



## گراف کاوی - تمرین سری دوم

موعد تحویل ۷ دی ۱۴۰۰

پیش از حل سوالات به موارد زیر دقت کنید:

- پاسخ تکلیف را به صورت یک فایل PDF آماده کنید و با نام HW2\_NAME\_STDNUM.pdf در سامانه آپلود کنید. (به جای NAME، فقط نام خانوادگی و به جای STDNUM، شماره دانشجویی قرار بگیرد و حتما رعایت شود!)
- در تحویل تکالیف به زمان مجاز تعیین شده دقت نمایید. موعد تکالیف قابل تمدید نمی باشند.
- در صورتی که مجموع تاخیر کل تکالیف شما کمتر از ۲۴ ساعت باشد نمره‌ای از شما کسر نمی گردد. در غیر این صورت به ازای هر روز تاخیر درصدی از نمره تکالیف شما کسر می گردد.
- پاسخ تکالیف را حتما در سامانه آپلود کنید و از ارسال تکالیف به ایمیل یا تلگرام اکیدا خودداری نمایید.
- در صورت وجود شباهت واضح، نمره‌ای به سوال تعلق نمی گیرد.
- در صورت وجود هرگونه ابهام می توانید در گروه تلگرام یا گروه اسکایپ سوالات خود را مطرح کنید.
- از طریق ایمیل های زیر می توانید با TA های مربوط به این تکلیف در ارتباط باشید.

– nazerimahdi2001@gmail.com

– arezo.h1371@yahoo.com

سوال ۱. یکی از روش‌هایی که به وسیله آن می‌توانیم میزان همسایه‌بودن راس‌ها (Node Neighborhood) را در یک گراف بیابیم، استفاده از Katz Index است. Katz Index میان دو راس  $u$  و  $v$  به صورت  $S_{uv} = \sum_{l=1}^{\infty} \beta^l A_{uv}^l$  تعریف می‌شود. در این تعریف  $\beta$  یک ضریب حقیقی است و discount factor نام دارد.

الف. بزرگی این معیار بیانگر چیست؟ چرا؟ فرم ماتریسی katz index چگونه است؟

ب. دلیل استفاده از discount factor چیست؟

ج. ماتریس  $A$  را در نظر بگیرید. مقادیر ویژه این ماتریس باید چگونه باشند تا  $\lim_{i \rightarrow \infty} A^i$  تماماً صفر باشد؟ ادعای خود را اثبات کنید.

د. می‌دانیم تحت شرایطی، رابطه  $\sum_{i=1}^{\infty} A^i = (I - A)^{-1}$  برقرار است. این رابطه را اثبات کنید و شرایط لازم را بیابید.

د. با توجه به قسمت‌های قبلی، نحوه انتخاب discount factor مناسب را توضیح دهید.

سوال ۲. یکی از روش‌هایی که به وسیله آن می‌توانیم ویژگی‌های یک گراف را شرح دهیم، استفاده از گرافلت‌ها است.

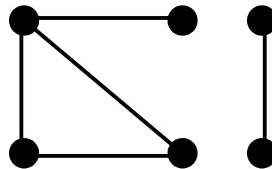
الف. گرافلت‌ها چیستند؟ ویژگی‌های آنان را شرح دهید.

ب. چند گرافلت با اندازه ۳ داریم؟ آن‌ها را رسم کنید. چند گرافلت با اندازه ۴ داریم؟

ج. آیا گرافلت‌های مورد استفاده برای راس‌ها با گرافلت‌های مورد استفاده برای گراف‌ها متفاوت هستند؟ تفاوت آنان را شرح دهید.

د. چگونه می‌توانیم با استفاده از گرافلت‌ها یک بردار ویژگی بسازیم؟

ه. گراف زیر را در نظر بگیرید. با استفاده از گرافلت‌ها، بردار ویژگی گراف و راس‌های آن را بیابید.



سوال ۳. ماتریس احتمال، یک Stochastic Matrix است، یعنی تمامی مقادیر آن نامنفی و جمع هر ستون آن برابر یک است.

الف. نشان دهید  $\lambda_i = 1$  یک مقدار ویژه برای هر ماتریس احتمال است.

ب. نشان دهید تمامی مقادیر ویژه هر ماتریس احتمال، کوچکتر یا مساوی یک هستند. ( $\lambda_i \leq 1$ )

ج. مقادیر و بردارهای ویژه ماتریس احتمال زیر را بدست آورید. بردارهای ویژه را نرمال کنید.

د. الگوریتم PageRank را بر روی ماتریس احتمال زیر اجرا کنید. برای این منظور می‌توانید از یک حالت اولیه دلخواه

مثبت شروع کنید و به کمک کامپیوتر یا ماشین حساب، حدود ۲۰ مرحله از الگوریتم را محاسبه کنید. دقت کنید که نیازی به استفاده از انتقال‌های تصادفی نیست.

ه. حالت پایدار حاصل شده در قسمت قبل، برابر با کدام یک از بردارهای ویژه است؟

$$A = \begin{bmatrix} 0.3 & 0.4 & 0.5 \\ 0.3 & 0.4 & 0.3 \\ 0.4 & 0.2 & 0.2 \end{bmatrix}$$

سوال ۴. فایل ضمیمه شده با نام 'soc-edges.txt' یال هایی از شبکه ی فیسبوک را شامل می شود. این شبکه لحظه ای را نشان می دهد که اعضای جدیدی به شبکه اضافه شدند و با اعضای دارای بیشترین ارتباطات (همسایه ها) در شبکه ارتباط برقرار کرده اند. هدف مشخص کردن اعضای جدید با استفاده از خصوصیات ساختاری رئوس شبکه است. در فایل 'soc-nodes.txt' برچسب رئوس برای اعضای جدید با یک و برای بقیه ی اعضا با صفر در ستون 'label' مشخص شده است.

راهنمایی: کد ضمیمه شده در فایل ML.py شامل یک مدل رده بندی است که برای پیشبینی و تشخیص رئوس مربوط به اعضای جدید از بقیه ی اعضا با استفاده از ویژگی (های) ورودی ایجاد شده است. برای آموزش این مدل ویژگی یا ویژگی هایی از رئوس شبکه را که می تواند در رده بندی رئوس شبکه تاثیرگذار باشد، بیابید. مقادیر این ویژگی ها را برای هر راس در یک data frame ذخیره کنید و کد را تکمیل کنید. دقت نهایی مدل با استفاده از ویژگی های منتخب باید برابر و یا بیشتر از 85% باشد.

سوال ۵. کدی بنویسید که معیار PageRank را برای رئوس شبکه با استفاده از power iterations پیاده سازی کند. با استفاده از این کد معیار PageRank را برای رئوس شبکه ی wiki-vote محاسبه کنید. در این شبکه هر راس متناظر با یک کاربر ویکی پدیا و هر یال مشخص کننده ی یک رای (اعتبار) از راس مبدا به راس مقصد است. فایل 'soc-wiki-vote.txt' شامل یال های این شبکه می شود. نوع یال ها در این شبکه جهت دار است و از راس اول به راس دوم وارد می شود. با تغییر پارامتر  $\beta$  در محدوده ی 0.1 تا 0.9 هر بار با اضافه کردن 0.1 به مقدار قبلی، تعداد تکرارهای لازم برای همگرایی و انحراف استاندارد مقادیر نهایی PageRank را مقایسه کنید.

راهنمایی: در این پیاده سازی اگر  $r_i^t$  رتبه ی راس  $i$  در مرحله ی  $t$ ،  $d_i$  تعداد یال های خروجی از راس  $i$ ،  $N$  تعداد کل رئوس شبکه و  $\beta$  پارامتر damping و یا احتمال دنبال کردن یال های مجاور به راس فعلی در هر مرحله باشد، مقادیر رتبه برای رئوس شبکه در هر مرحله با استفاده از فرمول زیر بروزرسانی می شود.

$$r_j^{t+1} = \sum_{i \rightarrow j} \beta \frac{r_i^t}{d_i} + (1 - \beta) \frac{1}{N} \quad (1)$$

در این پیاده سازی معیار همگرایی به این صورت تعریف می شود که میانگین اختلاف رتبه ی رئوس در تکرار فعلی و قبلی کمتر از  $tol = 10^{-7}$  باشد.