



ارائه یک روش جدید برای نمونه‌برداری از شبکه‌های اجتماعی

زینب سقطی جلالی^۱، محمدرضا میبدی^۲، علیرضا رضوانیان^۳

^۱ آزمایشگاه محاسبات نرم، دانشکده مهندسی کامپیوتر و فناوری اطلاعات، دانشگاه صنعتی امیرکبیر، تهران
s.jalali@aut.ac.ir

^۲ آزمایشگاه محاسبات نرم، دانشکده مهندسی کامپیوتر و فناوری اطلاعات، دانشگاه صنعتی امیرکبیر، تهران
mmeybodi@aut.ac.ir

^۳ آزمایشگاه محاسبات نرم، دانشکده مهندسی کامپیuter و فناوری اطلاعات، دانشگاه صنعتی امیرکبیر، تهران
a.rezvanian@aut.ac.ir

چکیده

در سال‌های اخیر تحلیل شبکه‌های اجتماعی توسعه چشمگیری داشته است، اما با توجه به بزرگی شبکه‌های اجتماعی، اعمال الگوریتم‌ها و تحلیل مستقیم بر روی شبکه وجود ندارد. بنابراین از نمونه‌برداری برای انجام تحلیل شبکه‌های اجتماعی استفاده می‌شود. اکثر روش‌های نمونه‌برداری به انتخاب تصادفی رؤوس یا یال‌ها تأکید دارند و به خوبی نمی‌توانند خصوصیات شبکه اولیه را حفظ کنند در حالی که هدف اصلی از نمونه‌برداری، تولید شبکه نمونه‌ای با شbahت بالا به شبکه اولیه است تا بتوان به تحلیل‌های بدست آمده به خوبی اعتماد کرد. در این مقاله، یک روش جدید برای نمونه‌برداری از شبکه‌های اجتماعی به کمک درخت‌های پوشای گراف‌های بدون وزن ارائه می‌گردد. در ابتدا چندین درخت پوشای نقاط شروع تصادفی بدست می‌آید، سپس یال‌هایی به دست آمده در این درخت‌های پوشای براساس تعداد دفاتر تکرارشان رتبه‌بندی شده و در نهایت گراف نمونه توسط مجموعه‌ای از یال‌هایی با رتبه بالا ایجاد می‌شود. ایده اصلی این روش بر این اساس استوار است که یال‌هایی که در درخت‌های پوشای مختلف تکرار می‌شوند دارای خصوصیات ساختاری مشابه با شبکه هستند و نمونه‌ای که حاوی این یال‌ها است، خصوصیات مختلفی از شبکه اولیه را دارد. برای ارزیابی کارایی روش پیشنهادی، آزمایشات متعددی ترتیب داده شده است و نتایج الگوریتم پیشنهادی با سایر الگوریتم‌های نمونه‌برداری مقایسه شده‌اند که نتایج بدست آمده بیان گر این است که روش پیشنهادی به مراتب بهتر از سایر روش‌های متداول عمل می‌کند.

کلمات کلیدی

شبکه‌های اجتماعی، تحلیل شبکه‌های اجتماعی، نمونه‌برداری شبکه، درخت پوشای

خروجی می‌شود که قادر نیست خصوصیات ارتباطاتی شبکه اولیه را حفظ کند.

(۲) گراف نمونه‌برداری شده به سمت رؤوس با درجات بالاتر متمایل است، زیرا

یال‌های همسایه این رؤوس به مراتب بیشتر از سایر رؤوس هستند [۲]. برای رفع این مشکلات و ارائه نمونه‌ای که نماینده مناسبی از گراف اولیه باشد، به جای اختصاص احتمال یکنواخت به یال‌ها، احتمالی متناسب با تعداد تکرارهای این یال‌ها در تعداد مناسبی از درخت‌های پوشای رؤوس اولیه متفاوت برای شبکه اختصاص داده می‌شود [۳] [۴] [۵]. در این مقاله یک روش نمونه‌برداری مؤثر مبتنی بر درخت پوشای منظور حفظ خصوصیات شبکه اولیه بر مبنای انتخاب یال پیشنهاد شده است. اهمیت درخت پوشای این اصل استوار است که حلقه‌ها را در گراف از بین می‌برد و مسیر کوتاهی بین راس اولیه و سایر رؤوس فراهم می‌کند [۶]، یک درخت پوشای حاوی

۱- مقدمه

در سال‌های گذشته شاهد پیشرفت چشمگیر روش‌های نمونه‌برداری از شبکه‌ها بوده‌ایم. اما اکثر این روش‌ها تنها بر نمونه‌برداری از رؤوس یا یال‌ها به صورت مستقل تأکید می‌کنند، بنابراین شبکه خروجی برای بازیابی خصوصیات شبکه اولیه به اندازه کافی مناسب نیست [۱]. اکثر روش‌های موجود مبتنی بر روش‌های انتخاب تصادفی یال یا نواد عمل می‌کنند. مشکلات زیادی در روش نمونه‌برداری انتخاب یال تصادفی وجود دارد که برخی از این مشکلات عبارتند از: ۱) گراف نمونه‌برداری شده مشکل از گراف‌های کوچک نامتصل زیادی است که باعث ایجاد قطر بزرگ در شبکه



نمونهبرداری یال‌ها و نمونهبرداری براساس همبندی^۶ تقسیم می‌کند. در نمونهبرداری براساس همبندی، انتخاب رئوس یا یال‌ها بیشتر به همبندی گراف اولیه بستگی دارد. اسنوبال^۷ نمونهای از نمونهبرداری براساس همبندی است [8].

از دیدگاه دوم، روش‌های نمونهبرداری به دو دسته پیمایش گراف^۸ و قدمزن تصادفی تقسیم می‌شوند. در دسته قدمزن تصادفی، رئوس می‌توانند مجدداً پیمایش شوند. این دسته شامل قدمزن تصادفی ساده و مشتقات این روش است. در حالت کلی روش‌های قدمزن تصادفی به رئوس با درجات بالا تمایل‌لند (اریب) که البته این مشکل به راحتی قابل حل است. در دسته پیمایش گراف، نمونهبرداری بدون جایگزینی انجام می‌شود و هر راس حداقل یک بار پیمایش می‌شود. تفاوت روش‌های این دسته در زمان پیمایش رئوس است. این دسته شامل نمونهبرداری اول سطح، نمونهبرداری اول عمق^۹، فارست فایر، اسنوبال و پاسخ‌گرا^{۱۰} است. در روش‌های نمونهبرداری بدون جایگزینی، وابستگی‌های پیچیده‌ای بین رئوس نمونهبرداری شده وجود دارد که کار را برای محاسبات ریاضی سخت می‌کند [9].

۳- الگوریتم پیشنهادی

شبکه‌های اجتماعی به صورت گراف $G = \langle V, E \rangle$ با مجموعه رئوس V و مجموعه یال E مدل می‌شوند: رئوس در شبکه بیان گر کاربران شبکه و یال‌ها بیان گر ارتباطات میان این کاربران است. یک نمونه بازنمایی^{۱۱} از یک شبکه اجتماعی بزرگ، یک گراف با مقیاس کوچک با خصوصیات آماری مشخصی است که ویژگی‌های گراف اولیه را حفظ می‌کند. از این نمونه، ویژگی‌های مهم گراف اولیه مانند توزیع، فعلیت کاربران و ارتباط کاربران قبل بازیابی است [10] [11]. علاوه بر این زمانی که نرخ نمونهبرداری پایین است، پیدا کردن روش نمونهبرداری مناسب برای به دست آوردن ساختار اطلاعاتی شبکه اولیه یک چالش بزرگ محسوب می‌شود. برای حل این مشکل، هدف ارائه یک روش نمونهبرداری است که از یال‌های مهم برای نمونهبرداری استفاده کند. فرض بر این است که اگر تعداد مناسبی درخت پوشای نقاط شروع متفاوت داشته باشیم، یال‌هایی که به دفعات بیشتر در این درخت‌ها تکرار می‌شوند را می‌توان به عنوان یال‌های مهم در نظر گرفت. براساس این فرض، الگوریتم جدیدی ارائه می‌شود که اطلاعات مهم گراف اولیه را حتی با نرخ نمونهبرداری پایین پوشش دهد.

روش پیشنهادی، یک الگوریتم با تکرار است که در هر تکرار یک درخت پوشای نقطه شروع تصادفی^{۱۲} به عنوان ریشه ساخته می‌شود. روش پیشنهادی بر پایه این فرض است که یال‌هایی که در اکثر درخت‌های پوشای تکرار می‌شوند، یال‌های مهم‌تری نسبت به سایر یال‌ها هستند و چون تعداد زیادی درخت پوشای در اختیار داریم، نمونه خوبی به دست خواهد آمد. پس از ساخت درخت‌های پوشای به هر یالی که در درخت‌های پوشای تکرار شده است، رتبه‌ای بر مبنای تعداد دفعات تکرار در درخت‌های مختلف تخصیص داده می‌شود. نمونه نهایی با در نظر گرفتن گرافی با مجموعه رئوس m درصدی از مجموعه رئوس اولیه با یال‌های از رتبه بالاتر تشکیل می‌گردد.

باید در نظر داشت که روش‌های نمونهبرداری برای شبکه‌های جهت‌دار و بی‌جهت مناسب هستند. اگر گراف $G = \langle V, E \rangle$ بیان گر گراف بی‌جهت و روودی باشد که در آن V مجموعه رئوس $\{v_1, v_2, \dots, v_n\}$ و $E = \{e_1, e_2, \dots, e_n\}$ مجموعه یال‌ها $E = \{e_1, e_2, \dots, e_n\}$ باشد. شبکه خروجی با

حداقل یک مسیر از هر راس به راس دیگر است و می‌توان یال‌هایی که در این مسیرها تکرار می‌شوند را به عنوان یال‌های مهم در نظر گرفت. بنابراین به نظر می‌رسد با در نظر گرفتن درخت‌های پوشای به عنوان هسته اولیه نمونهبرداری، می‌توان نمونهای داشت که شامل یال‌های مهم باشد و خصوصیات گراف اولیه را حفظ کند. به عبارتی می‌توان هر درخت پوشای گراف را به عنوان ساختاری که نماینده تاثیرات شبکه‌های اجتماعی در جامعه نظر گرفت. برای مثال در [7] که درباره تاثیرات شبکه‌های اجتماعی در می‌دانش آموزان در مدل سازی ۱۸ ماهه‌ای که از ارتباطات دانش آموزان مدرسه "وایت میدووسترن"^{۱۳} در شبکه‌های اجتماعی انجام شده است. ۵۲٪ روابط دوستانه بین دانش آموزان در یک درخت پوشای بزرگ که زنجیره‌ای از روابط است، قابل نمایش است. همچنین از آن جایی که بسیاری از روش‌های نمونهبرداری مشکل اریبی و تمایل به سمت رئوس یا یال‌های خاصی دارند که نمونهبرداری را غیر یکنواخت می‌کند، در مفهوم درخت پوشای تمامی رئوس شرکت دارند و مشکل اریبی به سمت رئوس با درجات پیش نمی‌آید. از طرفی با توجه به متفاوت بودن یال‌ها در درخت‌های پوشای مختلف، با بالا بردن تعداد درخت‌ها، مشکل اریبی به سمت یال‌های خاص نیز برطرف می‌شود.

بنابراین با داشتن تعداد کافی درخت پوشای به عنوان مبنای نمونهبرداری می‌توان انتظار داشت نمونه خوبی به دست آید. از مزایای استفاده از درخت پوشای به عنوان هسته نمونهبرداری می‌توان به موارد زیر اشاره کرد: (۱) سادگی ساختار درخت پوشای که پیاده‌سازی نمونهبرداری کارا را ساده می‌کند و (۲) نمونهبرداری مبتنی بر درخت پوشای به جزئیات ساختاری و هندسی گراف و استگی ندارد و به سادگی و به صورت مستقل عمل می‌کند و همچنین به تغییرات هندسی و ساختاری نیز وابسته نیست. یک درخت، ساختار ساده‌ای برای نمایش ارتباطات متقابل رئوس است بر همین اساس می‌توان مسئله نمونهبرداری را به مسئله یافتن مجموعه یال‌های مناسب که در درخت پوشای وجود دارند تبدیل نمود.

ادامه این مقاله به این صورت سازماندهی می‌شود. بخش ۲ یک دید کلی از کارهای انجام شده برای نمونهبرداری فراهم می‌کند. در بخش ۳ روش نمونهبرداری پیشنهادی ارائه می‌شود. کارایی روش پیشنهادی به کمک آزمایشات در بخش ۴ بررسی می‌شود و بخش ۵ یک تتجه‌گیری کلی را ارائه می‌دهد.

۲- کارهای انجام شده

از دیدگاه‌های مختلف، روش‌های نمونهبرداری به دسته‌های مختلفی تقسیم می‌شوند که در این بخش این طبقه‌بندی‌ها آمده است.

از دیدگاه اول روش‌های نمونهبرداری به دو دسته نمونهبرداری رئوس^{۱۴} و نمونهبرداری یال‌ها^{۱۵} تقسیم می‌شود. همان‌طوری که از نامشان مشخص است در نمونهبرداری رئوس، عملیات نمونهبرداری بر روی رئوس انجام می‌شود در حالی که یال‌هایی که بین این رئوس قرار دارند تغییری نمی‌کنند. نمونهبرداری اول سطح، قدمزن تصادفی متروپلیس^{۱۶}، نمونه‌هایی از این روش‌های نمونه‌برداری هستند. از طرفی دیگر، نمونهبرداری یال، گراف نمونه را با انتخاب از میان یال‌های گراف اصلی می‌سازد و سپس رئوس ابتدا و انتهای این یال‌ها را می‌افزاید. نمونهبرداری فراتری^{۱۷} مثالی از روش نمونهبرداری یال است. البته برخی به طبقه‌بندی دوم، دسته سومی نیز اضافه می‌کنند، در این طبقه‌بندی، روش‌های نمونهبرداری به سه دسته نمونهبرداری رئوس،

برای کلیه k های بزرگتر از 0 . توزیع ضریب خوشی بیان گر مجموعه‌ای از رؤوس با ضریب خوشی c برای کلیه c های بین 0 و 1 است. از نتایج این توزیع‌ها در آزمون‌های آماری کولموگروف و فاصله نرمال استفاده می‌شود.

۱. **کولموگروف اسمیرنوف^۳**: برای تخمین فاصله بین دو توزیع تجمعی به کار می‌رود. این معیار به طور گسترده برای مقایسه دو توزیع به کار می‌رود. پارامتر X بیانگر متغیر تصادفی و F_1, F_2 دو تابع چگالی تجمعی^۴ (cdf) را نشان می‌دهند.

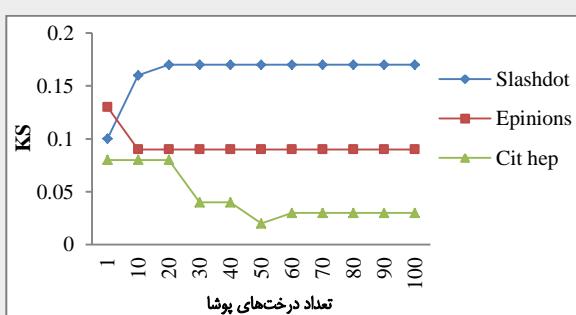
$$KS(F_1, F_2) = \max_x |F_1(x) - F_2(x)| \quad (1)$$

۲. فاصله نرمال شده L_2 : در برخی موارد، اگر مقادیر بردارها اعداد صحیح نباشند (کوچکتر از یک باشند) از فاصله نرمال شده L_2 استفاده می‌کنیم. برای مثال برای محاسبه فاصله بین دو بردار از مقادیر شبکه از فرمول زیر به دست می‌آید.

$$L_2(p, q) = \frac{\|p - q\|}{\|p\|} \quad (3)$$

۱-۴ آزمایش اول: تعداد درخت‌های پوشای مورد نیاز

این آزمایش برای به دست آوردن کمینه تعداد درخت‌های پوشای مورد نیاز برای رسیدن به نرخ نمونه‌برداری مورد نظر انجام می‌شود. بدین منظور، آزمایشات با تعداد درخت‌های پوشای مختلف از ۱ عدد تا ۱۰۰ عدد با نرخ افزایش ۱۰ تا ی بآ نرخ نمونه‌برداری 30% برای بررسی بهترین تعداد درخت‌ها تکرار می‌شوند. نتایج این آزمایش برای مجموعه داده‌های جدول ۱ در شکل‌های ۲ تا ۵ ارائه شده است. بررسی نتایج این آزمایش بیان گر این است که حداقل ۵۰ درخت پوشای برای نمونه‌برداری مورد نیاز است. زیرا نتایج آزمایشات از ۱ تا ۵۰ درخت تغییر می‌کند و از ۵۰ به بعد تقریباً ثابت می‌ماند.



شکل ۲: آزمون KS برای توزیع درجه برای نرخ نمونه‌برداری 30%

نمایش داده می‌شود. شبکه $G_s(V_s, E_s)$ نمونه‌ای از گراف $G(V, E)$ است که در آن $V = \{v_1, v_2, \dots, v_n\}$ مجموعه رؤوس و $f: G \rightarrow G_s$ مجموعه یال‌هایی است که با تابع $f(v_i, v_j) \subseteq V \times V$ به دست آمداند. همچنین $G_s(V_s, E_s)$ یک درخت پوشای گراف $G(V, E)$ است که در آن $V = \{v_1, v_2, \dots, v_n\}$ مجموعه رؤوس و $f: G \rightarrow G_s$ مجموعه یال‌ها با تابع $f(v_i, v_j) \subseteq V \times V$ است. با این تابع، G_s باید متصل بوده و G دارای مجموعه رؤوس یکسان باشد و مجموعه $E_s = |V_s| - 1$ تعداد یال‌هاست. شکل ۱ شبکه کد این الگوریتم پیشنهادی که SST نام دارد را نشان می‌دهد.

Input: Graph $G = \langle V, E \rangle$,

Assumptions

Let k denotes the number of computed spanning trees,
 L : list of selected nodes with size of n ,
 P : Sampling rate

Output: Sample graph $G_s = \langle V_s, E_s \rangle$

Begin

Let τ_t denotes the spanning tree at iteration t ,

$t \leftarrow 1$:

While ($t < k$)

Select v_s randomly as initial vertex of spanning tree;

Find the spanning tree τ_t from v_s ;

$t \leftarrow t + 1$

End While

$L_s \leftarrow$ Compute number of visiting edge

$L \leftarrow$ Sort L_s in descending order

$G_s \leftarrow$ Select top P percent from L

End algorithm

شکل ۱: شبکه کد الگوریتم

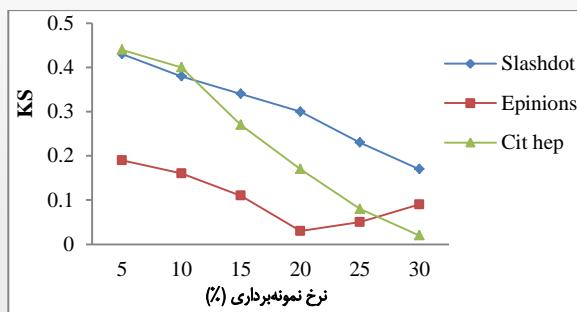
۴- نتایج شبیه‌سازی

برای نشان دادن کارایی الگوریتم پیشنهادی، این الگوریتم بر مجموعه داده‌های مختلف اجرا شده و نتایج با سایر روش‌های متداول نمونه‌برداری مقایسه شده است. مجموعه داده‌هایی که در آزمایشات به کار می‌روند در جدول ۱ آمده‌اند.

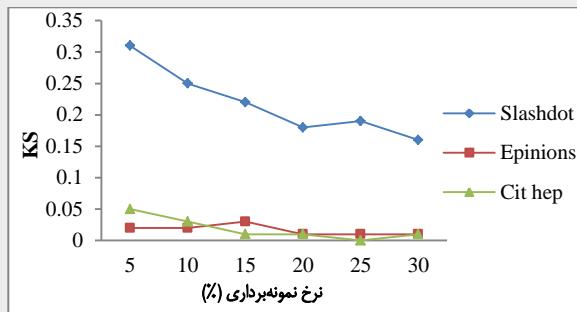
جدول ۱: مجموعه داده‌های آزمایشی

شبکه	تعداد یال‌ها	تعداد رؤوس	توضیح
Cit-HepPh	۳۴۵۴۶	۴۲۱۵۷۸	یک شبکه استناد مقاله فیزیک
Epinions	۷۵۸۷۹	۵۰۸۸۳۷	یک شبکه اجتماعی برخط مبتنی بر اعتماد
Slashdot0902	۸۲۱۶۸	۹۴۸۴۶۴	سایت خبری مرتبط با تکنولوژی

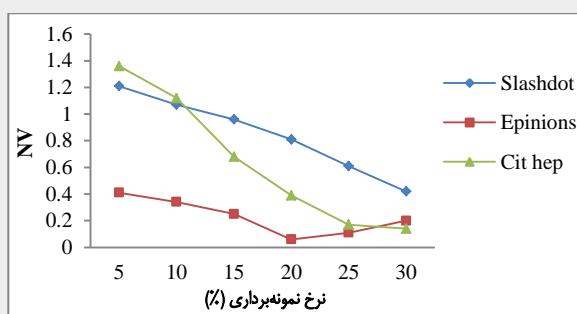
برای بررسی کارایی الگوریتم پیشنهادی، چندین آزمایش بر روی مجموعه داده‌های جدول ۱ انجام شد. کلیه آزمایشات براساس آزمون کولموگروف اسمیرنوف (KS) و آزمون بردار نرمال (NV) برای توزیع‌های آماری به دست آمده از گراف اولیه و نمونه بررسی می‌شوند. همچنین نتایج با نتایج مشابه به دست آمده از الگوریتم‌های نمونه‌برداری انتخاب راس تصادفی، انتخاب یال تصادفی و قدمزن تصادفی متropolیس مقایسه می‌گردند [12]. توزیع درجه به طور گسترده توسط بسیاری از محققان برای فهم میزان اتصال^۳ گراف مورد مطالعه قرار گرفته است و عبارت است از: مجموعه‌ای از رؤوس با درجه k



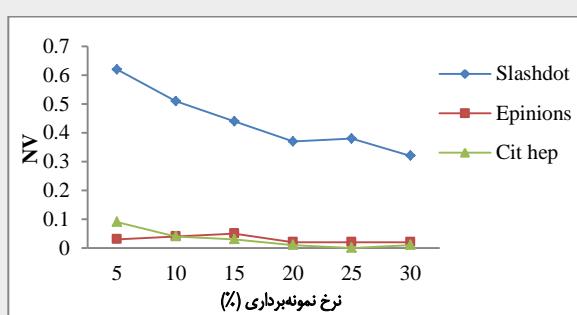
شکل ۶: آزمون KS برای ضریب خوشگی برای نرخ های نمونهبرداری ۵ تا ۳۰٪



شکل ۷: آزمون KS برای ضریب خوشگی برای نرخ های نمونهبرداری ۵٪ تا ۳۰٪



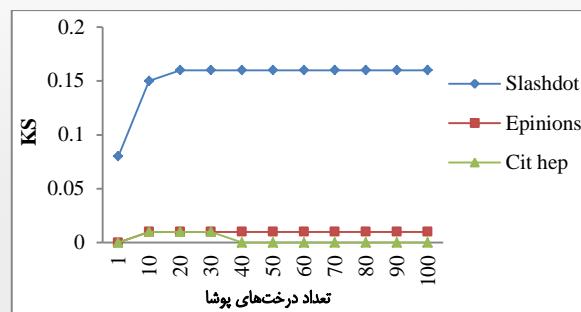
شکل ۸: آزمون NV برای توزیع درجه برای نرخ های نمونهبرداری ۵٪ تا ۳۰٪



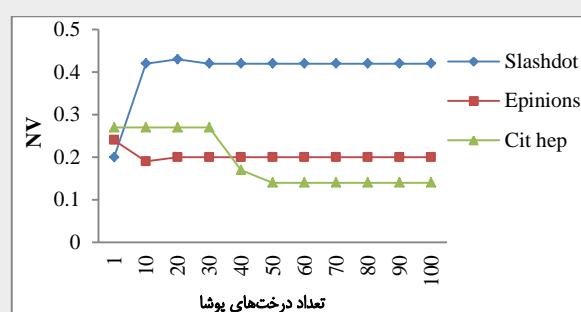
شکل ۹: آزمون NV برای ضریب خوشگی برای نرخ های نمونهبرداری ۵٪ تا ۳۰٪

۳-۳ آزمایش سوم: مقایسه کارایی الگوریتم پیشنهادی با سایر الگوریتمها

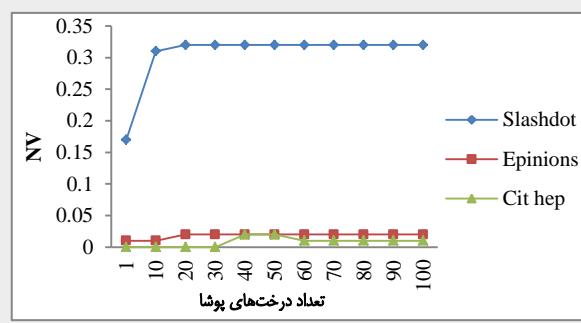
این آزمایش برای مقایسه کارایی الگوریتم پیشنهادی (SST) با الگوریتم‌های نمونهبرداری راس تصادفی (RN) [13]، نمونهبرداری یال تصادفی (RE) [13]، نمونهبرداری قدمزن تصادفی (RW) [13]، قدمزن تصادفی متروپلیس [14]



شکل ۱۰: آزمون KS برای ضریب خوشگی برای نرخ نمونهبرداری ۵٪ تا ۳۰٪



شکل ۱۱: آزمون NV برای توزیع درجه برای نرخ نمونهبرداری ۵٪ تا ۳۰٪



شکل ۱۲: آزمون NV برای ضریب خوشگی برای نرخ نمونهبرداری ۵٪ تا ۳۰٪

۲-۴ آزمایش دوم: بررسی تاثیر نرخ نمونهبرداری بر کارایی الگوریتم

این آزمایش برای مقایسه کارایی الگوریتم پیشنهادی با نرخ های نمونهبرداری مختلف طراحی شده است. نرخ های نمونهبرداری از ۰.۰۵٪ تا ۰.۳٪ با نرخ افزایش ۰.۰۵٪ در هر گام طراحی شده‌اند. نتایج آزمایشات که در شکل‌های ۶ تا ۹ نشان داده شده‌اند، بیان گر این است که برخلاف بسیاری از کارهای انجام شده مانند [14] [15] که نرخ نمونهبرداری در آن‌ها بالا است، الگوریتم پیشنهادی برای نرخ نمونهبرداری ۵٪ تا ۱۰٪ درصد نیز قابل قبول است. بنابراین هزینه الگوریتم پیشنهادی کمتر از بسیاری از روش‌های متداول است.

۵- نتیجه‌گیری

در این مقاله بر مبنای مفهوم درخت پوشان، روشی برای نمونه‌برداری از شبکه‌های اجتماعی ارائه شد. این روش بر این اصل استوار است که یال‌هایی که در درخت‌های پوشان تکرار می‌شوند، قادر به حفظ اطلاعات ساختاری شبکه هستند. کارایی روش پیشنهادی با اعمال آزمایشاتی بر مجموعه داده‌های مختلف بررسی شد. نتایج بیانگر این است که الگوریتم پیشنهادی نسبت به الگوریتم‌های متداول نمونه‌برداری (راس تصادفی، یال تصادفی، قدمزن تصادفی و قدمزن تصادفی متروبیلیس) بهتر عمل می‌کند.

مراجع

- [1] Lee, S., kim, P., Jeong, H., "Statistical properties of sampled networks", Physical Review, vol. 73, no. 1, p. 016102, 2006.
- [2] Maya, A. S., Berger-Wolf, T. Y., *Benefits of Bias: Towards Better Characterization of Network Sampling*, in Seventeenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2011.
- [3] Granovetter, M., "Network sampling: Some first steps", American Journal of Sociology, pp. 1287-1303, 1976.
- [4] Heckathorn, D., "Respondent-driven sampling: a new approach to the study of hidden populations", Social problems, pp. 174-199, 1997.
- [5] Kolaczyk, E., *Sampling and estimation in network graphs*, in *Statistical Analysis of Network Data*, Springer New York, 2009.
- [6] M. P. Stumpf, C. Wiuf, and R. M. May, *Subnets of scale-free networks are not scale-free: sampling properties of networks*, in Proceedings of the National Academy of Sciences of the United States of America, Vol.102, No.12, PP. 4221-4224, 2005.
- [7] Yoon, S., Lee, S., Yook, S., H., Kim, Y. "Statistical properties of sampled networks by random walks", Physical Review, vol. 75, no. 4, p. 046114, 2007.
- [8] Ahmed, N. K., Berchmans, F., Neville, J., Komella, R., *Time-based sampling of social network activity graphs*, in Eighth Workshop on Mining and Learning with Graphs, 2010.
- [9] Ahmed, N. K., Neville, J., Komella, R., "Network Sampling: From Static to Streaming Graphs" Accepted to appear In ACM Transactions on Knowledge Discovery from Data [Earlier version in arxiv preprint], 2013.
- [10] Pfeil, U., Arjan, R., Zaphiris, P., "Age differences in online social networking—A study of user profiles and the social capital divide among teenagers and older users in MySpace", Computers in Human Behavior, vol. 25, no. 3, p. 643–654, 2009.
- [11] Wilson, C., Boe, B., Puttaswamy, K. P. N., Zhao, B. Y., *User interactions in social networks and their implications*, in fourth ACM European Conference on Computer Systems, 2008.
- [12] Li, Y., Wu, C., luo, P., Zhang, W., "Exploring the characteristics of innovation adoption in social networks: structure, homophily and strategy", Entropy, vol. 15, no. 7, pp. 2662-2678, 2013.
- [13] Leskovec, J., faloutsos, C., *Sampling from Large Graphs*, in twelfth ACM SIGKDD International Conference of Knowledge Discovery and Data Mining, 2006.
- [14] Kurant, M., Gjoka, M., Wang, Y., Almquist, Z. W., Butts, C. T., Markopoulou, A., *Coarse-Grained Topology*

(MHRW) و نمونه‌برداری قدمزن تصادفی حلزونی (S1-S3) [16] با نرخ نمونه‌برداری ۱۵٪ اعمال شده است. مقایسات بر مبنای آزمون کولموگروف اسمیرنوف (KS) و بردار نرمال (NV) برای توزیع درجه (dd) و ضریب خوشگی (cc) انجام می‌شود.

نتایج الگوریتم پیشنهادی با ۱ بار اجرا با ۱۵۰ درخت پوشان و نتایج سایر الگوریتم‌ها با ۳۰ بار اجرا برای نرخ نمونه‌برداری ۱۵٪ به دست می‌آید. همچنین برای بررسی کارایی الگوریتم و بررسی این موضوع که نتایج به دست آمده تا چه اندازه بیانگر بهتر بودن نسبت به سایر الگوریتم‌ها است از آزمون t-test استفاده می‌شود. نتایج به دست آمده از این آزمایش در نتایج بیان گر این است که الگوریتم پیشنهادی کارایی بهتری نسبت به سایر الگوریتم‌ها دارد. نتایج به دست آمده از این آزمایش در جداول ۲ الی ۴ نشان داده شده است.

جدول ۲: مقایسه الگوریتم با سایر الگوریتم‌ها برای مجموعه داده CitHep

SST	MHRW	RE	RW	RN		
۰.۲۷±۰.۰۰۰	۰.۰۷±۰.۰۲	۰.۰۷±۰.۰۰۰	۰.۰۷±۰.۰۰۰	۰.۸۱±۰.۰۱	dd	KS
	×	×	×	✓		
۰.۰۱±۰.۰۰۰	۰.۰۵±۰.۰۰۱	۰.۰۰۰±۰.۰۰۰	۰.۰۲±۰.۰۰۰	۰.۰۵±۰.۰۰۳	cc	NV
	✓	×	✓	✓		
۰.۶۸±۰.۰۰۰	۰.۱۹±۰.۰۰۲	۰.۲۴±۰.۰۰۱	۰.۳۲±۰.۰۰۱	۰.۰۰۰±۰.۱۷	dd	KS
	×	×	×	✓		
۰.۰۳±۰.۰۰۰	۰.۰۸±۰.۰۰۰	۰.۰۱±۰.۰۰۰	۰.۰۰۴±۰.۰۰۰	۰.۰۹±۰.۰۰۰	cc	NV
	✓	×	✓	✗		

جدول ۳: مقایسه الگوریتم شماره ۱ با سایر الگوریتم‌ها برای مجموعه داده Epinion

SST	MHRW	RE	RW	RN		
۰.۰۰۰±۰.۳۹	۰.۰۲±۰.۳۹	۰.۰۰۰±۰.۴۵	۰.۰۰۰±۰.۱۷	۰.۰۲±۰.۱۸	dd	KS
	✓	✓	✓	✗		
۰.۰۰۰±۰.۰۸	۰.۰۱±۰.۰۸	۰.۰۰۰±۰.۱۰	۰.۰۰۰±۰.۰۲	۰.۰۳±۰.۱۰	cc	NV
	✓	✓	✗	✓		
۰.۰۰۰±۰.۵۸	۰.۰۰۲±۰.۵۸	۰.۰۱±۰.۷۱	۰.۰۱±۰.۳۲	۰.۰۶±۰.۳۴	dd	KS
	✓	✓	✓	✗		
۰.۰۰۰±۰.۱۴	۰.۰۰۰±۰.۱۴	۰.۰۰۰±۰.۱۷	۰.۰۰۰±۰.۰۴	۰.۰۰۰±۰.۱۸	cc	NV
	✓	✓	✗	✓		

جدول ۴: مقایسه الگوریتم با سایر الگوریتم‌ها برای مجموعه داده Slashdot

SST	MHRW	RE	RW	RN		
۰.۰۰۰±۰.۳۴	۰.۰۲±۰.۳۹	۰.۰۰۰±۰.۴۳	۰.۰۰۰±۰.۰۴	۰.۰۱±۰.۰۷۹	dd	KS
	✓	✓	✗	✓		
۰.۰۰۰±۰.۲۲	۰.۰۱±۰.۳۰	۰.۰۰۰±۰.۳۵	۰.۰۰۰±۰.۰۵	۰.۰۲±۰.۲۸	cc	NV
	✓	✓	✗	✓		
۰.۰۰۰±۰.۹۶	۰.۰۰۴±۰.۶۶	۰.۰۱±۰.۷۶	۰.۰۱±۰.۱۰	۰.۰۳±۰.۱۶	dd	KS
	✓	✓	✗	✓		
۰.۰۰۰±۰.۴۴	۰.۰۰۰±۰.۶۱	۰.۰۰۰±۰.۷۱	۰.۰۰۰±۰.۱۰	۰.۰۰۰±۰.۵۶	cc	NV
	✗	✗	✗	✓		



- Estimation via Graph Sampling*, in *ACM Workshop on Online Social Networks*, 2012.
- [15] Ribeiro, B., Figueiredo, D., deSouzaeSilva, E., Towsley, D., *Characterizing continuous-time random walks on dynamic networks*, in ACM SIGMETRICS joint international conference on Measurement and modeling of computer systems, 2011.
- [16] Pina-Garcia, C. A., Gu, D., “*Spiraling Facebook: an alternative Metropolis–Hastings random walk using a spiral proposal distribution*”, *Social Network Analysis and Mining*, vol. 3, no. 3, pp. 1869-5450, 2013.

زیرنویس‌ها

-
- ¹ White Midwestern
- ² Node sampling
- ³ Edge sampling
- ⁴ Metropolis hasting random walk
- ⁵ frontier
- ⁶ topology
- ⁷ snowball
- ⁸ Graph traversal
- ⁹ Depth first search
- ¹⁰ Respondent driven sampling
- ¹¹ Representative sample
- ¹² connectivity
- ¹³ Kolmogorov Smirnov statistics
- ¹⁴ Cumulative density function