

شبکه‌های اقتصادی و اجتماعی

دانشکده مهندسی کامپیوتر

مریم رضانی
بهار ۱۴۰۴



گراف‌ها و پیمایش تصادفی، تشخیص جوامع

تمرین سوم

تاریخ انتشار: ۱۷ اردیبهشت ۱۴۰۴

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

۲. لطفا پاسخ خود را با توضیحات کافی و با رسم شکل نگارش کنید.

سوالات تئوری (۷۰ نمره)

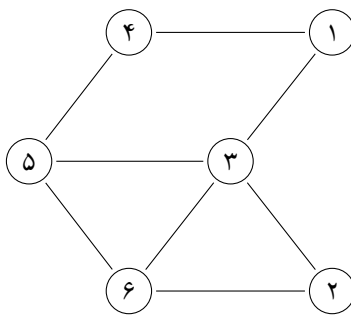
تاریخ تحویل: ۲۸ اردیبهشت ۱۴۰۴؛ با تاخیر: ۳۱ اردیبهشت ۱۴۰۴

پرسش ۱ (۱۰ نمره)

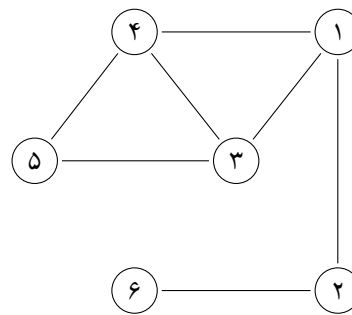
(آ) در مدل Erdős-Rényi با n راس و میانگین درجه λ نشان دهید که قطر گراف برابر است با $\frac{\log n}{\log \lambda}$.

(ب) اگر مقدار λ کمتر از ۱ یا بیشتر از ۱ باشد قطر گراف را تحلیل کنید.

پرسش ۲ (۱۰ نمره) شباهت بین دو گراف A و B را با استفاده از تمام Graphlet Kernels های سه تایی محاسبه کنید.



B (ب)



A (آ)

پرسش ۳ (۱۰ نمره) در یک شبکه اجتماعی، گره‌ها نشان‌دهنده افراد و یال‌ها نشان‌دهنده ارتباطات بین آن‌ها هستند. شما تصمیم دارید که ساختار این شبکه را مدل‌سازی کنید.

(آ) مدل‌های Erdős-Rényi و Watts-Strogatz را مقایسه کنید و توضیح دهید که کدام یک بیشتر شبیه یک شبکه اجتماعی واقعی است؟ چرا؟

(ب) مدل Kleinberg's geographical model چگونه می‌تواند توجیهی برای ارتباطات محلی و جهانی در شبکه‌های اجتماعی ارائه دهد؟ توضیح دهید که چگونه این مدل به پدیده "شش درجه جدایی" مرتبط است.

پرسش ۴ (۱۰ نمره) در بسیاری از شبکه‌های واقعی، توزیع درجه گره‌ها از نوع heavy-tailed distribution است.

(آ) تفاوت بین توزیع نرمال و Power-law Distribution در شبکه‌ها چیست؟ چگونه می‌توان فهمید که یک شبکه از Power-law Distribution پیروی می‌کند؟

(ب) مدل‌های Barabási-Albert و Forest Fire را مقایسه کنید. چگونه هر یک از این مدل‌ها می‌توانند هسته‌های بزرگ در شبکه‌ها را توضیح دهند؟

(ج) چگونه مدل Zipf & Pareto می‌تواند توزیع ارتباطات در شبکه‌های اجتماعی را توجیه کند؟

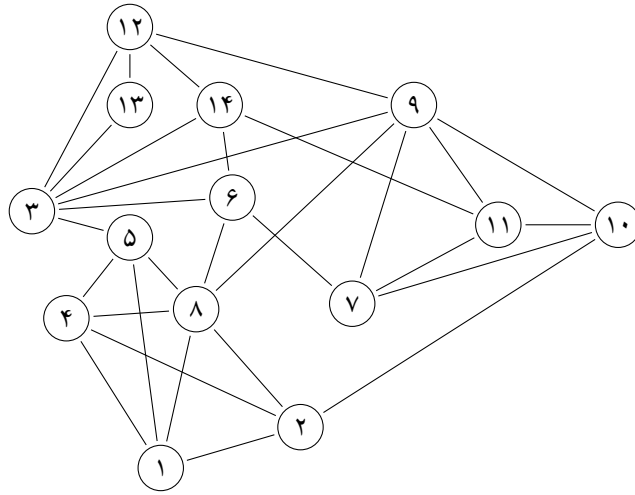
پرسش ۵ (۱۰ نمره) شما مسئول طراحی یک الگوریتم برای جستجوی اطلاعات در یک شبکه استنادی علمی هستید، که در آن گره‌ها مقالات علمی و یال‌ها استنادات هستند.

(آ) توضیح دهید که چگونه می‌توان از الگوریتم PageRank یا HITS برای یافتن مقالات مهم در این شبکه استفاده کرد؟ تفاوت این دو الگوریتم چیست؟

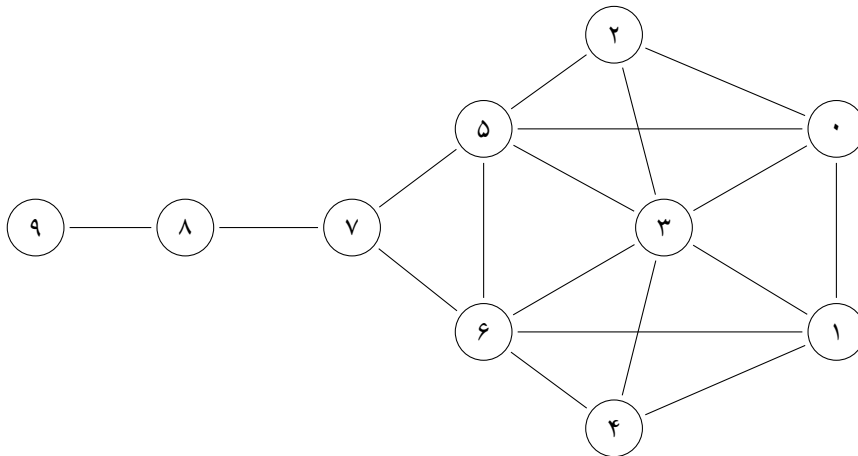
(ب) مدل Kronecker Graph Model چگونه می‌تواند برای پیش‌بینی ارتباطات آینده در یک شبکه مورد استفاده قرار گیرد؟

(ج) در شبکه‌های اطلاعاتی، مسیرهای کوتاه‌ترین و انتشار اطلاعات چه نقشی در بازیابی اطلاعات دارند؟ مثالی از یک کاربرد عملی ارائه دهید.

پرسش ۶ (۱۰ نمره) باتوجه به گراف زیر به سوالات پاسخ دهید.



- (آ) یک گراف ۳ منتظم همبند با ۸ رأس را پیدا کنید.
 (ب) یک گراف ۳ منتظم با ۸ رأس را پیدا کنید و نشان دهید که نسبت به اولی هم‌ریخت نیست.
 (ج) چهار گراف ۳ منتظم دیگر با ۸ رأس را بیابید و نشان دهید هیچ دو شکلی هم‌ریخت نیستند.
 پرسش ۷ (۱۰ نمره) با توجه به گراف زیر به سؤالات پاسخ دهید.



(آ) مشخص کنید هر کدام از موتیف‌های زیر چند بار تکرار شده است (با رسم شکل).



(ب) تعداد وقوع هر موتیف از ۱۰ گراف تصادفی آورده شده است. بیان کنید که آیا هر یک از این موتیف‌ها از نظر آماری over-represented شده‌اند یا خیر. اگر مقدار z در بیشتر از $1/150$ باشد، از نظر آماری over-represented در نظر گرفته می‌شود.

- Motif E: [9, 8, 1, 6, 8, 13, 6, 8, 3, 9]
- Motif R: [7, 5, 7, 6, 8, 5, 0, 7, 9, 7]

پرسش ۲ (۱۵ نمره) در این تمرین، هدف شما درک، پیاده‌سازی و مقایسه‌ی الگوریتم‌های مختلف تشخیص اجتماع^۱ در یک شبکه‌ی واقعی از مقالات علمی^۲ است. همچنین در قسمت دوم این سوال نیاز است تا شما با استفاده از ساختار شبکه، ویژگی‌های رئوس و برجسب‌های واقعی هر رأس استفاده کنید تا کیفیت خوشه‌بندی‌هایی که تعیین کردید را بررسی نمایید. داده مورد بررسی در این تمرین، داده Cora شامل ۲۷۰۸ مقاله علمی است که در یک شبکه استنادی با ۵۴۲۹ یال ارجاع‌دهی شده‌اند. همچنین این مقالات به یکی از ۷ کلاس مختلف موضوعی زیر تعلق دارند:

- Case Based
- Genetic Algorithms
- Neural Networks
- Probabilistic Methods
- Reinforcement Learning
- Rule Learning
- Theory

این داده شامل دو فایل اصلی است.

• **cora.content** که اطلاعات مربوط به ویژگی‌های هر مقاله به صورت بردارهای باینری از کلمات موجود در لغت‌نامه نشان داده و همچنین شامل موضوع اصلی هر مقاله است.

• **cora.cites** گراف استنادی که در آن هر سطر نشان‌دهنده یک ارتباط استنادی (ارجاع‌دهی) بین دو مقاله است.

به هر یک از بخش‌های زیر در یک سلول یک دفترچه ژوپیتر پاسخ دهید.

(آ) داده‌های موجود در فایل‌های **cora.cites** و **cora.content** را بارگذاری کنید.

(ب) تعداد مقالات در هر یک از دسته‌های موضوعی را با نمودار مناسب نشان دهید.

(ج) گراف استنادها را بر اساس داده‌های **cora.cites** بسازید.

(د) توزیع درجات این گراف را نمودار کنید. میانگین این درجات چقدر است؟

(ه) تعداد و اندازه مولفه‌های هم‌بندی را مشخص کنید.

(و) میانگین کوتاه‌ترین فواصل در مولفه‌های هم‌بندی این گراف چند است؟

(ز) با استفاده از الگوریتم‌های **Girvan Newman** و **Louvain** و **Lukes** جوامع موجود در این گراف را تشخیص دهید. می‌توانید از پیاده‌سازی‌های موجود در کتاب‌خانه **networkx** استفاده کنید.

(ح) با توجه به غیر نظارتی بودن مساله تشخیص جوامع، توضیح دهید چگونه می‌توان از معیار **F1** برای این مساله استفاده کرد. آن را پیاده‌سازی کنید و روش‌های فوق را با استفاده از آن مقایسه کنید.

(ط) با استفاده از معیار **Normalized Mutual Information** عملکرد این روش‌ها را بررسی کنید.