

# شبکه های اجتماعی و اقتصادی

دانشکده مهندسی کامپیوتر

مریم رضانی  
بهار ۱۴۰۴



تاریخ انتشار: ۱ خرداد ۱۴۰۴

## تمرین چهارم

سیستم های توصیه گر، گراف، شبکه

۱. سوالات خود در مورد این تمرین را در کوئرا مطرح کنید.

۲. لطفا پاسخ خود را با توضیحات کافی ارائه دهید.

تاریخ تحویل:

## سوالات تئوری (۸۰ نمره)

پرسش ۱ (۲۰ نمره)

سوال: تحلیل گراف های ناهمگن و کاربرد Meta-Path در خوشه بندی مدل سازی گراف های ناهمگن:

گرافی ناهمگن تعریف کنید که شامل نودهای متفاوت مانند مقاله ها، ژورنال ها و نویسندگان باشد. همچنین گراف دیگری شامل فیلم ها، کارگردان ها و بازیگران ایجاد نمایید. در هر دو گراف، نوع نودها و نوع یال ها (روابط میان نودها) را مشخص کنید.

تعریف Meta-Path:

مفهوم Meta-Path را در گراف های ناهمگن توضیح دهید. چرا این مفهوم برای تحلیل رابطه های غیر مستقیم بین نودها مهم است؟

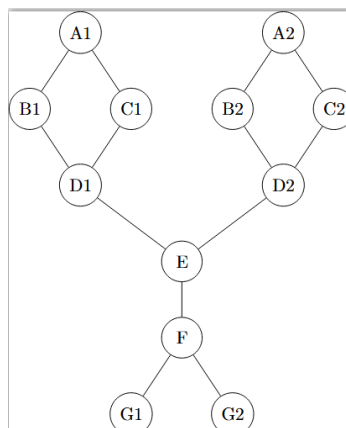
کاربرد Meta-Path در مدل سازی گراف:

نحوه استفاده از Meta-Path ها برای مدل سازی روابط میان نودهای هم نوع (مثلاً نویسنده با نویسنده یا فیلم با فیلم) را توضیح دهید. چگونه می توان از این مدل سازی برای اعمال الگوریتم های Community Detection استفاده کرد؟

طراحی Meta-Path برای گراف فیلم ها: حداقل سه Meta-Path مختلف پیشنهاد دهید که برای خوشه بندی (Clustering) فیلم ها مفید باشند. برای هر Meta-Path توضیح دهید چه نوع ارتباطی میان فیلم ها برقرار می شود و چرا برای کشف جامعه های معنایی مناسب است.

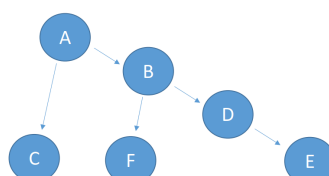
تعریف Meta-Path برای گراف مقالات جهت خوشه بندی نویسندگان: چگونه می توان از Meta-Path ها برای تعریف شباهت میان نویسندگان استفاده کرد؟ چند نمونه Meta-Path ارائه دهید که خوشه بندی نویسندگان بر اساس آن ها معنادار باشد (مثلاً همکاری مشترک، انتشار در ژورنال مشابه و ...).

پرسش ۲ (۲۰ نمره)



اجرای الگوریتم های Node2Vec، GraphSAGE و GAT را برای گراف داده شده به صورت گام به گام تا سه مرحله توضیح دهید. بر اساس پیچیدگی روابط بین رئوس، اندازه گراف و زمانی بودن توضیح دهید در هر حالت کدام الگوریتم مناسب تر است و چرا.

پرسش ۳ (۲۰ نمره)



احتمال	یال
۰,۹	$A \rightarrow B$
۰,۴	$A \rightarrow C$
۰,۶	$B \rightarrow D$
۰,۳	$B \rightarrow F$
۰,۵	$D \rightarrow E$

### مرکزیت در فضای تعبیه شده (Embedding Space) در برابر مرکزیت ساختاری

هر گره در یک فضای دوبعدی نهفته تعبیه شده که از طریق راهپیمایی‌های بایاس دار (Node2Vec) به دست آمده است. مختصات گره‌ها به صورت زیر است:

گره	(x, y)
A	(9.0, 1.0)
B	(8.0, 2.0)
C	(7.0, 0.0)
D	(6.0, 3.0)
E	(5.0, 4.0)
F	(7.5, 2.5)

(آ) فاصله اقلیدسی گره F تا سایر گره‌ها را محاسبه کرده و میانگین فاصله‌ها را گزارش دهید.

(ب) آیا مرکزیت در فضای تعبیه شده لزوماً با پتانسیل تأثیرگذاری در گراف ساختاری همبستگی دارد؟ با ذکر مثال مشخص از این گراف توضیح دهید.

### تحلیل انتشار تأثیر (مدل IC)

(آ) با استفاده از مدل انتشار مستقل (IC) تعداد مورد انتظار گره‌های فعال شده با شروع از گره A را محاسبه کنید. فرض کنید فرایند انتشار در حداکثر سه مرحله انجام می‌شود. فرمول  $\text{Expected}(X \rightarrow Y) = P(X \text{ activates } Y) \times P(X \text{ is active})$  همچنین موارد زیر را نیز محاسبه و تحلیل کنید:

• احتمال فعال شدن گره E

• تفاوت در انتظار تأثیر زمانی که گره B به جای A به عنوان seed انتخاب می‌شود

(ب) توضیح دهید که چرا از الگوریتم‌های حریصانه در بیشینه‌سازی تأثیر استفاده می‌شود و به صورت خلاصه نشان دهید که تابع تأثیر زیرمجموعه‌گرا (submodular) است.

### دینامیک - GRL هم‌رخدادی متنی در Node2Vec

فرض کنید تعبیه‌های فوق با استفاده از الگوریتم Node2Vec به دست آمده‌اند با تنظیمات زیر: طول راهپیمایی: ۳، دو راهپیمایی برای هر گره، پارامتر بازگشت  $p = 1$ ، پارامتر in-out برابر  $q = 0.25$  (تمایل به اکتشاف)

(آ) دو راهپیمایی نمونه از گره A را شبیه‌سازی کنید. وزن انتخاب‌ها و مسیرها را مشخص کنید. مثال:  $A \rightarrow B \rightarrow F \rightarrow B$

(ب) بر اساس راهپیمایی‌ها، ماتریس هم‌رخدادی بین گره‌ها را بسازید. این ماتریس چه تأثیری بر شباهت نقطه‌ای بین A و سایر گره‌ها در آموزش skip-gram دارد؟

(ج) توضیح دهید که چگونه تغییر پارامتر  $q$  در گراف‌های پراکنده و متراکم تعبیه‌ها را تحت تأثیر قرار می‌دهد. این مسئله چه ارتباطی با سیستم‌های توصیه‌گر حساس به تأثیر دارد؟

### ملاحظات راهبردی و نمایش چندوجهی

(آ) محدودیت‌های استفاده از تعبیه‌های ایستا (مانند Node2Vec یا DeepWalk) در فرآیندهای پویا مثل انتشار تأثیر را توضیح دهید. یک روش جایگزین پیشنهاد دهید (مانند GRL زمانی یا مدل‌های مبتنی بر توجه).

(ب) در گراف‌های چندوجهی که تأثیر تنها ساختاری نیست (مثلاً وابسته به محتوا یا زمان یا اعتماد اجتماعی است)، چگونه می‌توان معماری‌های GRL را گسترش داد؟ یک اسکچ از نحوه انتقال پیام در GNN را که شامل احتمال تأثیر است، ارائه دهید.

به تمامی پرسش‌ها با استدلال دقیق پاسخ دهید. در صورت نیاز به مفاهیم نظری مانند زیرمجموعه‌گرایی، توزیع پایدار راهپیمایی تصادفی یا گرادیان‌های تابع هزینه skip-gram ارجاع دهید.

### پرسش ۹ (۲۰ نمره) سیستم‌های پیشنهاددهنده

یک گراف دو بخشی user-item را در نظر بگیرید که در آن هر یال بین کاربر  $U$  و آیت  $I$  نشان‌دهنده‌ی این است که کاربر  $U$  آیت  $I$  را پسندیده است. همچنین ماتریس امتیازدهی را برای این مجموعه از کاربران و آیت‌ها با  $R$  نمایش می‌دهیم، که در آن هر سطر از  $R$  مربوط به یک کاربر و هر ستون مربوط به یک آیت است. اگر کاربر  $i$  آیت  $j$  را پسندیده باشد،  $R_{i,j} = 1$ ، در غیر این صورت  $R_{i,j} = 0$ . همچنین فرض می‌کنیم  $m$  کاربر و  $n$  آیت داریم، بنابراین ماتریس  $R$  ابعادی برابر  $m \times n$  خواهد داشت.

ماتریس  $P$  با ابعاد  $m \times m$  را تعریف می‌کنیم به عنوان یک ماتریس قطری که در آن عنصر قطری  $i$ ام برابر است با درجه‌ی رأس کاربر  $i$ ، یعنی تعداد آیت‌هایی که کاربر  $i$  آن‌ها را پسندیده است. به صورت مشابه، ماتریس  $Q$  با ابعاد  $n \times n$  یک ماتریس قطری است که در آن عنصر قطری  $i$ ام برابر است با درجه‌ی رأس آیت  $i$ ، یا به عبارتی تعداد کاربرانی که آیت  $i$  را پسندیده‌اند. برای مثال به شکل زیر توجه کنید.

(الف)

ماتریس شباهت کاربر به صورت غیر نرمال شده را به صورت  $T = RR^T$  تعریف می‌کنیم (یعنی حاصل ضرب ماتریس  $R$  و ترانپوز آن). مفهوم عناصر  $T_{ij}$  (برای  $i \neq j$ ) را از منظر ساختارهای گراف دو بخشی (مانند درجه‌ی رأس‌ها، مسیر میان رأس‌ها و ...) توضیح دهید (به شکل ۱ مراجعه کنید).

**شباهت کسینوسی:**

یادآوری می‌کنیم که شباهت کسینوسی بین دو بردار  $\mathbf{u}$  و  $\mathbf{v}$  به صورت زیر تعریف می‌شود:

$$\cos - \text{sim}(\mathbf{u}, \mathbf{v}) = \frac{\mathbf{u} \cdot \mathbf{v}}{\|\mathbf{u}\| \|\mathbf{v}\|}$$

(ب)

ماتریس شباهت آیتم را به صورت  $S_I$  با ابعاد  $n \times n$  تعریف می‌کنیم، به طوری که عنصر سطر  $i$  و ستون  $j$  برابر است با شباهت کسینوسی بین آیتم  $i$  و آیتم  $j$  که متناظر با ستون‌های  $i$  و  $j$  از ماتریس  $R$  هستند. نشان دهید که:

$$S_I = Q^{-1/2} R^T R Q^{-1/2}$$

که در آن  $Q^{-1/2} = 1/\sqrt{Q_{rc}}$  به صورت  $Q^{-1/2}$  برای تمام درایه‌های غیر صفر تعریف شده است و برای سایر درایه‌ها برابر صفر است.

همین سؤال را برای ماتریس شباهت کاربران، یعنی  $S_U$  تکرار کنید؛ به طوری که عنصر سطر  $i$  و ستون  $j$  برابر است با شباهت کسینوسی بین کاربر  $i$  و کاربر  $j$  که متناظر با سطرها  $i$  و  $j$  از ماتریس  $R$  هستند. عبارت  $S_U$  را نیز به صورت یک عملیات ماتریسی بر حسب  $P$ ،  $R$  و  $Q$  بیان کنید؛ توجه داشته باشید که نباید به صورت درایه به درایه  $S_U$  را تعریف نمایید.

$$R = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

$$P = \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$Q = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 3 \end{bmatrix}$$

\* شکل ۱: گراف دو بخشی کاربر-آیتم.\*

پاسخ شما باید نحوه‌ی استخراج این رابطه‌ها را نشان دهد. (نکته: برای جذر درایه به درایه‌ی یک ماتریس، می‌توان آن را به توان  $1/2$  نوشت).

(ج)

روش پیشنهادی بر اساس فیلترینگ مشارکتی کاربر-کاربر  $u$  به این صورت است: برای تمام آیتم‌ها  $s$  محاسبه کنید

$$r_{u,s} = \sum_{x \in \text{کاربران}} \cos(\text{sim}(x, u)) \cdot R_{xs}$$

و سپس  $k$  آیتمی را پیشنهاد دهید که مقدار  $r_{u,s}$  در آن‌ها بیشینه است.

به صورت مشابه، روش پیشنهادی بر اساس فیلترینگ مشارکتی آیتم-آیتم برای کاربر  $u$  به این صورت است: برای تمام آیتم‌ها  $s$  محاسبه کنید

$$r_{u,s} = \sum_{x \in \text{آیتم‌ها}} R_{ux} \cdot \cos(\text{sim}(x, s))$$

و سپس  $k$  آیتمی را پیشنهاد دهید که مقدار  $r_{u,s}$  در آن‌ها بیشینه است.

ماتریس پیشنهادی  $\Gamma$  را تعریف می‌کنیم که ابعادی برابر با  $m \times n$  دارد، به طوری که  $\Gamma(i, j) = r_{i,j}$ . ماتریس  $\Gamma$  را برای هر دو روش فیلترینگ کاربر-کاربر و فیلترینگ آیتم-آیتم بر حسب  $R$ ،  $P$  و  $Q$  بیابید. پاسخ نهایی شما باید شامل عملیات در سطح ماتریسی باشد و نه تعریف درایه به درایه.

راهنمایی: در حالت آیتم-آیتم داریم:

$$\Gamma = RQ^{-1/2}R^T.$$

پاسخ شما باید روند رسیدن به این عبارت‌ها را (حتی در مورد آیتم-آیتم که فرمول نهایی داده شده) توضیح دهد.

(د)

جدول امتیازدهی زیر را بین پنج کاربر و شش آیتم در نظر بگیرید:

شناسه آیتم	۱	۲	۳	۴	۵	۶
۱	۵	۶	۷	۴	۳	؟
۲	۴	؟	۳	؟	۵	۴
۳	؟	۳	۴	۱	۱	؟
۴	۷	۴	۳	۶	؟	۴
۵	۱	؟	۳	۲	۲	۵

جدول ۱: امتیازدهی پنج کاربر (ردیف‌ها) به شش آیتم (ستون‌ها).

- (د-۱) مقادیر نامشخص امتیازدهی کاربر ۲ را با استفاده از الگوریتم فیلترینگ مشارکتی مبتنی بر کاربر پیش‌بینی کنید. از ضریب همبستگی پیرسون با میانگین‌گیری استفاده کنید. فرض کنید اندازه گروه هم‌تایان حداکثر ۲ است و همبستگی‌های منفی را حذف نمایید.
- (د-۲) مقادیر نامشخص امتیازدهی کاربر ۲ را با استفاده از الگوریتم فیلترینگ مشارکتی مبتنی بر آیتم پیش‌بینی کنید. از شباهت کسینوسی تعدیل‌شده استفاده کنید. فرض کنید اندازه گروه هم‌تایان حداکثر ۲ است و همبستگی‌های منفی را حذف نمایید.

اکنون، یک سامانه پیشنهاددهنده فیلم را در نظر بگیرید که در آن فیلم‌ها با ژانرها مرتبط هستند و امتیازدهی یک کاربر خاص مشخص است:

شناسه فیلم	کمدی	درام	عاشقانه	هیجانی	اکشن	ترسناک	دوست داشتن یا نداشتن
۱	۱	۰	۱	۰	۰	۰	دوست ندارد
۲	۱	۱	۱	۰	۱	۰	دوست ندارد
۳	۱	۱	۰	۰	۰	۰	دوست ندارد
۴	۰	۰	۰	۱	۱	۰	دوست دارد
۵	۰	۱	۰	۱	۱	۱	دوست دارد
۶	۰	۰	۰	۰	۱	۱	دوست دارد
آزمون-۱	۰	۰	۰	۱	۰	۱	؟
آزمون-۲	۰	۱	۱	۰	۰	۰	؟

جدول ۲: رابطه ژانر-فیلم و بازخورد کاربر در قالب دوست داشتن یا نداشتن.

تمام قوانین انجمنی را با حداقل پشتیبانی ۳۳٪ و اطمینان ۷۵٪ استخراج کنید. بر اساس این قوانین، آیا آیتم آزمون-۱ یا آزمون-۲ را به کاربر پیشنهاد می‌کنید؟

موارد مورد نیاز برای تحویل:

- تفسیر  $T_{ij}$  و  $T_{ii}$ .
- بیان  $S_I$  و  $S_U$  برحسب ماتریس‌های  $P$ ،  $R$  و  $Q$  به همراه توضیح.
- بیان  $\Gamma$  برحسب  $P$ ،  $R$  و  $Q$  به همراه توضیح.
- محاسبات مبتنی بر تکنیک‌های فیلترینگ مشارکتی.
- پاسخ به سوالات.

تاریخ تحویل:

سوالات عملی (۴۰ نمره)

پرسش ۱ (۲۰ نمره) به فایل جویپتر Q1.ipynb مراجعه کنید.

پرسش ۲ (۲۰ نمره) به فایل جویپتر Q2.ipynb مراجعه کنید.