است که بر اساس جفت مقدار-کلید استوار است.

روش‌های مبتنی بر مرکزیت: مهم‌ترین معیارها در زمینه‌ی شناسایی تأثیرگذارترین گره، مرکزیت‌ها هستند. با این حال روش‌های ساده‌ی مرکزیت از دقت کافی برخوردار نمی‌باشند، و روش‌های دقیق مانند مرکزیت میانی و مرکزیت نزدیکی پیچیدگی محاسباتی بالایی را به همراه دارند [۱۰]. در مرکزیت‌های طیفی نیز، مانند مرکزیت نیمه محلی (SLC)، مرکزیت بردار ویژه (EC)، رتبه‌بندی صفحه (PR) و رتبه‌بندی رهبر (LR) محدودیت‌های نسبت به توپولوژی شبکه وجود دارد [۱۱] [۱۲].

در برخی از پژوهش‌ها از مرکزیت‌های کلاسیک به طور مستقیم برای کشف گره‌های تأثیرگذار استفاده شده است. به عنوان مثال در [۱۳] با استفاده از مرکزیت میانی، مرکزیت نزدیکی، مرکزیت درجه و مرکزیت بردار ویژه، گره‌های یک شبکه تروریستی در وب تاریک رتبه‌بندی شده است. در [۱۴] و [۱۵] نیز پنج مرکزیت مذکور برای فعالیت‌های بازاریابی در زمینه‌ی زیبایی و سلامت و کشف گره‌های تأثیرگذار فریبکار استفاده شده‌اند. دستیابی به گره‌های تأثیرگذار در OppNet به دلیل پویایی توپولوژی و عدم دید سراسری نسبت به شبکه، آسان نیست. بنابراین [۱۶] از روش‌های آماری توقف بهینه برای انتخاب بین گره‌های مختلف کاندیدای تحویل استفاده نموده است.

به علت محدودیت‌های روش‌های کلاسیک برخی از پژوهشگران سعی در استفاده از ایده‌ی اولیه‌ی آن‌ها نموده‌اند. در [۱۷] به منظور استفاده از مفهوم مرکزیت نزدیکی به جای محاسبه‌ی نزدیکی یک گره با کل گره‌های شبکه، کوتاه‌ترین مسیرها فقط تا گره‌های مرکزی جوامع اندازه‌گیری شده است. بنابراین پیچیدگی زمانی مرکزیت نزدیکی از $O(n^3)$ به $O(n^2)$ کاهش یافته است.

در بخشی از پژوهش‌ها محققین به ارائه‌ی روش‌های فرا ابتکاری پرداخته‌اند. به عنوان مثال: H-index میزان تأثیرگذاری گره‌ها را بر اساس توانایی انتشار آن‌ها در نظر می‌گیرد اما بخشی از اطلاعات مربوط به گره‌های همسایه در این روش نادیده گرفته می‌شود [۱۸]. جهت رفع این نقص در [۱۹] متدی ارائه شده است که براساس همه‌ی اطلاعات همسایگان، قابلیت گسترش گره را تعیین می‌نماید و توانایی بالایی در تمایز بین گره‌ها و رتبه‌بندی آن‌ها، در مقایسه با سایر روش‌ها دارد.

محلی گره برای رتبه‌بندی گره‌ها استفاده نموده است. الگوریتم GLS نیاز به دانش قبلی و تنظیم پارامتر ندارد و به طور خودکار قادر به کشف گره‌های تأثیرگذار است. طبق نتایج، متد GLS نسبت به متدهای: رتبه‌بندی صفحه، بردار ویژه، مرکزیت میانی، مرکزیت نزدیکی، K-shell، H-index، Hits، WFCA دارای بالاترین ضریب کندال τ (بالاتر از ۰.۸) است که این مسئله نشانگر دقت بالای آن نسبت به روش‌های نام برده است. GLS نیز مانند سایر روش‌های دارای محدودیت است. این متد دارای پیچیدگی زمانی از مرتبه‌ی نمایی ($O(n^2)$) است؛ بنابراین با توجه به رشد لحظه‌ای شبکه‌های اجتماعی استفاده از این الگوریتم در شبکه‌های بزرگ هزینه‌بر و در برخی موارد ناممکن است.

بنابر دقت و قدرت الگوریتم GLS، هدف این پژوهش کاهش پیچیدگی زمانی این متد است تا بتوان از آن در شبکه‌های بزرگ نیز استفاده نمود. بخش محاسبه‌ی تأثیر سراسری این الگوریتم بیشترین مقدار زمان اجرا را به خود اختصاص داده است. در این پژوهش با در نظر گرفتن این فرض که هر گره تنها با همسایه‌هایش تا شعاع دو قادر به داشتن همسایه‌ی مشترک است، ساختار این متد به گونه‌ای تغییر یافته است که پیچیدگی زمانی آن به $O(nmk)$ کاهش یافته، بدون اینکه از دقت آن کاسته شده باشد.

ساختار ادامه‌ی این مقاله بدین شرح است: در بخش ۲ پژوهش‌های پیشین در زمینه‌ی کشف گره‌های تأثیرگذار ارائه شده است. در بخش ۳ ایده‌ی بکار رفته در این پژوهش به همراه مراحل الگوریتم پیشنهادی ارائه شده است. بخش ۴ شامل نتایج حاصل از پیاده‌سازی و آزمایشات مربوط به زمان و دقت الگوریتم است و در بخش ۵ مقاله با جمع‌بندی و نتیجه‌گیری به پایان رسیده است.

۲- پژوهش‌های پیشین

روش‌های مبتنی بر آنتروپی، اجماع و تصمیم‌گیری چند معیاره: در بخشی از پژوهش‌ها در زمینه‌ی کشف گره‌های تأثیرگذار، از مفهوم آنتروپی استفاده شده است، به عنوان نمونه در [۴]، متد (CbM) آنتروپی یک پیاده-روی تصادفی از یک گره به هر جامعه را در نظر می‌گیرد. از طرفی، برای حل کمبود معیار واحد در ارزیابی اهمیت گره، چندین روش رتبه‌بندی براساس تصمیم‌گیری چند معیاره (MCDM) پیشنهاد شده است. در [۵] از VIKOR و توزین آنتروپی برای ترکیب مرکزیت درجه، مرکزیت نزدیکی و مرکزیت میانی استفاده شده است و طبق نتایج VIKOR خروجی دقیق‌تری نسبت به متدهای کلاسیک به دست آورده است. اما همچنان گره‌هایی با رتبه‌ی یکسان دارد. یکی دیگر از تکنیک‌های MCDM، AHP است [۶]، که با تقلیل تصمیمات پیچیده به یک سری از مقایسه‌های دو به دو، و سپس ترکیب آن‌ها عمل می‌کند. آزمایشات حاکی از آن است که AHP نسبت به متدهای کلاسیک و W-TOPSIS عملکرد بهتری داشته است.

ایده‌ی مرکزیت مشهود شبیه به MCDM است اما فقط قابل اعمال روی شبکه‌های وزن دار است. در [۷] از طریق اصلاح قدرت تخصیص

$$\text{averD}(v_i) = \frac{\sum_{v_j=\eta 1(v_i)} d(v_j)}{d(v_i)} \quad (2)$$

که $d(v_i)$ و $d(v_j)$ به ترتیب درجه گره v_i و v_j را نشان می‌دهند و $\eta 1(v_i)$ مجموعه‌ی نزدیکترین همسایه‌های v_i است.

تعریف ۳ (همسایگان مشترک): گره‌های v_i و v_j نشانگر دو گره متفاوت در شبکه G هستند. گره‌های همسایه که بین این دو گره مشترک هستند توسط $\text{Com}(v_i, v_j)$ نشان داده می‌شوند، که به شرح فرمول (۳) تعریف می‌شود.

$$\text{Com}(v_i, v_j) = \text{Nei}(v_i) \cap \text{Nei}(v_j) \quad (3)$$

$\text{Nei}(v_i)$ و $\text{Nei}(v_j)$ نشانگر نزدیکترین همسایه‌های گره v_i و v_j هستند. **تعریف ۴ (تأثیر سراسری):** توانایی گره v_i برای به اشتراک گذاشتن منابع با گره‌های دیگر یکی از مؤلفه‌های محاسبه‌ی تأثیر گره است. تأثیر سراسری گره v_i با $\text{glbInflu}(v_i)$ مشخص می‌شود و به صورت فرمول (۴) تعریف شده است:

$$\text{glbInflu}(v_i) = d(v_i) * \sum_{v_j=\eta 1} \text{pow}(A, \#\text{Com}(v_i, v_j)) \quad (4)$$

در فرمول (۴)، v_j نشانگر همسایه‌های گره v_i است، $\#\text{Com}(v_i, v_j)$ تعداد همسایه‌های مشترک دو گره v_i و v_j را مشخص می‌کند و A یک ثابت است که ۱.۰۱ در نظر گرفته شده است.

تعریف ۵ (احتمال مشارکت): معکوس متوسط درجه به عنوان احتمال مشارکت گره‌های همسایه در تأثیر گره v_i در نظر گرفته شده است و با $p(v_i)$ به صورت فرمول (۵) مشخص می‌شود.

$$p(v_i) = \frac{1}{\text{averD}(v_i)} \quad (5)$$

تعریف ۶ (تأثیر محلی): تأثیر محلی به تأثیر گره v_i در ساختار محلی (تا شعاع یک) اشاره دارد و با $\text{locInflu}(v_i)$ مشخص می‌شود، که به شرح فرمول (۶) تعریف می‌شود:

$$\text{locInflu}(v_i) = \sum_{v_j=\eta 1(v_i)} \text{DC}(v_j) p(v_i) \quad (6)$$

در اینجا مرکزیت درجه‌ی هر گره همسایه به عنوان نفوذ آن، و میزان مشارکت هر گره در گره v_i برابر با احتمال p در نظر گرفته شده است.

تعریف ۷ (تأثیر نهایی): تأثیر گره v_i نه تنها به نفوذ خود گره مربوط است، بلکه به محیط اطراف آن نیز بستگی دارد. بنابراین از ترکیب تأثیر

از دیگر متدهای فرابتنکاری مشابه، k -shell است. این متد براساس درجه‌ی گره‌ها آن‌ها را در لایه‌های مختلف قرار داده و بر این اساس به آن‌ها یک شاخص نسبت می‌دهد. ایراد اصلی این متد قرار دادن تعداد زیادی از گره‌ها را در یک لایه است، [۲۰] به وسیله‌ی یک پارامتر آزاد، ضعف ذکر شده را رفع کرده است. تنظیم این پارامتر نیازمند آزمون و خطا می‌باشد و بار پردازشی زیادی را به همراه دارد، بنابراین، در [۳] یک مدل ریاضی بر اساس شعاع همسایگی ارائه شده است. همچنین مقدار ksh هر گره را برابر مجموع ksh همسایگان آن قرار داده است که نام این متد $kshi+$ می‌باشد.

در گروهی از پژوهش‌ها ترکیب معیارهای مرکزیت مورد بررسی قرار گرفته‌اند. [۲۱] با ایده از رابطه گرانش نیوتون، مرکزیت درجه‌ی گره‌ها را به عنوان جرم، و برای تأثیر خود گره شاخص k -shell آن را به کار برده است. در نهایت مقدار مرکزیت گرانشی نهایی هر گره را برابر با مجموع مرکزیت گرانشی نزدیکترین همسایگان آن قرار داده است.

گروهی از محققین از معیارهای مرکزیت با توجه به سراسری بودن یا محلی بودن معیارها استفاده نموده‌اند. در [۱۰] میزان مشارکت همسایه‌ها تا شعاع ۲ در اثرگذاری گره مورد نظر اندازه‌گیری شده، و سپس با تأثیر خود گره بر نزدیک‌ترین همسایه‌ها، ترکیب شده است. این الگوریتم در برخی از شبکه‌ها به علت عدم توجه به ویژگی‌های سراسری، ضعیف عمل نموده است. در مقابل در [۲۲] مدل کمی GIN از بعد سراسری به کشف گره‌های تأثیرگذار پرداخته است. این متد در شبکه‌های غیرمتصل در رتبه‌بندی گره‌های تأثیرگذار بسیار توانمند است. در نهایت در [۱۱] الگوریتمی ارائه شده است که هم بعد سراسری هم بعد محلی گره را مدنظر قرار داده است، و هر دوی آن‌ها را با هم ترکیب نموده است. نمودار بلوکی این الگوریتم در شکل (۱) آورده شده است.

قبل از توضیح روش پیشنهادی، به برخی تعاریف اساسی مورد استفاده در GLS و روش پیشنهادی پرداخته شده است. با فرض اینکه $G=(V, E)$ یک شبکه‌ی بدون وزن و جهت، دارای $|V|$ گره و $|E|$ لبه باشد:

تعریف ۱ (مرکزیت درجه): مرکزیت درجه ساده‌ترین شاخص برای اندازه‌گیری تأثیر یک گره در شبکه است، که با $\text{DC}(v_i)$ معرفی شده است و در فرمول (۱) تعریف شده است.

$$\text{DC}(v_i) = \frac{d(v_i)}{n-1} \quad (1)$$

که v_i یک گره در شبکه G است، n تعداد کل گره‌ها است و $d(v_i)$ نشان دهنده‌ی درجه‌ی گره v_i است.

تعریف ۲ (متوسط درجه): $\text{averD}(v_i)$ متوسط درجه تمام گره‌های همسایه گره v_i است که به صورت فرمول (۲) تعریف می‌شود:

به طور کلی الگوریتم NGLS شامل بخش‌های زیر است:
محاسبه تأثیر سراسری: جریان اطلاعات بین گره‌های موجود در شبکه تأثیر گره‌ها را نشان می‌دهد. بنابراین، باید تأثیر قابلیت اشتراک اطلاعات گره با گره‌های دیگر در کل شبکه لحاظ شود. محاسبه تأثیر سراسری گره در روش پیشنهادی بدین شرح است: ۱- شناسایی تعداد همسایه‌های مشترک در شعاع یک و محاسبه σ_1 بر اساس فرمول (۸).

$$\sigma_1 = \sum_{v_j=1}^n \text{pow}(A, \#Com(v_i, v_j)) \quad (8)$$

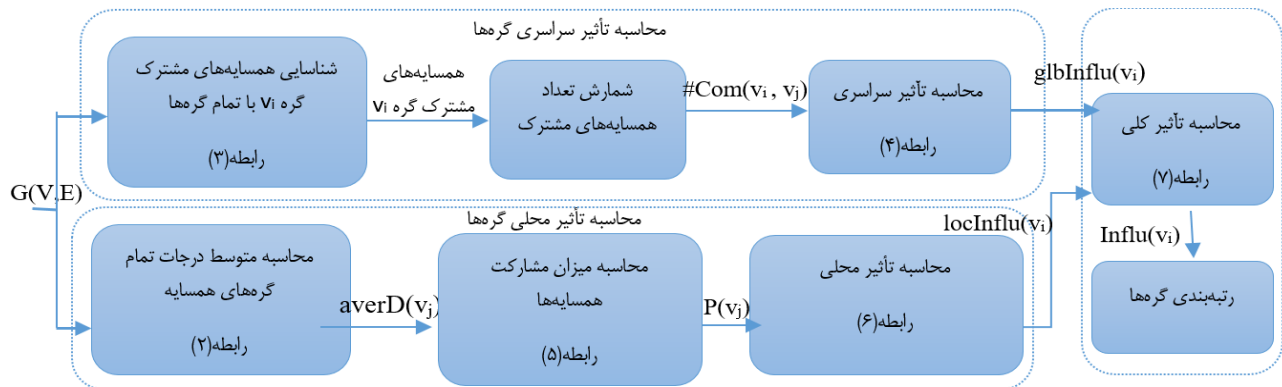
۲- شناسایی تعداد همسایه‌های مشترک در شعاع دو و محاسبه σ_2 بر اساس فرمول ۳.۸-۳.۰ صفر در نظر گرفتن تعداد همسایه‌های مشترک با سایر گره‌ها (در شعاع بیش از دو) و محاسبه σ_3 بر اساس فرمول ۵.۸-۵. محاسبه مجموع σ_i ها ۴- محاسبه تأثیر سراسری گره طبق فرمول (۹).

سراسری و محلی به عنوان شاخصی برای اندازه‌گیری تأثیر گره v_i استفاده می‌شود که با $Influ(v_i)$ مشخص می‌شود و در فرمول (۷) تعریف شده است.

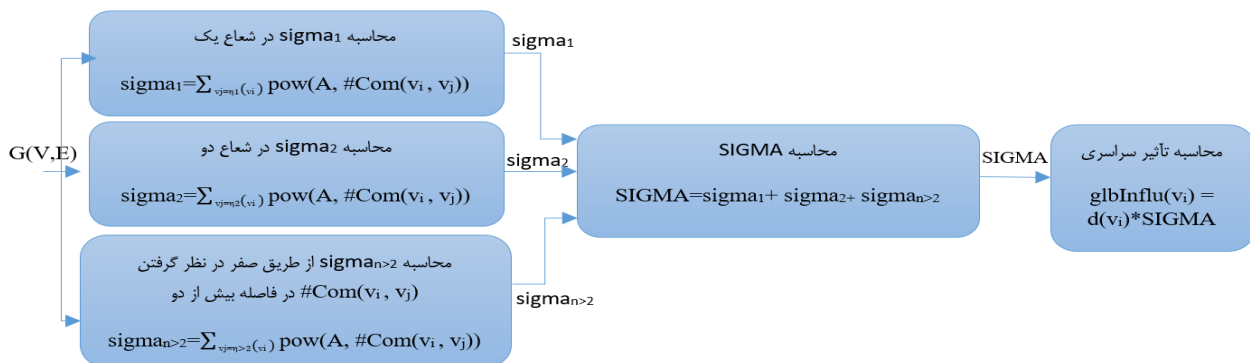
$$Influ(v_i) = \text{glbInflu}(v_i) \text{ locInflu}(v_i) \quad (7)$$

۳- روش پیشنهادی

چنانچه اشاره شد GLS به دلیل پیچیدگی زمانی بالا در شبکه‌های بزرگ قابل استفاده نیست. در روش پیشنهادی با در نظر گرفتن این فرض که هر گره فقط با همسایگان تا شعاع دو قادر است همسایه‌ی مشترک داشته باشد، و تعداد همسایگان مشترک آن با سایر گره‌ها برابر صفر است، می‌توان محدودیت زمانی الگوریتم GLS را رفع نمود. علت این فرض آن است که احتمال داشتن همسایه‌ی مشترک بین هر گره با نزدیک‌ترین همسایگان آن وجود دارد، در حالی که با همسایگان تا شعاع دو، قطعاً دارای حداقل یک همسایه‌ی مشترک است؛ اما با همسایگان در مسیرهای طولانی‌تر امکان وجود هیچ همسایه‌ی مشترکی نیست. بنا بر این فرض، مدت GLS که دارای دو حلقه‌ی تو در تو به سایز n در محاسبه تأثیر سراسری است، به سه حلقه-ی تو در تو با اندازه‌ی k, m, n تبدیل شده است که پیچیدگی زمانی الگوریتم را به $O(nmk)$ کاهش می‌دهد. نمودار بلوکی بخش محاسبه تأثیر سراسری در NGLS در شکل (۲) نمایش داده شده است.



شکل (۱): نمودار بلوکی الگوریتم GLS



شکل (۲): نمودار بلوکی تأثیر سراسری الگوریتم NGLS

۴-۲- زمان اجرای الگوریتم

برای مقایسه‌ی زمان اجرای الگوریتم‌های GLS و NGLS در نمودار شکل ۳ زمان اجرای الگوریتم‌ها بر روی شبکه‌های دنیای واقعی بر حسب ثانیه، نمایش داده شده است. در این نمودار محور عمودی نشانگر زمان بر حسب ثانیه، و محور افقی نشانگر تعداد گره‌های شبکه‌ها است. طبق نتایج زمان اجرای الگوریتم GLS به صورت نمایی با افزایش اندازه‌ی شبکه در حال افزایش است. به همین علت با ویژگی‌های سیستم مذکور، زمان اجرای GLS در شکل (۳) تنها روی شبکه‌هایی تا اندازه ۳۲۴۳۰ قابل اندازه‌گیری است. NGLS روی شبکه‌های بزرگ‌تر تا ۳ میلیون گره نیز اعمال شده است. طبق نتایج زمان اجرای NGLS روی هشت مجموعه داده‌ی اولیه نسبت به الگوریتم GLS حداقل ۵۰٪ کاهش یافته است. زمان اجرای NGLS دارای الگوی رشد ثابت نیست، علت این رخداد این است که پیچیدگی زمانی تنها تابعی از اندازه‌ی شبکه نیست؛ پیچیدگی زمانی الگوریتم NGLS از مرتبه‌ی $O(nmk)$ است بنابراین این مسئله به ویژگی‌های توپولوژیکی شبکه نظیر متوسط درجات، مدل توزیع درجات و تعداد مثلثات در شبکه وابسته است. جهت بررسی ویژگی‌های شاخص شبکه‌های شبیه‌سازی شده در زمان اجرای الگوریتم‌ها، زمان اجرای دو الگوریتم بر اساس سه مدل شناخته شده‌ی

$$glbInflu(v_i) = d(v_i) * SIGMA \quad (9)$$

محاسبه‌ی تأثیر محلی: محاسبه‌ی تأثیر محلی شامل مراحل زیر است: ۱- محاسبه میانگین درجات تمام گره‌های همسایه طبق فرمولی ۲. ۲- محاسبه‌ی میزان مشارکت گره‌های همسایه در تأثیر گره توسط فرمولی ۵ ۳- محاسبه‌ی مقدار نهایی تأثیر محلی بر اساس فرمول (۶).
رتبه‌بندی گره‌ها: ۱- در این مرحله تأثیر کلی هر گره که ترکیبی از تأثیر سراسری و محلی گره است براساس فرمول (۷) محاسبه می‌گردد. ۲- گره‌های موجود در شبکه بر اساس تأثیر بدست آمده برای آن‌ها مرتب‌سازی می‌شوند.

۴- ارزیابی نتایج

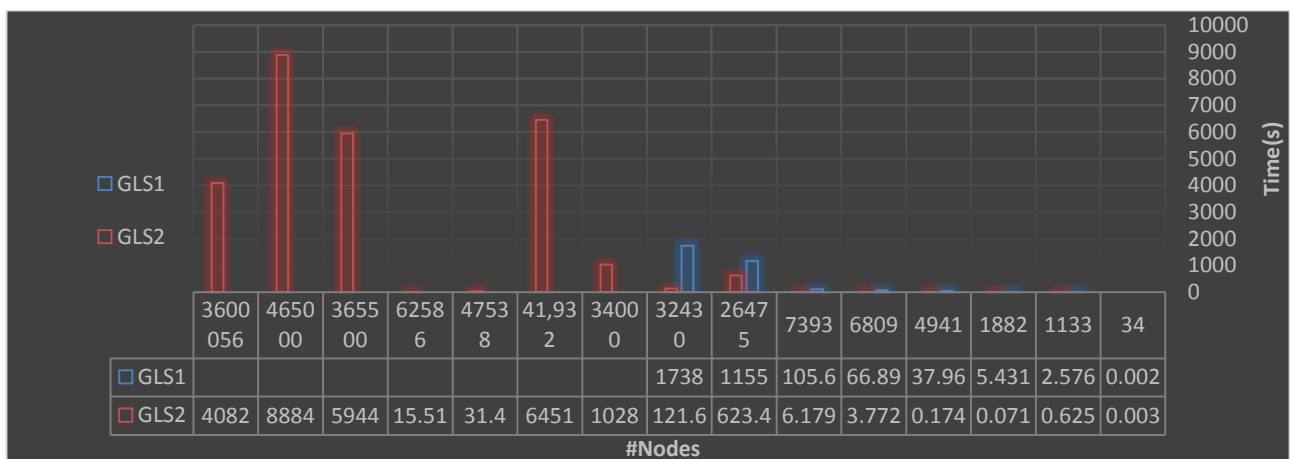
ارزیابی NGLS در این مقاله در دو بخش ارزیابی دقت و زمان اجرا انجام شده است. این الگوریتم‌ها با زبان پایتون نسخه‌ی 3.8.5 با استفاده از Visual StudioCode روی سیستم عامل windows10 پیاده‌سازی شده‌اند، همچنین سیستم مورد استفاده برای ارزیابی دارای 8 گیگابایت رم و پردازنده‌ی اینتل Core i7 6700HQ است.

۴-۱- دقت الگوریتم

تغییر الگوریتم GLS جهت رفع محدودیت زمانی آن، هیچ تغییری در خروجی آن ایجاد نکرده، و دقت بالای الگوریتم حفظ شده است. این مسئله از طریق دریافت کندال τ برابر با یک، از خروجی NGLS و GLS روی داده‌ی دنیای واقعی و شبکه‌های شبیه‌سازی شده با پارامترهای متفاوت اثبات شده است. همچنین کندال τ روی خروجی هر دو الگوریتم و مدل انتشار SIR با $t=500$ و $\lambda=0.1$ محاسبه شده است. نتایج آزمایشات در جدول (۱) نشان داده شده است.

جدول (۱): نتایج آزمایشات سنجش دقت NGLS

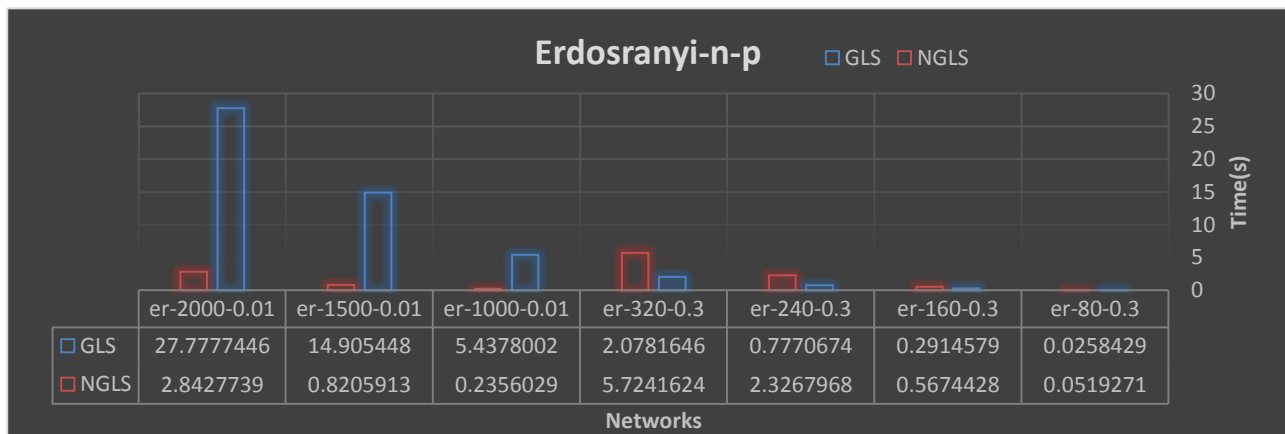
شبکه‌ها	گره‌ها	یال‌ها	τ (NGLS, GLS)	τ (SIR, NGLS)	τ (SIR, GLS)
karate	۳۴	۷۸	۱	۰.۸۲۱۷۴۶	۰.۸۲۱۷۴
EMAIL	۱۱۳۳	۵۴۵۱	۱	۰.۸۵۱۳۳	۰.۸۵۱۳۳
Powergrid	۱۷۸۲	۱۷۴۰	۱	۰.۸۴۳۱۳	۰.۸۴۳۱۳
er-80-0.3	۸۰	۳۷	۱	۰.۷۹۳۲۲	۰.۷۹۳۲۲
wat-100-3-0.2	۱۰۰	۱۰۰	۱	۰.۹۶۵۳۲	۰.۹۶۵۳۲



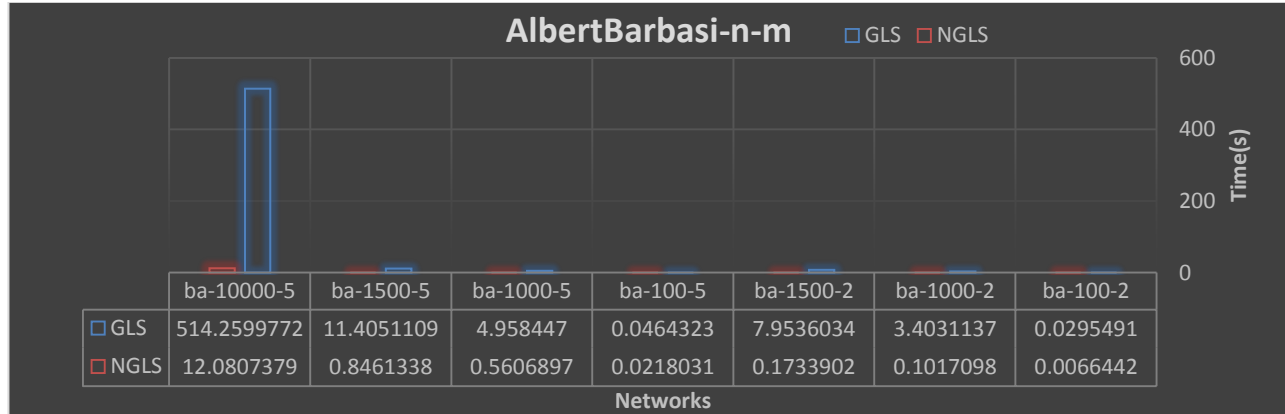
شکل (۳): نمودار مقایسه‌ی زمان اجرای الگوریتم GLS و NGLS روی شبکه‌های دنیای واقعی

رشد نمایی آن رشد خطی دارد. اما با نزدیک شدن p به یک متوسط طول مسیر کاهش می‌یابد و تجمع گره‌ها در یک خوشه‌ی بزرگ زیاد می‌گردد. همچنین توزیع درجه‌ی این شبکه‌ها براساس توزیع پواسون است. بنابراین ارتباطات گره‌ها در شعاع یک و دو زیاد شده بار محاسباتی الگوریتم NGLS

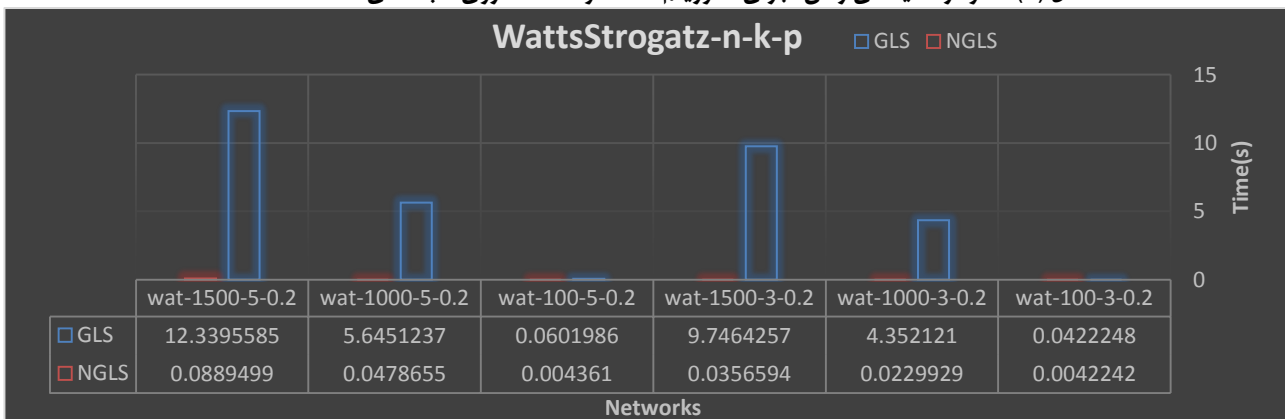
تصادفی، جهان کوچک و مستقل از مقیاس در شکل‌های (۴) و (۵) و (۶) نیز بررسی شده است. در تمام نمودارهای مذکور محور عمودی نشانگر زمان بر حسب ثانیه و محور افقی نشانگر شبکه‌های شبیه‌سازی شده است. شبکه‌ها تصادفی بر اساس مدل Erdős-ranyi: طبق نتایج برای p کوچک‌تر از ۰.۰۵ زمان الگوریتم NGLS کمتر از GLS است و در مقابل



شکل (۴): نمودار مقایسه‌ی زمان اجرای الگوریتم GLS و NGLS روی شبکه‌های Erdos-Ranyi



شکل (۵): نمودار مقایسه‌ی زمان اجرای الگوریتم GLS و NGLS روی شبکه‌های Albert-Barbasi



شکل (۶): نمودار مقایسه‌ی زمان اجرای الگوریتم GLS و NGLS روی شبکه‌های Wattst-Strogatzi

- [1] S. Rosenthal, "Detecting influencers in social media discussions," 21, XRDS: Crossroads, The ACM Magazine for Students, 1, 2014.
- [2] S. Gao, J. Ma, Z. Chen, G. Wang, and C. Xing, "Ranking the spreading ability of nodes in complex networks based on local structure," *Physica A: Statistical Mechanics and its Applications*, vol. 403, pp. 130-147, 2014.
- [3] G. Maji, A. Namtirtha, A. Dutta, and M. C. Malta, "Influential spreaders identification in complex networks with improved k-shell hybrid method," *Expert Systems with Applications*, vol. 144, p. 113092, 2020.
- [4] M. M. Tulu, R. Hou, and T. Younas, "Identifying influential nodes based on community structure to speed up the dissemination of information in complex network," *IEEE Access*, vol. 6, pp. 7390-7401, 2018.
- [5] Y. Yang, L. Yu, X. Wang, Z. Zhou, Y. Chen, and T. Kou, "A novel method to evaluate node importance in complex networks," *Physica A: Statistical Mechanics and its Applications*, vol. 526, p. 121118, 2019.
- [6] T. Bian, J. Hu, and Y. Deng, "Identifying influential nodes in complex networks based on AHP," *Physica A: Statistical Mechanics and its Applications*, vol. 479, pp. 422-436, 2017.
- [7] T. Bian and Y. Deng, "A new evidential methodology of identifying influential nodes in complex networks," *Chaos, Solitons & Fractals*, vol. 103, pp. 101-110, 2017.
- [8] X. Chen, M. Tan, J. Zhao, T. Yang, D. Wu, and R. Zhao, "Identifying influential nodes in complex networks based on a spreading influence related centrality," *Physica A: Statistical Mechanics and its Applications*, vol. 536, p. 122481, 2019.
- [9] D. Mittal, P. Suthar, M. Patil, P. Pranaya, D. P. Rana, and B. Tidke, "Social Network Influencer Rank Recommender Using Diverse Features from Topical Graph," *Procedia Computer Science*, vol. 167, pp. 1861-1871, 2020.
- [10] J. Dai, B. Wang, J. Sheng, and G. Duan, "Identifying Influential Nodes in Complex Networks Based on Local Neighbor Contribution," *IEEE Access*, vol. 7, pp. 131719-131731, 2019.
- [11] J. Sheng et al., "Identifying influential nodes in complex networks based on global and local structure," *Physica A: Statistical Mechanics and its Applications*, vol. 541, p. 123262, 2020.
- [12] T. Bian and Y. Deng, "Identifying influential nodes in complex networks: A node information dimension approach," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 28, no. 4, p. 043109, 2018.
- [13] A. Farooq, G. J. Joyia, M. Uzair, and U. Akram, "Detection of influential nodes using social networks analysis based on network metrics," in *2018 International Conference on Computing, Mathematics and Engineering Technologies (iCoMET)*, 2018, pp. 1-6: IEEE.
- [14] R. Y. Siti Nurulain Mohd Rum, Lilly Suriani Affendey, "Detecting Influencers in Social Media Using Social Network
- [15] Analysis (SNA)," *International Journal of Engineering & Technology*, vol. 10, 2018.

را زیاد نموده و باعث می‌شود در این نمونه‌ی خاص زمان اجرای NGLS بیشتر از قبل شود.

شبکه‌های مستقل از مقیاس بر اساس مدل Albert-Barbasi: آزمایشات در این شبکه‌ها با اندازه‌های متفاوت و دو متوسط درجه‌ی ۲ و ۵ انجام شده است. همانطور که از شکل (۵) مشخص است با افزایش مقدار m به عنوان متوسط درجات، این زمان در هر دو الگوریتم افزایش یافته است. اما همچنان زمان الگوریتم NGLS نسبت به GLS بسیار ناچیز است و حداقل ۸۰ درصد کاهش داشته است. توزیع درجه‌ی توان-قانون و کوچک بودن متوسط طول مسیر از ویژگی‌های مشترک شبکه‌های دنیای واقعی و شبکه‌های Albert-Barbasi است که در بخش‌های قبلی این کاهش زمان در شبکه‌های دنیای واقعی اثبات شده است. علت این کاهش زمان ساختار این شبکه‌ها است که تعداد گره با روابط بالا در آن‌ها کم است و رفع کردن شمارش تعداد همسایه-های مشترک تمام گره‌ها با یکدیگر منجر به کاهش بسیار زیاده زمان اجرا شده است.

شبکه‌های جهان کوچک بر اساس مدل Watts-Strogatz: در این ارزیابی زمان اجرا روی شبکه‌ها با اندازه‌های یکسان و متوسط درجه‌ی ۳ برای $p=0.2$ و $p=0.8$ اندازه‌گیری شده است. این شبکه‌ها دارای ضریب خوشگی بالا و متوسط طول مسیر کوتاه است همچنین توزیع درجه‌ی آن‌ها از نوع پواسون است. متوسط طول مسیر و ضریب خوشگی به مقدار p وابسته است و هرچه قدر به یک نزدیکتر باشد به شبکه‌های Erdős- renyi شباهت می‌یابد. طبق نمودار شکل (۶) زمان اجرای NGLS نسبت به نسخه‌ی GLS بسیار ناچیز است.

۵- جمع بندی

در این پژوهش ساختار الگوریتم GLS به گونه‌ای تغییر کرده است که ضمن حفظ دقت، ضعف پیچیدگی زمانی آن رفع شده و امکان اعمال آن برای شبکه‌های بزرگ فراهم شده است. زمان اجرای NGLS در شبکه‌های دنیای واقعی حداقل ۵۰٪ نسبت به GLS کاهش داشته است و در برخی از شبکه‌های بزرگ زمان اجرای الگوریتم بسیار ناچیز است. علت این رویداد این است که زمان اجرا تنها تابع اندازه‌ی شبکه نیست و عوامل توپولوژیک از جمله مدل توزیع درجات، متوسط توزیع درجات، متوسط طول مسیر و ضریب خوشگی در آن دخیل هستند. طبق نتایج آزمایشات روی شبکه‌های شبیه‌سازی شده با پارامترهای متفاوت زمان NGLS بیشترین کاهش را ابتدا روی شبکه‌های جهان کوچک و بعد از آن شبکه‌های مستقل از مقیاس و تصادفی داشته است. NGLS الگوریتمی با دقت بالا و بدون محدودیت زمانی است که در انواع شبکه‌ها در زمینه‌های مختلف اجتماعی، سیاسی، پزشکی و... قابل استفاده است. ایجاد امکان اعمال NGLS روی شبکه‌های وزن‌دار جزء برنامه‌های آینده‌ی این پژوهش است.

مراجع

- [16] A. M. Pudjajana, D. Manongga, A. Iriani, and H. D. Purnomo, "Identification of Influencers in Social Media using Social Network Analysis (SNA)," in *2018 International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)*, 2018, pp. 400-404: IEEE.
- [17] C. Borrego, J. Borrell, and S. Robles, "Hey, influencer! Message delivery to social central nodes in social opportunistic networks," *Computer Communications*, vol. 137, pp. 81-91, 2019.
- [18] C. Salavati, A. Abdollahpouri, and Z. Manbari, "Ranking nodes in complex networks based on local structure and improving closeness centrality," *Neurocomputing*, vol. 336, pp. 36-45, 2019.
- [19] L. Lü, T. Zhou, Q.-M. Zhang, and H. E. Stanley, "The H-index of a network node and its relation to degree and coreness," *Nature communications*, vol. 7, no. 1, pp. 1-7, 2016.
- [20] A. Zareie and A. Sheikahmadi, "EHC: Extended H-index Centrality measure for identification of users' spreading influence in complex networks," *Physica A: Statistical Mechanics and Its Applications*, vol. 514, pp. 141-155, 2019.
- [21] A. Namtirtha, A. Dutta, and B. Dutta, "Identifying influential spreaders in complex networks based on kshell hybrid method," *Physica A: Statistical Mechanics and its Applications*, vol. 499, pp. 310-324, 2018.
- [22] J. Wang, C. Li, and C. Xia, "Improved centrality indicators to characterize the nodal spreading capability in complex networks," *Applied Mathematics and Computation*, vol. 334, pp. 388-400, 2018.
- [23] J. Zhao, Y. Wang, and Y. Deng, "Identifying influential nodes in complex networks from global perspective," *Chaos, Solitons & Fractals*, vol. 133, p. 109637, 2020.