

به نام خدا  
دانشگاه صنعتی امیرکبیر (پلی تکنیک تهران)  
دانشکده مهندسی کامپیوتر



دانشگاه صنعتی امیرکبیر  
(پلی تکنیک تهران)

## تحلیل شبکه‌های پیچیده تمرین اول

دانشجو: رضا ساجدی

۴۰۰۱۳۱۰۷۲

استاد: دکتر چهرقانی

پاییز ۱۴۰۱

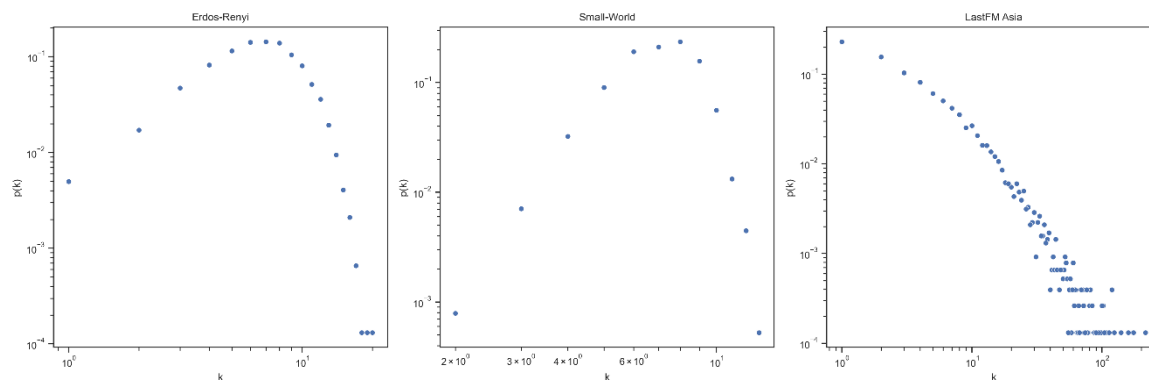
## فهرست مطالب

سؤال اول	..... ۳
سؤال دوم	..... ۴
سؤال سوم	..... ۵
سؤال چهارم	..... ۶
سؤال پنجم	..... ۷

## سؤال اول

توزیع درجه گراف‌های دنیای واقعی (LastFM) از نوع پاورلو<sup>۱</sup> است. نمودار توزیع درجه مدل دنیای کوچک<sup>۲</sup> و مدل تصادفی ER تقریباً شبیه یکدیگر است و با توزیع درجه گراف‌های دنیای واقعی تفاوت دارد. توزیع درجه آنها از خانواده توزیع‌های نرمال (شبیه توزیع دو جمله‌ای) است. برای ایجاد مدل دنیای کوچک از یک حلقه دایره‌ای شکل استفاده شده است. مقدار احتمال تصادفی‌سازی در مرحله سیم‌کشی مجدد<sup>۳</sup> برابر با ۲۴٪ در نظر گرفته شده است. مقدار ضریب خوشه‌بندی مدل دنیای کوچک، مشابه گراف دنیای واقعی شده است. در واقع هدف اصلی ارائه این مدل، داشتن ضریب خوشه‌بندی بالا در کنار طول مسیر میانگین پایین بوده است. این هدف با قرار دادن تعداد کمی از یال‌های تصادفی در کنار یال‌های منظم تحقق یافته است. ضریب خوشه‌بندی مدل ER با گراف دنیای واقعی سازگاری ندارد و کمتر است.

Erdos-Renyi's Clustering Coefficient: 0.0008  
Small-World's Clustering Coefficient: 0.2861  
LastFM Asia's Clustering Coefficient: 0.2850



<sup>1</sup> Power law

<sup>2</sup> Small-world

<sup>3</sup> Rewiring

## سؤال دوم

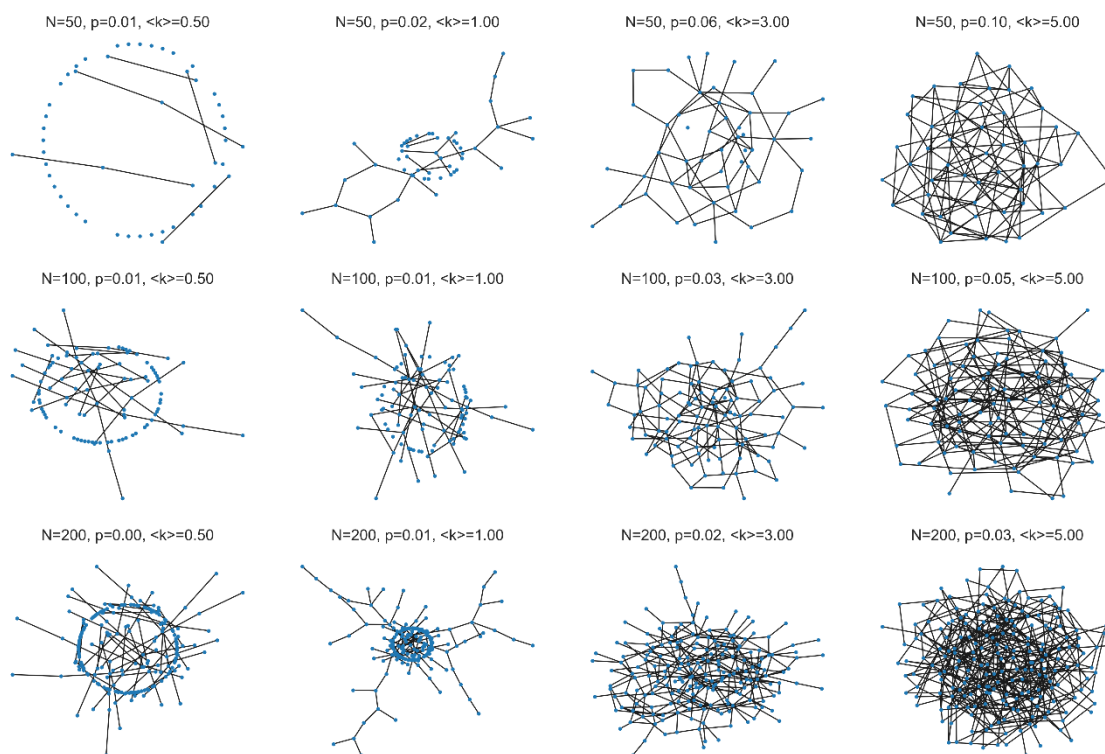
در شکل زیر نحوه تغییر اندازه نسبی بزرگترین مؤلفه گراف‌های تصادفی با توجه به افزایش میانگین درجه نمایش داده شده است. برای ایجاد گراف‌های تصادفی از مدل ER استفاده شده است. مقادیر آرگومان‌های انتخابی برای هر مورد در بالای آن نمایش داده شده است. میانگین درجه در گراف‌های ER از رابطه زیر به دست می‌آید:

$$\langle k \rangle = p \times (N - 1)$$

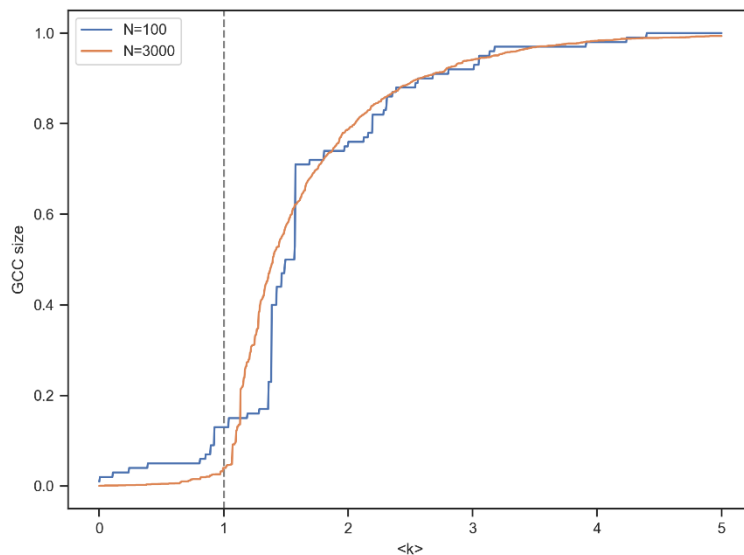
برای اینکه مقادیر میانگین درجه، مشابه صورت سؤال شود، مقادیر احتمال به صورت زیر تعریف شده است:

$$p_1 = \frac{0.5}{N-1}, \quad p_2 = \frac{1}{N-1}, \quad p_3 = \frac{3}{N-1}, \quad p_4 = \frac{5}{N-1}$$

پارامتر N نیز با سه مقدار ۵۰، ۱۰۰ و ۲۰۰ آزمایش شده است. همان‌طور که مشاهده می‌کنید، نتایج با گزاره‌های مطرح شده در چهار رژیم، مطابقت دارد.



در نمودار زیر اندازه نسبی بزرگترین مؤلفه بر حسب میانگین درجه برای یک گراف با تعداد گره ۱۰۰ و گرافی دیگر با تعداد گره ۳۰۰۰ نمایش داده شده است. از مقایسه این دو منحنی می‌توان نتیجه گرفت که هرچه تعداد گره‌ها بیشتر باشد، دقت گزاره‌ها بیشتر است. منحنی نارنجی که مربوط به گراف با ۳۰۰۰ گره است، شباهت بیشتری به منحنی موجود در صورت سؤال دارد.



## سؤال سوم

(الف) بیشترین تعداد لینک می‌تواند برابر با  $N1 \cdot N2$  باشد.

(ب)

$$\frac{(N1 + N2)(N1 + N2 - 1) - 2N1N2}{2} = \frac{(N1^2 + N2^2) - (N1 + N2)}{2}$$

(ج) چگالی کل شبکه تقریباً برابر با چگالی قسمتی که دارای  $N2$  نود است می‌شود. درواقع وابسته به مقدار  $N2$  و احتمال وجود نیم‌یال از این قسمت می‌باشد.

(د) فرض کنید  $m$  یال بین این دو نوع از رؤوس وجود داشته باشد. میانگین درجه بخش اول برابر با  $k1 = m/N1$  و میانگین درجه بخش دوم برابر با  $k2 = m/N2$  می‌شود. بنابراین خواهیم داشت:

$$\frac{k1}{k2} = \frac{m/N1}{m/N2} = \frac{N2}{N1} \rightarrow k1 = k2 \frac{N2}{N1}$$

## سؤال چهارم

پیچیدگی زمانی الگوریتم حریصانه<sup>۴</sup>  $O(k.n.R.M)$  است. در هر تکرار این الگوریتم، لازم است سود حاشیه‌ای<sup>۵</sup> برای همه گره‌های باقی‌مانده محاسبه شود و گره‌ای که بیشترین سود حاشیه‌ای دارد به مجموعه گره‌های انتخابی اضافه شود. اما در الگوریتم CELF با توجه به اینکه در تکرارهای بعد سود حاشیه‌ای گره‌ها کمتر شده یا یکسان می‌ماند، در بیشتر مواقع در عمل دیگر لازم نیست سود حاشیه‌ای برای همه گره‌های باقی‌مانده محاسبه شود و تنها در نظر گرفتن عنصر بالای لیست مرتب‌شده، کفایت می‌کند. همین امر منجر به افزایش سرعت الگوریتم CELF می‌شود. نتایج آزمایش ما نشان می‌دهد که الگوریتم CELF تقریباً ۳۰۰ برابر سریع‌تر از الگوریتم حریصانه بوده است و هرچقدر تعداد دفعات شبیه‌سازی (R) بیشتر شود، این نسبت نیز بیشتر می‌شود. سه مورد از گره‌های انتخابی توسط این دو الگوریتم، یکسان بوده است و هرچقدر تعداد دفعات شبیه‌سازی افزایش پیدا کند، احتمالاً نتایج الگوریتم‌ها بیشتر با یکدیگر همخوانی پیدا کند. در این آزمایش، تعداد دفعات شبیه‌سازی برابر ۱۰۰ و احتمال فعال‌سازی یال‌ها برابر با ۰/۰۰۱ در نظر گرفته شده است.

Greedy result: 93504, 62821, 81286, 67303, 37653, 64054, 92790, 124566, 35290, 53213

CELF result: 44866, 62821, 34608, 53213, 67410, 92790, 115607, 50808, 45530, 38109

Similar nodes: 53213, 62821, 92790

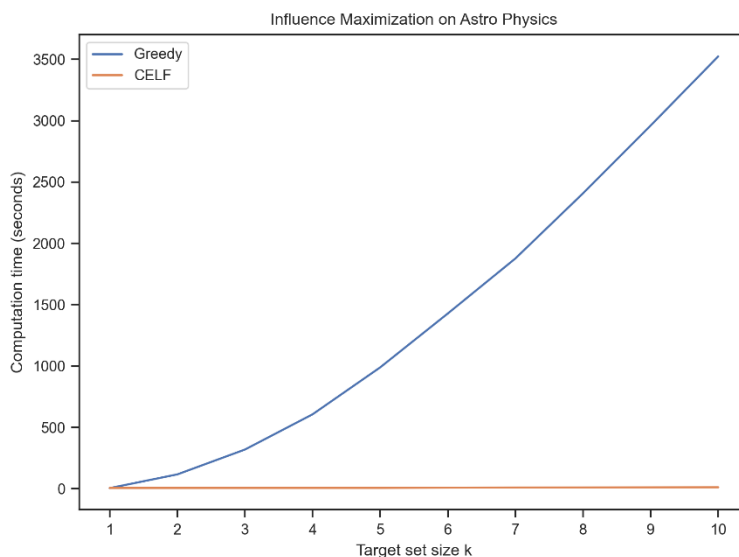
Greedy computation time: 3524.07s

CELF computation time: 11.82s

---

<sup>4</sup> Greedy

<sup>5</sup> Marginal gain



## سؤال پنجم

مسئله کشف شیوع<sup>۶</sup>، مشابه مسئله بیشینه‌سازی تأثیر<sup>۷</sup> است. تفاوت اصلی آن در این است که به جای تعیین تعداد گره‌های انتخابی، یک بودجه تعیین می‌شود و مجموع هزینه گره‌های انتخابی نباید از بودجه تعیین شده فراتر رود. برای استفاده از الگوریتم حریصانه‌ای که در مسئله بیشینه‌سازی تأثیر معرفی گردید، می‌توان هزینه گره‌ها را نادیده گرفت و گره‌هایی که بیشترین پاداش را دارند را انتخاب کرده و در پایان شرط هزینه بررسی شود. در این حالت در برخی موارد ممکن است انتخاب‌های بهتری وجود داشته باشد که پاداش نهایی را افزایش دهد و نادید گرفته شود.

برای حل این مشکل، در الگوریتم CELF یک بار مشابه الگوریتم حریصانه عمل می‌شود و برای انتخاب گره‌ها به هزینه آنها توجه نمی‌شود. یک بار هم بهترین گره‌ها با توجه به نسبت پاداش به هزینه آنها انتخاب می‌شود. در نهایت از بین این دو مجموعه گره، مجموعه‌ای که بیشترین پاداش را داشته باشد به عنوان جواب نهایی در نظر گرفته می‌شود.

برای افزایش سرعت در الگوریتم CELF نیز با توجه به اینکه در تکرارهای بعد سود حاشیه‌ای گره‌ها کمتر شده یا یکسان می‌ماند، در بیشتر مواقع در عمل دیگر لازم نیست سود حاشیه‌ای برای همه

<sup>۶</sup> Outbreak detection

<sup>۷</sup> Influence maximization

گره‌های باقی‌مانده محاسبه شود و تنها درنظر گرفتن عنصر بالای لیست مرتب‌شده، کفایت می‌کند. همین امر منجر به افزایش سرعت الگوریتم می‌شود.

در این آزمایش، مقدار بودجه برابر با ۰/۰۰۱ و تعداد شیوع‌ها نیز برابر با ۵۰ درنظر گرفته شده است. پاداش حاصل از الگوریتم CELF تقریباً ۵ برابر الگوریتم حریصانه بوده است. در این مورد، الگوریتم CELF تقریباً ۱۳ برابر سریع‌تر از الگوریتم حریصانه عمل کرده است.

Greedy benefit: 3.33

Greedy nodes: 93504, 81761, 9667, 69643, 23596, 75223, 35290, 86651, 53213, 32191

Greedy computation time: 692.61s

CELF benefit: 16.6

CELF nodes: 124928, 34945, 70275, 115076, 12680, 11913, 63625, 59277, 59278, 85779, 110740, 70291, 102806, 5527, 66707, 32664, 118170, 6548, 32796, 133142, 58654, 120732, 2856, 27690, 108330, 125612, 116397, 100778, 57649, 38450, 55729, 63284, 33203, 73271, 12218, 131133, 95680, 39108, 48073, 42572, 107852, 60108, 57295, 69202, 126550, 125399, 70488, 34524, 19166, 33886, 111072, 85986, 28003, 101994, 69228, 47213, 2159, 78191, 76784, 8561, 67955, 89461, 118520, 85883, 55805

CELF computation time: 50.91s

