

پروژه نهایی درس شبکه‌های اجتماعی

دی ماه ۱۴۰۱

مهلت تحویل تا ۱۲ بهمن

در این تمرین شما باید با استفاده از ابزارهای تحلیل شبکه مثل Gephi یا networkx شبکه ضمیمه را تحلیل نمایید. شبکه ضمیمه، شبکه بازنشر (ریتویت) بین اکانت‌های توییتری است که طی ۶۴ روز از ۲۳ شهریور ماه تا ۲۵ آبان ۱۴۰۱ در موضوع اعتراضات مربوط به مهسا امینی (با هشتگ یا عبارت "مهسا امینی" یا هشتگ‌های مرتبط) فعالیت توییتری داشته اند. یال‌های این شبکه جهت دار است و جهت آنها از اکانت ریتویت کننده (ارجاع دهنده) به اکانت منتشر کننده توییت اصلی (مرجع) رسم شده است. وزن یال‌ها نشان دهنده تعداد دفعات بازنشر توییت از اکانت ارجاع دهنده به اکانت مرجع است.

به دلیل بزرگ بودن شبکه (بیش از ۱,۶ میلیون نود) فقط نودهایی که حداقل یک یال با وزن حداقل ۵ داشته اند در گراف ضمیمه آورده شده اند و ما بقی نودها از شبکه حذف شده اند. در نتیجه حدود ۲۴ هزار نود و بیش از ۵۰۰ هزار یال بین آنها بدست آمده است.

ستونهای مربوط به نودها عبارتند از:

- ۱- name : شناسه یکتای اکانت که یک شناسه عددی است.
- ۲- label : برچسب انگلیسی کانال که در حقیقت username توییتری اکانت است. توجه داشته باشید که بعضا دو یا چند اکانت با شناسه های عددی مختلف در بازه‌های زمانی مختلف یوزرنیم یکسانی داشته باشند.
- ۳- tweets : تعداد توییت‌های منتشر شده توسط اکانت در موضوعات مرتبط با مهسا امینی
- ۴- replies : تعداد پاسخ‌های منتشر شده توسط اکانت به توییت‌های دیگران. منظور فقط پاسخ‌هایی است که حاوی عبارات مرتبط با موضوع مهسا امینی است.
- ۵- retweets : تعداد توییت‌های بازنشر شده توسط اکانت در موضوعات مرتبط با مهسا امینی
- ۶- quotes : تعداد نقل قول‌های منتشر شده توسط اکانت در موضوعات مرتبط با مهسا امینی
- ۷- followers : تعداد دنبال کنندگان اکانت در لحظه گزارش. این تعداد ممکن است در حال حاضر تغییر کرده باشد. تعداد دنبال کنندگان برخی اکانتها ممکن است صفر باشد. در این حالت این عدد یا واقعا نشان دهنده این است که این اکانت دنبال کننده‌ای ندارد یا به هر دلیلی امکان گرفتن تعداد دنبال کنندگان این اکانت توسط سامانه پایش وجود نداشته است.
- ۸- impresssion : جمع لایک و ریتوییت توییت‌های منتشر شده توسط این اکانت در موضوعات مرتبط با مهسا

امینی

۹- age: سن اکانت به "روز" از لحظه ایجاد اکانت تا ۲۵ ابان در قالب یک عدد اعشاری. قسمت اعشاری این عدد، کسری از ۲۴ ساعت را نشان میدهد؛ مثلاً سن کمتر از یک روز یعنی اکانت در همان روز ۲۵ ابان ایجاد شده و تنها چند ساعت از ایجاد آن گذشته است.

۱۰- activedays: نشان میدهد این اکانت در چند روز از کل بازه ۶۴ روزه فعالیت داشته است؟ اگر یک اکانت در

یک روز حداقل یک فعالیت -حتی در حد بازنشر داشته- آن روز برای این اکانت، روز فعالیت منظور شده است.

۱۱- reaction: جمع عکس‌العمل‌های این اکانت به توییت‌های دیگران که مساوی است با جمع ستون‌های بازنشر،

پاسخ، و نقل قول

۱۲- activity: جمع فعالیت‌های این اکانت در موضوع مهسا امینی که مساوی است با جمع ستون‌های توییت، و بازنشر

۱۳- robot: احتمال ربات بودن یا سازمانی بودن این اکانت در قالب یک عدد بین صفر و یک. این عدد از روی فاصله

زمانی بین یک توییت و توییت بعدی یا یک ریتوییت و ریتوییت بعدی اکانت محاسبه شده است. اگر درصد زیادی از

فعالیت‌های این اکانت فاصله زمانی کمی دارند به گونه‌ای که چنین فاصله کمی برای یک کاربر عادی امکان پذیر

نباشد، احتمال ربات بودن یا سازمانی بودن اکانت بالا در نظر گرفته شده است. سازمانی بودن اکانت به این معنی

است که اکانت عضو یک گروه سازمان دهی شده است که تولید محتوا در آن توسط یک تیم صورت می‌پذیرد و

اکانت‌های عضو آن سازمان برای تولید محتوا وقت چندانی لازم ندارند و برای بازنشر هم به صورت دستوری عمل

میکند. این اکانت‌ها در فعالیت‌های خود، فقط اهداف سازمانی را در نظر می‌گیرند و افکار شخصی خود را دخیل

نمیکند؛ به همین دلیل نیازی نیست روی ارسال توییت یا ریتوییت، تامل کنند. توجه داشته باشید که این عدد

فقط برای اکانت‌هایی محاسبه شده که تعداد توییت یا ریتوییت بالایی داشته‌اند (تا از نظر آماری عدد احتمال بدست

آمده قابل اعتماد باشد). برای اکانت‌هایی که توییت یا ریتوییت بالایی نداشته‌اند این عدد صفر لحاظ شده است (هر

چند ممکن است همین اکانت‌ها اگر در بازه بزرگتری بررسی شوند مشخص شود ربات‌اند یا عضو یک سازمان

هستند).

۱۴- btsexo: این فیلد دو مقداری (صفر و یک) نشان دهنده این است که این اکانت علاقمند به موسیقی پاپ

آسیایی (معمولاً کره‌ای یا حتی ژاپنی، چینی و ...) است یا نه. مقدار یک به این معنی است که نام انگلیسی این

اکانت یا بیوگرافی این اکانت حاوی کلماتی مانند **otx**, **army**, **exo**, **bts** (x یک عدد است مثلاً 5) است.

اکانت‌هایی که مقدار این فیلد برای آنها صفر لحاظ شده لزوماً غیرعلاقمند به این نوع موسیقی نیستند؛ بلکه در

یوزرنیم یا بیوگرافی آن اکانت‌ها کلمات مرتبط یافت نشده است. ممکن است با بررسی عمیق توییت‌های این افراد،

شواهدی مبنی بر علاقمندی آنها به این نوع موسیقی بدست آید.

شما در این پروژه باید با بررسی و تحلیل شبکه با روشها و الگوریتمهای مختلف به سوالات زیر به صورت مبسوط پاسخ دهید:

سوال ۱: لیست ده اکانت برتر از نظر فیلدهای activity, impression, quotes, replies, retweets, tweets و reaction و followers را بیابید و محتوای توییت‌های این اکانتها را در توییت‌ر بررسی و تحلیل نمایید.

سوال ۲: لیست ده اکانت برتر از منظر معیارهای Hub, Authority, PageRank و In Degree را بیابید و تفسیر کنید اشتراکها و اختلافهای این ۴ لیست به چه دلیل است؟ احتمالا مجبور باشید به محتوای توییتها هم نگاهی بیندازید.

سوال ۳: با حذف جهت از گراف، الگوریتم k-core را روی این دیتاست اجرا نموده و k-shell نودها را محاسبه نمایید. بررسی کنید نودهایی که بالاترین k-shell را در کل گراف دارند چه ویژگیهای موضوعی یا محتوایی داشته اند.

سوال ۴: الگوریتم Louvain را برای استخراج انجمنها روی گراف اجرا نمایید. پس از اجرای الگوریتم، انجمنهای بدست آمده را با layout مناسب بازنمایی و نتیجه را تحلیل کنید (احتمالا نام اکانت‌های داخل هر انجمن بتواند دلیل تشکیل آن انجمن را برای شما مشخص نماید). فقط انجمنهایی را تحلیل کنید که تعداد نودهای آن انجمن بیش از ۲ درصد کل گراف باشد.

سوال ۵: الگوریتمهای infomap یا یکی از الگوریتمهای مبتنی بر label propagation را روی گراف اجرا کنید و انجمنهای بدست آمده را با روش louvain مقایسه کنید. فقط ۱۰ انجمن بزرگتر هر روش را با روش دیگر مقایسه کنید.

سوال ۶: با استفاده از فیلدهای robot و btsexo و ترکیب آن با اطلاعات بدست آمده از الگوریتمهای تشخیص انجمن تحلیل کنید کدام انجمنها فعالیت سازمان یافته داشته‌اند و تحقیق کنید این انجمنها متعلق به چه گروههای سیاسی هستند.

سوال ۷: اکانت‌هایی که سن کمی دارند بخصوص آنهایی که از ۲۳ شهریور به بعد ایجاد شده‌اند را با استفاده از داده‌های تحلیل انجمن بدست آمده به نحو مناسب بازنمایی کنید و تحلیل کنید این اکانتها بیشتر در کدام انجمنها قرار میگیرند.

سوال ۸: اکانت‌هایی که روزهای فعالیت (activedays) زیادی داشته‌اند متعلق به کدام انجمنها هستند؟