

# شبکه‌های اجتماعی و اقتصادی

دانشکده مهندسی کامپیوتر

دکتر مریم رضائی

ترم: بهار ۱۴۰۳-۱۴۰۴



تاریخ انتشار: ۱۶ فروردین ماه ۱۴۰۴

## تمرین دوم

ارسال بدون تاخیر: ۲۷ فروردین، ارسال با تاخیر: ۲۹ فروردین

سوالات عملی (۷۰ نمره)

در این سوال شبکه‌ای جهت دار و وزن دار در ضمیمه تمرین به شما داده شده است، این شبکه که به صورت فایل قابل پردازش در نرم افزار Gephi (فایل HW2-dataset.gdf) به شما داده شده است<sup>۱</sup>، حاوی شبکه‌ی باز نشر میان اکانت‌های ایکس (توییتر سابق) هست که در یک بازه‌ی زمانی چند روزه، حداقل یک پست با مضامینی که نشان‌دهنده حمایت از دولت اسرائیل است منتشر یا باز نشر کرده‌اند. یال‌های این شبکه همانطور که گفته شد وزن دار هستند و وزن آن‌ها نشان‌دهنده‌ی تعداد دفعاتی است که اکانت source از اکانت target باز نشر کرده است. (یعنی جهت یال از ارجاع دهنده به مرجع است) همچنین نودهای شبکه حاوی اطلاعاتی همچون: نام کاربری (username) انگلیسی، تعداد پست‌ها، ریپلای‌ها، کوت‌های منتشر شده توسط آن کاربر، تعداد فالوورها، میزان امتیاز (لایک بر پست) دریافتی توسط پست‌های کاربر، میزان اهمیت کاربر در موضوع مورد بررسی (حمایت از اسرائیل) و سن اکانت کاربر است.

**پرسش ۱ (۵ نمره)** ابتدا نرم افزار Gephi دانلود و نصب کرده و سپس ابتدا فایل ضمیمه شده را با استفاده از آن باز کرده و رتوس و یال‌ها را به همراه پارامترهای مختلف مورد نیاز آماده نمایید

**پرسش ۲ (۱۰ نمره)** شبکه بارگذاری شده را از منظر معیارهای مرکزیت زیر تعیین نمایید.

(آ) معیارهای درجه، درجه ورودی و درجه خروجی

(ب) معیار نزدیکی (Closeness) یا نزدیکی هارمونیک (Harmonic Closeness)

(ج) معیار بینابینی (Betweenness)

(د) معیار بردار ویژه (Eigenvector)

(ه) معیار HIT

(و) معیار رتبه صفحه (Page Rank)

**پرسش ۳ (۱۰ نمره)** برای هر معیار، با توجه به حساب‌های کاربری استخراج شده در سوال قبل به صورت کامل و دقیق با استفاده از اطلاعاتی که از بررسی ساختار شبکه و نیز از فعالیت نودها در شبکه اجتماعی ایکس به دست می‌آوردید تحلیل کنید چرا این نودها برای هر معیار بالاترین میزان مرکزیت را کسب کرده‌اند و این مقدار بدست آمده نشان‌دهنده چیست؟

**پرسش ۴ (۱۰ نمره)** در مجموعه داده تهیه شده، ستون اهمیت نودها (importance) سعی دارد، اهمیت محتوایی نودها و تأثیر آن محتواها را محاسبه کند:

(آ) نودهایی که از منظر این شاخص حائز عدد بالاتری هستند را تحلیل کنید.

(ب) حدس می‌زنید این شاخص چگونه بدست آمده است؟ حدس خود را از طریق بررسی میزان شاخص همبستگی (correlation) بین ستون‌های اطلاعاتی مختلف (نظیر سن، تعداد فالوور و ...) مجموعه داده و همچنین شاخص‌های ساختاری (تعداد همسایه‌ها معیارهای مختلف مرکزیت و ...) با میزان importance به طور کامل بررسی نمایید.

(ج) اگر شما می‌خواستید اهمیت نودها را از منظر تأثیرگذاری در موضوع طرفداری از دولت اسرائیل بدست بیاورید چه شاخص ساختاری (ساختار در گراف مورد بررسی) پیشنهاد می‌کردید؟ آن را با استفاده از پایتون پیاده‌سازی و سپس ترتیب نودها را از منظر این شاخص مشخص و نتیجه را تحلیل کنید.

**پرسش ۵ (۳۵ نمره)** در ادامه با توجه به گراف مورد بررسی می‌خواهیم خوشه‌بندی (Graph Clustering) بر روی رتوس گراف انجام داده و به تحلیل نودهای پر اهمیت در هر دسته بپردازیم:

(آ) ابتدا ماتریس لاپلاسیان (Laplacian Matrix) شبکه را با توجه به وزن دار و جهت دار بودن گراف بسازید. (مشخص کنید کدام نوع لاپلاسیان (نرمالیزه شده یا Random-walk) را برای این کار انتخاب کرده‌اید) سپس با کمک کتابخانه‌هایی نظیر numpy.linalg، مقادیر ویژه و بردارهای ویژه‌ی ماتریس لاپلاسیان را محاسبه کنید.

(ب) با بررسی شکاف بین مقادیر ویژه (Eigen Gap) تعداد خوشه‌های مناسب را تخمین بزنید و نمودار مقادیر ویژه (از کوچک به بزرگ) را رسم کرده و سپس موقعیت بزرگ‌ترین پرش (Gap) را مشخص نمایید.

(ج) در ادامه با استفاده از K-Means بردارهای ویژه‌ی متناظر با کوچکترین k مقدار ویژه (با توجه به تعداد خوشه) را خوشه‌بندی کنید. (هر نود را به خوشه‌ی متناظر اختصاص دهید و برچسب خوشه‌ی به دست آمده را به عنوان یک ویژگی جدید در داده‌ی گراف ثبت کنید. سپس این ویژگی را به فرمت gdf (یا CSV) خروجی گرفته و در گام بعد وارد Gephi کنید.

<sup>۱</sup> این مجموع داده با کمک و همراهی جناب آقای دکتر مسعود اسدپور و از پلتفرم هشتگ (tag 8) خروجی گرفته شده است و مجوز استفاده از آن تنها برای این تمرین داده شده است، لطفاً از نشر و یا استفاده از این دیتاست در موارد دیگر بپرهیزید.

- (د) داده‌ی به‌روزشده را در Gephi بارگذاری کرده و گراف را براساس خوشه‌هایی که در مراحل قبل بدست آوردید، رنگ‌دهی کنید. سپس اندازه‌ی نودها را براساس میزان اهمیت کاربر که در دیتاست نمایش داده شده و همچنین در گام بعدی بر اساس معیارهای مرکزیت که بدست آوردید تنظیم کنید تا نودهای اثرگذار هر خوشه و الگو و رفتارهای احتمالی درون خوشه‌ها نمایان شود.
- (ه) براساس نتایج فوق و اصلی‌ترین نودهای هر خوشه (نودهای دارای میزان مرکزیت یا importance بالا)، تفسیر کنید که هر خوشه چه ویژگی‌های محتوایی یا رفتاری را داراست (مثلا ریات‌ها، حامیان اسرائیل، مخالفان اسرائیل، خیرگزاری‌ها و ...)
- (و) خوشه‌بندی را اینبار براساس الگوریتم‌های آماده در Gephi نظیر Modularity انجام دهید، آیا تعداد خوشه‌های تعیین شده و همچنین رفتار کلی خوشه‌های ایجاد شده با خوشه‌بندی صورت گرفته توسط شما یکسان است و چه تفاوت‌های دارد؟
- (ز) (امتیازی) فرض کنید بخواهیم نوعی خوشه‌بندی طیفی پویا در گراف زمانی (Temporal Graph) فعالیت در ایکس، انجام دهیم که تغییرات خوشه‌ها را در طول زمان نشان دهد. ایده‌ی عملی یا روش تحقیقاتی برای تحقق این موضوع را پیشنهاد دهید و بررسی کنید.

### اقدام قابل تحویل:

۱. فایل گزارش مبسوط تمرین با قالب گرافیکی مناسب (معیار قضاوت در مورد تمرین همین گزارش است)
۲. نوت بوک (Notebook) کلیه قسمت‌های پیاده‌سازی شده با پایتون
۳. یک فایل CSV حاوی یوزرنیم نودها و عدد مرکزیت نود براساس شاخص طراحی شده شما، مرتب شده به صورت نزولی براساس مقدار شاخص (نودهای مهم‌تر در ردیف‌های بالاتر)

ارسال بدون تاخیر: ۲۷ فروردین، ارسال با تاخیر: ۲۹ فروردین

## سوالات تئوری (۳۰ نمره)

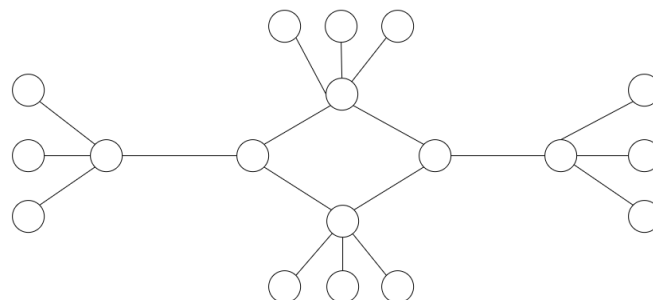
### پرسش ۱ (۱۵ نمره)

یکی از ویژگی‌های مهم در انواع گراف‌ها مخصوصاً گراف‌های مرتبط با شبکه‌های اجتماعی پویا بودن و به اصطلاح Dynamic بودن آن‌ها است. بدین معنا که ویژگی‌های مختلف گراف نظیر تعداد و درجه رئوس، ارتباطات و وزن یال‌ها و سایر ویژگی‌های ساختاری گراف در حالت‌های مختلف در حال تغییر هستند. این تغییرات گاهی در طول بازه‌های زمانی مختلف رخ می‌دهند و گراف در سری زمانی‌های مختلف در حال تغییر است که به آنها گراف‌های سری زمانی یا Temporal Graph می‌گوییم. حال گرافی را در نظر بگیرید که در طول بازه‌های زمانی مختلف در حال تغییر است، برای سادگی در این سوال فرض می‌کنیم که این تغییرات در یال‌های گراف ایجاد می‌شوند و رئوس گراف در طول زمان تغییری نمی‌کنند؛ همچنین یال‌ها بدون جهت هستند. به این ترتیب مرکزیت راس‌ها ثابت نبوده و وابسته به زمان است. حال قصد داریم مقایسه‌ای بین پنج تا از معیارهای مختلف مرکزیت (degree, closeness, betweenness, PageRank) و Eigenvector) در این گراف‌ها انجام دهیم. با توجه به این موضوع به سوالاتی که در ادامه مطرح شده اند پاسخ دهید:

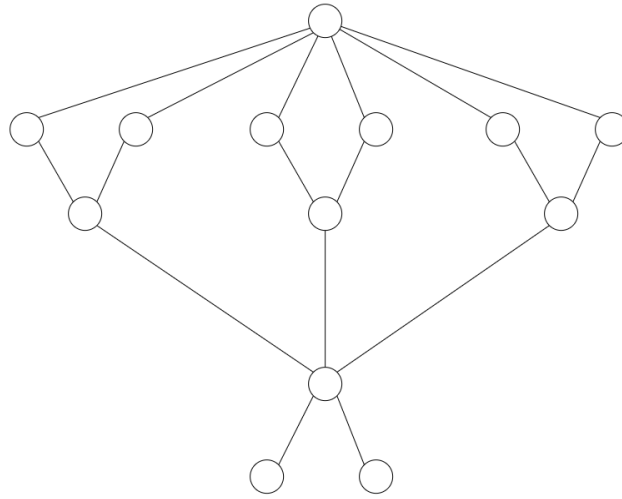
- سه مثال و کاربرد از گراف‌های سری زمان (Temporal Graph) به طور کامل توضیح دهید.
- فرض کنید تعداد یال‌های اضافه (یا حذف) شده از گراف در زمان  $t+1$  نسبت به زمان  $t$  از مرتبه  $O(1)$  باشد. کدام یک از معیارهای مرکزیت تغییر چندانی نخواهد داشت؟
- با همان فرض قسمت قبل آیا امکان دارد که میزان مرکزیت یک راس تغییر چشمگیری داشته باشد؟ اگر بله مثال بزنید اگر نه اثبات کنید.
- در نظر گرفتن هر کدام از معیارهای مرکزیت فوق در گراف‌های پویا چه مزیت‌ها و معایبی دارد؟ به طور کامل شرح داده و مثال بزنید

پرسش ۲ (۱۰ نمره) فرض کنید نقشه یک شهر را با گرافی بدون جهت نشان می‌دهیم که در آن هر رأس نشان دهنده‌ی یک محله در شهر است و یال‌ها نیز نشان دهنده‌ی ارتباط بین محله‌ها که همگی وزن یکسانی نیز دارند:

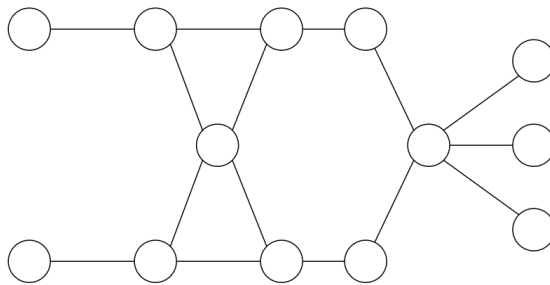
(آ) فرض کنید در یکی از محله‌های این شهر می‌خواهیم بیمارستانی بنا کنیم و نکته‌ی مهم برای محل بیمارستان نزدیک بودن آن به کلیه نقاط شهر است به طوری که از هر نقطه‌ای بتوانیم در کمترین زمان (کمترین فاصله) به بیمارستان برسیم؛ با توجه به گراف زیر کدام رأس بهترین انتخاب برای محل بیمارستان است و براساس کدام معیار در گراف این محل را انتخاب کردید؟  
توجه: مجموع فاصله از محله‌ها مهم نیست، مهم دسترسی پذیری از هر محله است.



(ب) این بار شما صاحب یک مجتمع تجاری هستید که می‌خواهید بیشترین تعداد مشتریان را داشته باشید، برای جذب مشتریان بیشتر مهم‌ترین عامل نزدیکی به مشتریان است، با فرض اینکه در گراف زیر تعداد افراد در هر محله برابر باشد کدام محله مناسب‌ترین مکان برای راه‌اندازی مجتمع تجاری است؟ بر اساس کدام معیار در گراف این رأس را انتخاب کردید؟



(ج) اگر در این گراف بخواهید نقطه‌ای را برای راه‌اندازی مجتمع تجاری انتخاب کنید، آن محله کدام رأس خواهد بود؟



(د) فرض کنید در گراف بالا، این بار رقیبی دیگر می‌خواهد مجتمع تجاری‌ای راه‌اندازی کند که اگر به محله‌ای نزدیک‌تر باشد مشتریان در آن محله جذب آن مجتمع تجاری خواهند شد حالا برای هر رأس مشخص کنید در صورت راه‌اندازی مجتمع تجاری رقیب در آن رأس چه تعداد مشتری کمتر یا بیشتری نسبت به محل مجتمع تجاری بخش «ج» خواهیم داشت و بهترین رأس برای مجتمع تجاری رقیب کدام خواهد بود؟

(ه) کدام رأس برای رقیب سوم بهتر است؟

با توجه به نتیجه این سه بخش توضیح دهید آیا در این گراف اولین کسی که مجتمع تجاری را راه‌اندازی می‌کند (first mover) سود کرده است یا ضرر؟ و همچنین با توجه به آن دلیل به وجود آمدن نقاط متمرکز با تعداد بالای از یک صنف (بورس کالاهای مختلف) را توضیح دهید.

پرسش ۳ (۵ نمره) آیا ممکن است در گرافی Median، Jordan Center، و Centroid آن گراف در رئوس متفاوتی باشند؟ مثال بزنید و توضیح دهید.