

گزارش تکلیف اول درس الگوریتم شبکه های پیچیده.

اعضای گروه :

۲. امیرحسین سهراب بیگ ۹۳۳۱۰۶۵

۳. امیرحسین باوند ۹۳۳۱۰۲۸

استاد : دکتر امیرحائری.

سوال 2

احتمال را برابری P فرض کنیم

$$\begin{aligned}
 &= 3 \times (1-P)^6 + 4 \times (1-P)^5 \times (1-(1-P)^2) \\
 &+ 4 \times (1-P)^5 \times (1-(1-P)^2) + 4 \times (1-P)^4 \times \\
 &\quad (1-(1-P)^3) \times (1-P) + 5 \times (1-P)^4 \times (1-(1-P)^3) \\
 &\quad \times P + 5 \times ((1-P)^3 \times (1-(1-P)^2) \times (1-(1-P)^2)) \\
 &+ 5 \times \\
 &+ 6 \times ((1-(1-P)^3) \times (1-(1-P)^2) \times ((1-P) \times (1-(1-P)^2)))
 \end{aligned}$$

سوال 3

الف) «این» حرف خوشی g, j, i, n, m, h, c, d
از آوایی گای را به یادگیری برسد

ب) اگر گره ai را اضافه کنیم حرف کاملاً آوده می شود

ج) اگر صد آواز $\frac{1}{3}$ شود کل حرف آوده می شود

سوال چهار :

در ابتدا چند نکته : از آن جایی که گراف بدون جهت و bipartite است اصلا درجه ورودی و خروجی معنی ندارد. همچنین گراف نیز اصلا conected نیست بنابراین برای این گراف قطر نیز معنا ندارد اما فرض میکنیم منظور از قطر قطر بزرگ ترین کامپوننت متصل به آن است و برای آن محاسبات را انجام میدهیم

همچنین برای محاسبه ضریب خوشه بندی نیز از مقاله [1] استفاده میکنیم

ابتدا برای random edge محاسبات را انجام میدهیم

در این گراف میانگین ضریب خوشه بندی را محاسبه میکنیم

روش یال تصادفی	میانگین ضریب خوشه بندی		
پنج درصد کل گراف	0.36669265673 2429		
ده درصد کل گراف	0.38043605253 912205		
پانزده درصد کل گراف	0.36661372163 808403		
بیست درصد کل گراف	0.35347952499 74704		
گراف کلی	0.28395745331 27902		

حال برای random node این کار را انجام میدهیم به این صورت که تعدادی از گره ها را از هر یک از دو طرف انتخاب میکنیم و گراف ها را می سازیم

	میانگین ضریب خوشه بندی		
پنج درصد کل گراف	0.06865572279 092409		

		0.11087231411 958479	ده درصد کل گراف
		0.13793152676 665715	پانزده درصد کل گراف
		0.14855649573 598873	بیست درصد کل گراف
		0.28395745331 27902	گراف کلی

حال با استفاده از روش پرش تصادفی گراف را تشکیل می‌دهیم

		میانگین ضریب خوشه بندی	
		0.21330333265 990947	پنج درصد کل گراف
		0.17735191123 990715	ده درصد کل گراف
		0.16363295164 872338	پانزده درصد کل گراف
			بیست درصد کل گراف
		0.28395745331 27902	گراف کلی

حال با استفاده از الگوریتم جستجوی اول عمق گراف را تشکیل می‌دهیم

		میانگین ضریب خوشه بندی	
		0.36735102042 81517	پنج درصد کل گراف
		0.39229332722 306803	ده درصد کل گراف

		0.45320473109 74926	پانزده درصد کل گراف
		0.52703970476 7855	بیست درصد کل گراف
		0.28395745331 27902	گراف کلی

نکاتی که در این سوال مطرح است شامل دو مورد میشود
اول این که برای بدست آوردن قطر گراف الگوریتم دی اف اس از همه بهتر عمل میکند زیرا دی اف اس گراف را به صورت عمقی تا جایی که ممکن است پیمایش میکند و معمولا برای یک مولفه همبند جواب های خوبی به دست می آورد و میتواند قطر گراف را به دست آورد.
در مورد ضریب خوشه بندی نیز با توجه به مقاله ای که بر اساس آن ضریب خوشه بندی را به دست آورده ایم روش یال تصادفی چون تقریباً از همه یال ها به طور یکنواخت استفاده میکند میتواند موثر تر باشد و در نمونه گیری نیز جواب های دقیق تری به دست آمده است

سوال پنج :

برای این که بفهمیم که کدام گره ها بیشترین مقدار را میتوانند بگیرند به این صورت عمل میکنیم که گره های پر درجه را در نظر میگیریم و آن ها را به عنوان گره های کاندیدا به حساب می آوریم و عملیات را بر روی آن ها انجام میدهم این کار باعث میشود که بسیاری از گره ها حذف شوند و دایره انتخاب ما محدود شود.
همچنین در این سوال ما از روش threshold استفاده میکنیم و threshold را نیز برابر یک دهم در نظر میگیریم و به این ترتیب مساله را حل میکنیم هم چنین در این سوال مشکل دیگری نیز وجود داشت و آن خود فرآیند انتشار بود که به شدت زمان گیر بود به این منظور ما از یک مقدار اندکی از گراف در حدود ده درصد از گراف استفاده میکنیم نکته ای که در این سوال وجود دارد این است که گراف دارای نود های پر درجه زیادی است در واقع قسمت فیلم ها اکثراً پر درجه هستند بنابراین انتشار زیادی میتواند اتفاق بیفتد

threshold را برابر چهاردهم در نظر میگیریم و برای هر کدام از درصدهای گفته شده اعداد زیر را به دست می آوریم

یک درصد : {0: 45597, 1: 1190}

دو درصد : {0: 44977, 1: 1810}

سه درصد : {0: 44511, 1: 2276}

چهار درصد : {0: 43905, 1: 2882}

پنج درصد : {0: 43285, 1: 3502}

شش درصد : {0: 42606, 1: 4181}

هفت درصد : {0: 41011, 1: 5776}

هشت درصد : {0: 14695, 1: 32092}

نه درصد : {0: 14655, 1: 32132}

ده درصد : {0: 14628, 1: 32159}

برای این که انتشار در بیش از نودو پنج درصد گراف اتفاق بیفتد باید حدود بیست و پنج درصد گراف را آلوده کنیم
{0: 1837, 1: 44950}

سوال شیش :

در این جا میتوان انتشار را برای کلاسترها جدا در نظر گرفت یعنی ابتدا گراف را به چند کلاستر تقسیم کنیم سپس برای هر کلاستر به صورت جدا عملیات انتشار را انجام دهیم تا در نهایت به یک جواب قابل قبول برسیم. با جدا فرض کردن کلاسترها دیگر برای هر کدام نیاز نیست که در همه ی گراف بگردیم و صرف پیدا کردن درون هر کلاستر کافی است تعداد نقاط اولیه در هر کلاستر را نیز با توجه به اندازه کلاستر ها تعیین میکنیم
همچنین اگر باز هم تعداد نقاط درون یک کلاستر زیاد شد میتوانیم ایده مساله قبل را برای هر کلاستر تکرار کنیم و در نهایت جواب هایی که به دست می آید به صورت زیر میشود

همچنین میتوان تعداد نقاط آلوده اولیه در هر خوشه را دستکاری کرد که ما به دلیل کمبود وقت از این کار اجتناب کردیم اما قابل پیاده سازی است

```
{iteration': 199, 'status': {}, 'node_count': {0: 22, 1: 10}, 'status_delta': {0: 0, 1: 0}}  
[(a6', 1'), (b20', 1')]  
{iteration': 199, 'status': {}, 'node_count': {0: 30, 1: 2}, 'status_delta': {0: 0, 1: 0}}  
[(b21', 1'), (b23', 1'), (a8', 1'), (a9', 2'), (a7', 2'), (b22', 3')]  
{iteration': 199, 'status': {}, 'node_count': {0: 26, 1: 6}, 'status_delta': {0: 0, 1: 0}}
```

سوال هفت :

سوال هشت :

در این جا گراف را مانند تکلیف یک میسازیم و سپس از روش knn برای پیش بینی استفاده میکنیم
دلیل این که از knn استفاده میکنیم این است که داده ها عددی هستند و در چند بعد پراکنده اند بنابراین knn با وجود این که مدل بسیار ساده ای است در این مساله بسیار خوب عمل میکند و دقتی برابر با نود و نه درصد به دست می آورد که برای یک مساله چند کلاسه بسیار عالی میباشد
در واقع میبینیم از بین همسایه های آن کدام یک بیشترین تعداد را دارد و این گونه گراف را پیش بینی میکنیم

Confusion matrix را محاسبه میکنیم که به صورت زیر میشود
خطا برابر است با یک درصد

```
[0 0 0 0 0 0 0 0 0 42]  
[0 0 0 0 0 0 0 0 36 0]  
[0 0 0 0 0 0 0 31 0 0]  
[0 1 0 0 0 0 33 0 0 0]  
[0 0 0 0 0 32 0 0 0 0]  
[1 0 0 1 36 0 0 0 0 0]  
[0 0 0 35 0 0 0 0 0 0]  
[0 0 33 0 0 0 0 0 0 0]  
[0 37 0 0 0 0 2 0 1 0]  
[[35 0 0 0 0 0 3 0 1 0]
```

Latapy, Matthieu, Clémence Magnien, and Nathalie Del Vecchio (2008).
Basic notions for the analysis of large two-mode networks. *Social Networks*
30(1), 31–48.