

سوالات میانترم

صدف فتح الهي

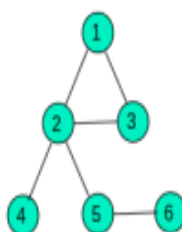
۸ اردیبهشت ۱۴۰۲



نام استاد: خانم دکتر طاهری
نام درس: گراف کاوی

۱ سوال اول

گراف زیر را به عنوان یک شبکه اجتماعی از کاربران و تعاملات آنها در نظر بگیرید



الف) معیارهای زیر را برای هر کاربر در این شبکه اجتماعی محاسبه کنید

degree
closeness centrality
betweenness centrality

کدام کاربران دارای بیشترین و کمترین مقادیر هر یک از این معیارها هستند؟ هر یک از مقادیر فوق در به‌کارگیری این معیارها در مهندسی ویژگی برای الگوریتم‌های یادگیری ماشین که رفتار شبکه اجتماعی یا تعامل کاربر را پیش‌بینی می‌کنند چگونه تفسیر می‌شوند؟ (کاربران را از این نظر با یکدیگر مقایسه کنید)

ب) معیارهای زیر را برای هر کاربر در شبکه اجتماعی محاسبه کنید

clustering coefficient
Graphlets of size 3 vector

هر یک از مقادیر فوق چگونه تفسیر می‌شوند؟ کاربران را از این نظر با یکدیگر مقایسه کنید

ج) فرض کنید می‌خواهید با هدف تبلیغات هدفمند، تأثیرگذارترین کاربر در شبکه اجتماعی را شناسایی کنید. کدام کاربر بیشترین تأثیر را بر معیارهای تعامل شبکه خواهد داشت و چرا؟ (بر اساس معیارهای مطرح شده در درس پاسخ دهید)

د) تعاملات (بال‌های) موجود در شبکه اجتماعی مورد اشاره را از نظر اهمیت با استفاده از حداقل ۳ شاخص با یکدیگر مقایسه کنید. کدام تعامل (تعاملات) از بقیه با اهمیت‌تر هستند؟ (در مورد تفسیر شاخص‌ها به‌طور دقیق توضیح دهید)

۱.۱ پاسخ الف

node	degree	clossness centrality	betweenness centrality
1	2	0.142857143	0
2	4	0.166666667	8
3	2	0.111111111	0
4	1	0.1	0
5	2	0.125	4
6	1	0.083333333	0

۱. degree :

- node 2 : Max
- node 6, node 4 : Min
- این معیار تعداد همسایه های یک راس را بدون در نظر گرفتن اهمیت آنها می‌شمارد.

۲. closeness centrality :

- node 2 : Max
- node 6 : Min
- این معیار مشخص می‌کند که یک راس چقدر به مرکز شبکه نزدیک است. به عبارتی یک راس مهم است اگر به همه ی راس ها نزدیک باشد و مسیر کوتاهی تا بقیه رئوس داشته باشد.

۳. betweenness centrality :

- node 2 : Max
- node 6, node4, node 3, node 1 : Min
- این معیار مشخص می‌کند که یک راس به عنوان اتصال دهنده چقدر مهم است. به عبارتی یک راس مهم است اگر به دفعات زیاد در کوتاه ترین مسیر سایر رئوس قرار گرفته شده باشد.

با توجه به تمام این معیارها میتوان گفت راس 2 مهمترین راس در شبکه است و راس 6 از همه کم اهمیت تر است.

۲.۱ پاسخ ب

node	clustering coefficient	graphlet of size 3
1	1	3
2	0.16666667	7
3	1	4
4	0	3
5	0	4
6	0	1

۱. clustering coefficient :

- node 3, node 1 : Max
- node 6, node 5, node 4 : Min
- این معیار تمایل راس های گراف برای تشکیل خوشه را مشخص میکند.

۲. Graphlets of size 3 vector :

- node 2 : Max
- node 6 : Min
- معیاری برای سنجش یک راس از لحاظ موقعیت توپولوژیکی است. به عبارتی این ویژگی برای سنجش شباهت دو گراف معیار مهمتری از بقیه است.

بنابراین این دو معیار برای پیش بینی کردن نقش خاصی که یک راس در گراف دارد مفید هستند. مثلا تعامل پروتئین ها طبق نتایج راس 1 و راس 3 تمایل بیشتری برای ایجاد خوشه با دیگر راس ها دارند.

۳.۱ پاسخ ج

راس 2. زیرا تقریبا در تمام معیارها مقدار بالایی داشت.

۴.۱ پاسخ د

۱. distance based feature :

• Shortest path distance between two nodes

این روش اهمیت راس هایی که درجه بالا دارند را در نظر نمیگیرد.

۲. local neighborhood overlap:

مشکل این مدل این است که اگر دو راس همسایه مشترک نداشته باشند صفر برمیگرداند در حالی که ممکن است آن دو راس در آینده بهم وصل شوند.

• common neighbor مشکل این روش این است که راس هایی که درجه بالایی دارند احتمال بیشتری هست که بهم وصل باشند.

• Jaccard's coefficient نرمالایز شده روش بالا.

• Adamic Adar index اگر همسایگان مشترک زیادی دارید که دارای درجه پایینی هستند، بهتر از این است که افراد مشهور بسیار مرتبط زیادی به عنوان مجموعه ای از همسایگان مشترک داشته باشید. بنابراین این یک ویژگی است که در یک شبکه اجتماعی واقعاً خوب کار می کند.

۳. Global neighborhood overlap

این روش علاوه بر اینکه مشکل روش قبل را ندارد بلکه تعداد قدم هایی از هر طولی را بین دو راس می شمارد.

• Katz index

edge	Distance based feature	Common neighbors	Jaccard's coefficient	Adamic Adar index
[1,2]	1	1	0.2	3.321928095
[1,3]	1	1	0.333333333	1.660964047
[1,4]	2	1	0.5	1.660964047
[1,5]	2	1	0.333333333	1.660964047
[1,6]	3	0	0	0
[2,3]	1	1	0.2	3.321928095
[2,4]	1	0	0	0
[2,5]	1	0	0	0
[2,6]	2	1	0.25	3.321928095
[3,4]	2	1	0.5	1.660964047
[3,5]	2	1	0.333333333	1.660964047
[3,6]	3	0	0	0
[4,5]	2	1	0.5	1.660964047
[4,6]	3	0	0	0
[5,6]	1	0	0	0

بنابراین قدم بعدی میتواند راس های ۳ یا ۴ یا ۵ باشد.

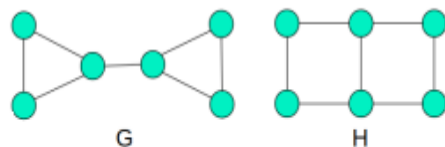
۳.۲ ج

دو الگوریتم node2vec و deepwalk را از نظر نقاط قوت و ضعف با هم مقایسه کنید. کدام الگوریتم را پیشنهاد میکنید.

در node2vec در هر مرحله میدانیم که در مرحله قبل در کدام راس بوده ایم ولی در deepwalk اینگونه نیست. از طرفی پیچیدگی زمانی node2vec خطی است و تمام مراجل ان میتواند به صورت موازی انجام شود. ولی نقطه ضعف ان این است که برای نتورک های بزرگ مناسب نیست چون باید برای تک تک راس ها embedding بدست آورد. با توجه به نوع کاربرد میتوان از یکی از این دو روش استفاده کرد ولی node2vec برای کلسیفیکشن در راس ها مناسب تر است.

۳ سوال سوم

مسئله یکرختی گراف ها یک مسئله بسیار دشوار در ریاضیات است. الگوریتم وایسفالر-لیمن به عنوان یکی از قدرتمندترین روش ها برای تشخیص عدم یکرختی دو گراف شناخته می شود. دو گراف غیریکرخت زیر را در نظر بگیرید و به سوالات پاسخ دهید



الف) آیا الگوریتم وایسفالر-لیمن قادر به تشخیص عدم یکرختی دو گراف بالا هست؟ (جوابی پاسخ خود را به طور دقیق با شکل توضیح دهید)

ب) فرض کنید که همه راس ها بردارهای ویژگی اولیه یکسانی دارند و ما از یک نسخه ساده شده از شبکه عصبی گرافی استفاده می کنیم (بدون تابع فعال سازی و بدون شبکه عصبی). همچنین رابطه زیر برای محاسبه فیچر هر راس در هر لایه از شبکه عصبی گرافی در نظر گرفته می شود

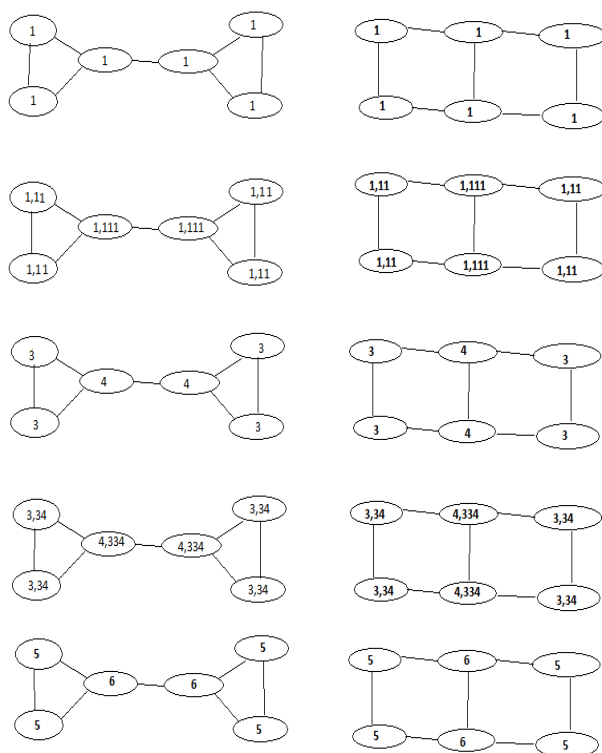
$$h_v^l = \text{Mean} \left(\left\{ h_u^{l-1} \right\}_{u \in N(v)} \right) + h_v^{l-1}$$

آیا شبکه عصبی گرافی با تعریف بالا قادر به تشخیص عدم یکرختی دو گراف بالا هست؟ (جوابی پاسخ خود را به طور دقیق با شکل توضیح دهید)

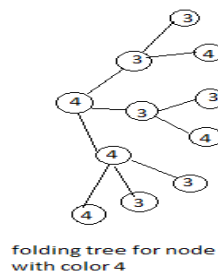
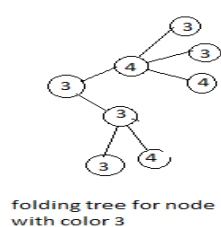
ج) با به کارگیری یکی از روش های مطرح شده در درس (روشی غیر از دو روش مطرح شده در قسمت های الف و ب و همچنین بدون استفاده از روش های معمول ریاضیاتی)، نشان دهید که دو گراف بالا یکرخت نیستند

۱.۳ الف

در ابتدا به همه ی راس ها عدد 1 را نسبت میدهیم. سپس رنگ راس های همسایه را به هر راسی نسبت میدهیم و رنگ بندی را مرتب میکنیم.



مشاهده میشود که بعد از یک مرحله هر دو گراف فیچر راسی یکسانی دارند و اگر باز هم ادامه دهیم تفاوتی ایجاد نمیشود و این یعنی این الگوریتم در تشخیص ایزومورف بودن نا توان است. زیرا اگر folding tree برای راس با رنگ 3 و 4 رسم کنیم. در هر دو گراف folding tree یکسانی به دست میآید.

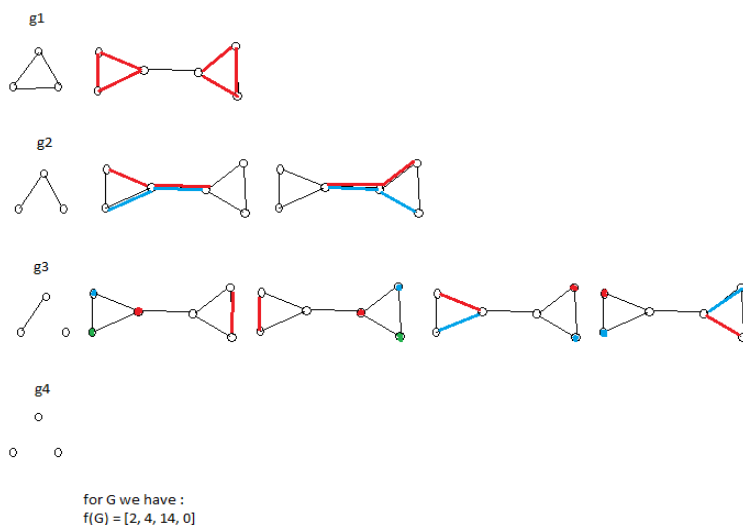


۲.۳ ب

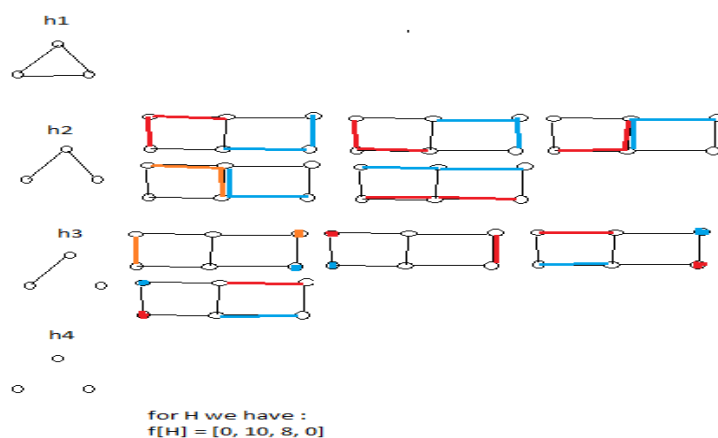
خیر زیرا همانطور که در قسمت الف مشخص شد هر دو گراف computation graph یکسانی دارند از طرفی فیچر به دست آمده برای رئوس در دو گراف دو به دو یکسان هستند یعنی راس با رنگ ۳ در هر دو گراف با توجه به فرمول دارای فیچر ۳ است. به همین ترتیب برای راس با رنگ ۴ نیز فیچر برابر ۳ است

۳.۳ ج

برای گراف G داریم :



برای گراف H داریم :



برای هر دو گراف Graphlet با سایز ۳ را رسم کردیم و فیچر گرافی را به دست آوردیم همانطور که مشخص است این دو فیچر یکی نیستند پس دو گراف ایزومورف نیستند.

۴ سوال امتیازی

شبکه های عصبی گرافی پیچشی و دو روش بندهای الف و ج را از نظر نقاط قوت و ضعف با هم مقایسه کنید:

۱. Graphlet kernel:

- بر اساس bag of word که در NLP هست ساخته شده که در اینجا به جای word از graphlet استفاده میکنند.
- از لحاظ محاسباتی پیچیده است زیرا شمردن تعداد graphlet ها سخت و وقتگیر است.
- با اضافه شدن یک راس به گراف باید از اول تمام مراحل را انجام داد.

۲. WL kernel:

- بر اساس bag of word که در NLP هست ساخته شده که در اینجا به جای word از color استفاده میکنند.
- از لحاظ محاسباتی به صرفه است.
- با اضافه شدن یک راس به گراف باید از اول تمام مراحل را انجام داد.

۳. GCN:

- پیچیدگی محاسباتی آن خطی است زیرا تعداد پارامترهایی که باید اپدیت شوند از مرتبه تعداد رئوس است
- با اضافه شدن یک راس به گراف لازم نیست از اول تمام مراحل را انجام داد.