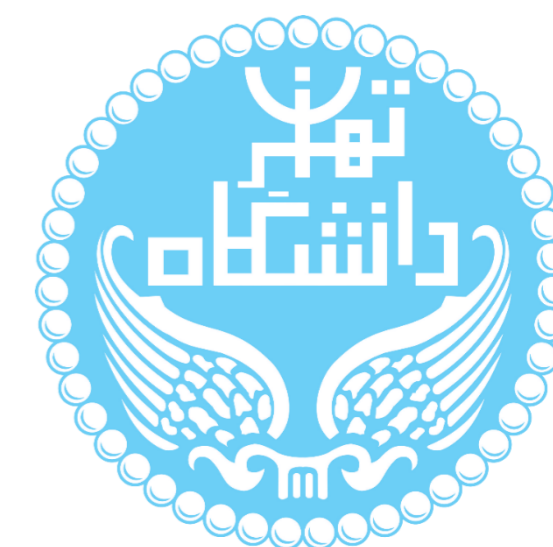


پیدا کردن افراد تأثیرگذار در شبکه‌ی اجتماعی اینستاگرام



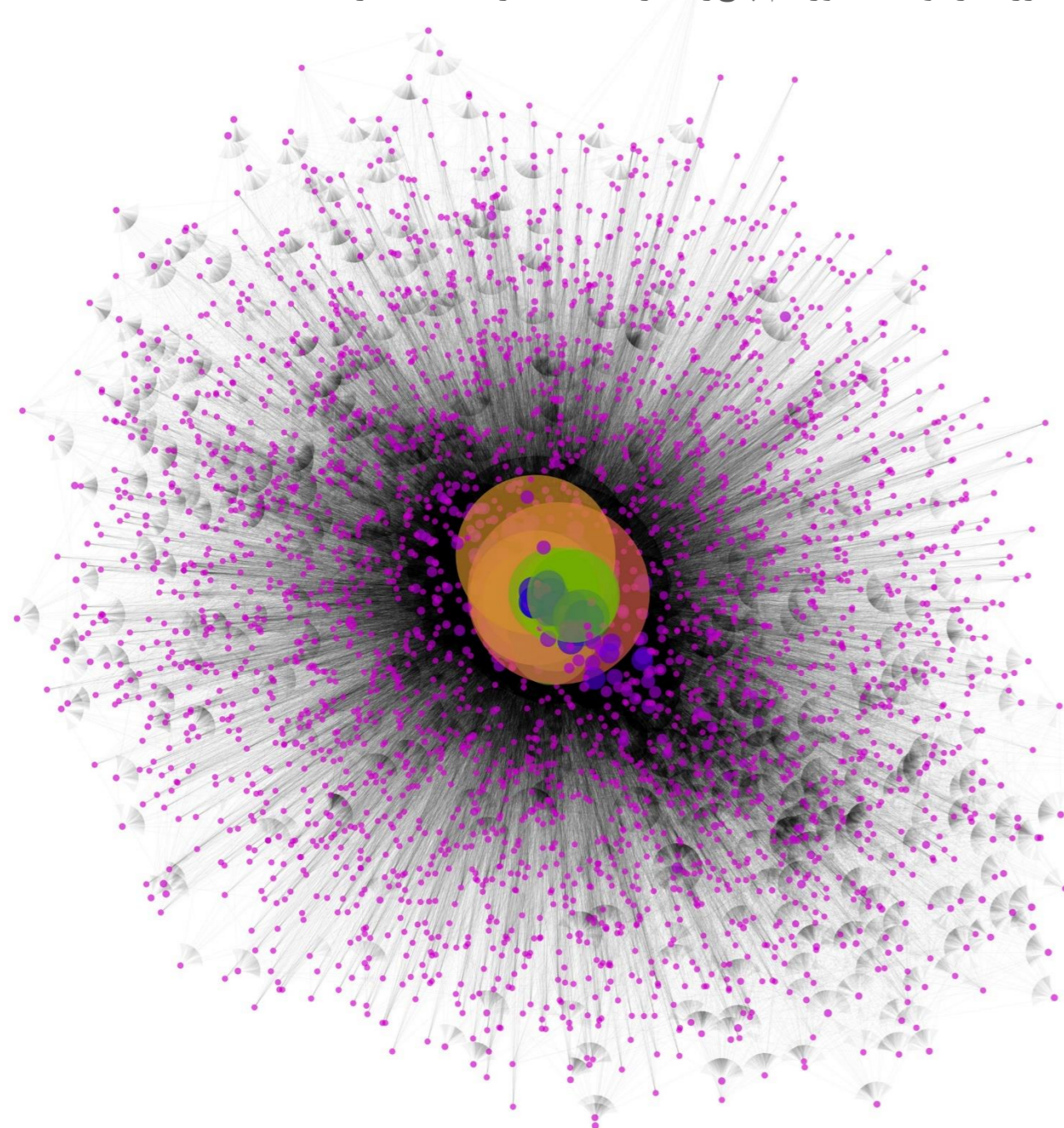
دانشجو: محمدحسین نوروزی
استاد راهنما: دکتر بهنام بهرک
دانشکده مهندسی برق و کامپیوتر، دانشگاه تهران



تحلیل گراف و استخراج افراد تأثیرگذار

به منظور شناسایی افراد تأثیرگذار در گراف کاربران ایرانی اینستاگرام از سه معیار مرکزیت میانوندی، مرکزیت درجه و مرکزیت بر اساس الگوریتم پیچ‌رنک استفاده کردیم. با وجود کوچک بودن اندازه‌ی گراف مدل‌سازی شده، کاربرانی که توسط معیار مرکزیت درجه و پیچ‌رنک به عنوان رهوس تأثیرگذار شناسایی شدند همگی جزو افراد مشهور و دنبال‌کننده‌ی زیاد در بین ایرانیان اینستاگرام بودند. اما نتایج حاصل از معیار مرکزیت میانوندی به علت پراکندگی رهوس جمع‌آوری‌شده، کوچک بودن اندازه‌ی نمونه و نرمال نبودن فرایند جمع‌آوری قابل قبول نبود و انحراف قابل توجه‌ای به سمت کاربران ایرانی موجود در نمونه جمع‌آوری‌شده داشت.

زیرگرافی از گراف کاربران ایرانی اینستاگرام
(رهوس با اندازه‌ی بزرگ‌تر توسط الگوریتم پیچ‌رنک رتبه‌ی بیشتری کسب کرده‌اند)



نام کاربری ده رتبه‌ی اول
rambodjavan1
mahnaz_afshar
taraneh_alidoosti
behnoushtatababai
elnazshakerdoost
behnoosh_bakhtiari
saharghoreyshiofficialpage
rezagolzar
parvizparastouei
golfarahani

مقدمه

شبکه‌های اجتماعی نقش مهمی را در جامعه‌ی امروز ایفا می‌کنند و به بستری برای ارتباط افراد و اشتراک افکار آنان تبدیل شده‌اند. شبکه‌های اجتماعی، به‌نوعی، ارتباط میان انسان‌ها را در فضای مجازی مدل‌سازی کرده‌اند و اینستاگرام یکی از پرستفاده‌ترین موارد آن‌هاست که میلیون‌ها کاربر ساعات زیادی از وقت خود را در آن سپری می‌کنند. در نتیجه، یافتن افراد و موضوعات تأثیرگذار در این شبکه استفاده‌های فراوانی، به‌ویژه در حوزه‌های بازاریابی، خبررسانی و تحلیل رویدادهایی مثل انتخابات‌ها دارد و داده‌های ارزشمندی را فراهم می‌آورد. به‌همین دلیل یافتن افراد تأثیرگذار بین کاربران ایرانی اینستاگرام می‌تواند استفاده‌های تجاری و آماری زیادی داشته باشد. پیش از این فعالیت‌های اندکی در رابطه با استخراج گراف کاربران ایرانی اینستاگرام صورت گرفته است و حتی آمار دقیقی از تعداد کاربران ایرانی اینستاگرام وجود ندارد. در این پروژه سعی شده است روش‌های مختلف برای استخراج گراف کاربران ایرانی اینستاگرام تحلیل شود و چالش‌های موجود در جمع‌آوری این داده‌ها و تشکیل گراف مورد بررسی قرار گیرد. هم‌چنین سعی شده است تا رهوس تأثیرگذار این گراف بر اساس داده‌های جمع‌آوری‌شده به کمک الگوریتم‌های مرکزیت در شناسایی شود.

استخراج داده‌ها و ساخت گراف

برای خزش اطلاعات مربوط به کاربران اینستاگرام، از API که اینستاگرام در دسترس قرار داده است استفاده کردیم و گراف کاربران را با استفاده از کتابخانه‌ی networkx مدل کردیم. به‌منظور تشخیص کاربران ایرانی ابتدا سه آزمون طرح کردیم:

- آزمون حساب کاربری فارسی:** کاربرانی که در حساب کاربری خود متن‌های فارسی به کار برده باشند را ایرانی در نظر گرفتیم. برای تشخیص زبان متن از مترجم گوگل استفاده کردیم. از آنجایی که اکثر کاربران ایرانی از زبان فارسی استفاده نمی‌کنند، استفاده از این آزمون نمی‌توانست تنها راه‌حل ما برای استخراج کاربران ایرانی باشد. هم‌چنین به دلیل نیاز بالای این آزمون به ارتباط با API اینستاگرام، سرعت پیشروی آن نیز کند بود.
- آزمون همسایگی:** فرض کردیم که اگر کاربری ۵ کاربر با شروط آزمون اول را دنبال کند، ایرانی است. در ۴۵۰ نمونه‌ی بررسی شده این آزمون از دقت بالایی برخوردار بود ولی از نظر سرعت مزیتی نسبت به آزمون اول نداشت و با مشکل بزرگ‌شدن بیش از حد گراف روبرو بود. تمام کاربرانی که حداکثر دو رأس از یک کاربر ایرانی فاصله داشتند توسط این آزمون به گراف اضافه می‌شدند و این امر باعث شد که تعداد رهوس موجود در گراف بعد از شناسایی ۵۰۰ ایرانی به ۸۰۰۰۰۰ رأس برسد.
- آزمون همسایگی داخلی:** فرض کردیم که اگر کاربری ۵ کاربر ایرانی را دنبال کند، ایرانی است. نسبت به آزمون دوم سرعت شناسایی ایرانیان بیش از ۱۰ برابر و سرعت رشد اندازه‌ی گراف کمتر از نصف شد. اما هم‌چنان رضایت‌بخش نبود، زیرا هم‌چنان نیاز به ارتباط با سرور بیرونی زیاد بود. شایان ذکر است که با بزرگ‌تر شدن گراف و شناسایی ایرانیان بیشتر، دقت این آزمون نیز کم می‌شد و می‌بایست این مسئله به‌مرور دنبال می‌شد و در صورت نیاز درباره‌ی شروط بازنگری صورت می‌گرفت.

مشاهداتمان نشان می‌داد که علت قبولی در آزمون سوم برای اکثر کاربران همسایگی با عده‌ی محدودی از کاربران ایرانی پرتعداد اینستاگرام است. هم‌چنین می‌دانستیم که آزمون سوم در شرایطی که تعداد کاربران ایرانی نمونه‌ی ما زیاد باشد سریع‌تر عمل می‌کند. به همین منظور تصمیم گرفتیم تا از طریق اشتراک میان دنبال‌کنندگان چند کاربر ایرانی پرتعداد اینستاگرام تعداد ایرانیان موجود در نمونه را افزایش دهیم و سپس به آزمون‌های طراحی‌شده بازگردیم.

دنبال‌کنندگان ۱۰ کاربر ایرانی پرتعداد را استخراج کردیم. اگر فرض کنیم کاربری که دو نفر از آن‌ها را دنبال می‌کند ایرانی است، ۸.۴ میلیون کاربر ایرانی شناسایی شدند و اگر فرض کنیم کاربری که سه نفر از آن‌ها را دنبال می‌کند ایرانی است، ۴.۹ میلیون کاربر ایرانی شناسایی شدند. برای سنجش این ادعا، نمونه‌ای تصادفی شامل ۴۰۰ کاربر از میان این ۴.۹ میلیون نام کاربری انتخاب شد و هیچ کاربر غیر ایرانی‌ای میان آن‌ها نبود. در ادامه اطلاعات مربوط به ۱۲ هزار کاربر ایرانی را استخراج کردیم و به گرافمان اضافه کردیم. نمونه‌ی به‌دست‌آمده شامل ۴.۷ میلیون رأس و ۸.۳ میلیون یال بود.

جمع‌بندی

- در این تحقیق روش‌های مختلفی برای استخراج کاربران ایرانی اینستاگرام بررسی شد و نشان داده شد که روش‌های شناسایی مبتنی بر نزدیکی کاربران ایرانی به یکدیگر در نمونه‌های کوچک عملکرد کندی دارند و سریع‌ترین روش برای تشکیل حداکثر بخش‌های گراف خزش اطلاعات دنبال‌کنندگان کاربرهای ایرانی پرتعداد است. هم‌چنین عملکرد قابل قبول الگوریتم پیچ‌رنک در شناسایی افراد تأثیرگذار مشاهده شد.
- کاربردها: شناسایی افراد تأثیرگذار می‌تواند در زمینه‌ی مؤثرتر کردن و کاهش هزینه‌های تبلیغات و هم‌چنین تحلیل رویدادهای کلان اجتماعی مثل انتخابات‌ها مفید باشد.
- از جمله محدودیت‌های موجود در تحقیق می‌توان به سرعت پاسخ‌گویی پایین API اینستاگرام برای خزش اطلاعات و هم‌چنین تعداد بالای کاربران خصوصی اشاره کرد.
- هم‌چنین بسیاری از رهوس غیرایرانی گراف درجه‌ی کمی دارند و با حذف آن‌ها می‌توان، بدون ایجاد تغییر در ویژگی‌های کلان گراف، حجم گراف را به‌شدت کاهش داد.

مراجع اصلی

- Newman, Mark. *Networks: an introduction*. Oxford university press, 2010.
- Bakshy, Eytan, Jake M. Hofman, Winter A. Mason, and Duncan J. Watts. "Everyone's an influencer: quantifying influence on twitter." In *Proceedings of the fourth ACM international conference on Web search and data mining*, pp. 65-74. ACM, 2011.
- L. Manikonda, Y. Hu, and S. Kambhampati, "Analyzing user activities, demographics, social network structure and user-generated content on instagram," *arXiv preprint arXiv:1410.8099*, 2014.
- Bakshy, S., Shamma, D. A., & Gilbert, E. Faces Engage Us: Photos with Faces Attract More Likes and Comments on Instagram. *CHI 2014*, ACM Press (2014), 965-974.
- A. Pal, A. Herdagdelen, S. Chatterji, S. Taank, and D. Chakrabarti, "Discovery of Topical Authorities in Instagram," *Proceedings of the 25th International Conference on World Wide Web - WWW 16*, 2016.