



زن زندگی آزادی



دانشگاه تهران
دانشکده مهندسی برق و کامپیوتر
درس شبکه‌های اجتماعی

تمرین سری اول

سیاوش رزمی	نام و نام خانوادگی
۸۱۰۱۰۰۳۵۲	شماره دانشجویی
۰۱/۰۹/۱۸	تاریخ ارسال گزارش

فهرست گزارش سوالات

- 1..... توضیحات کد
- 2..... روش پیاده سازی
- 4..... تحلیل نتایج

۱- کد برنامه با استفاده از کتابخانه‌های زیر نوشته شده است:

۱. NetworkX: for Graph Creation

توابع استفاده شده:

`complete_graph`: for making a complete graph as the basis

`add_nodes_from`: add multiple nodes to the graph

`shortest_path_length`: used for finding shortest path between two given nodes

۲. Numpy: for matrix calculation

۳. matplotlib: for plotting the results

۴. random: for random number generation

۵. joblib: for parallelizing the code

از کتابخانه `joblib` جهت موازی سازی و کاهش زمان فرایند ساخت گراف استفاده شده است.

تعداد ۴ `job` مجزا به شکل موازی بر روی پردازنده نتیجه کد را محاسبه میکنند.

روش پیاده سازی

بطور کلی برای اثبات درستی فرمول ها، گراف هایی با مشخصات داده شده (گاما و متوسط درجه) و با تعداد گره های افزایشی تولید میکنیم سپس مقدار متوسط فاصله را در این گراف ها محاسبه کرده و نشان می دهیم هر چه تعداد گره ها در گراف افزایش پیدا میکند مقادیر محاسبه شده به مقادیر بدست آمده نزدیک تر می شود.

روش ساخت گراف:

طبق صورت سؤال بایستی گراف هایی با توزیع powerlaw و مقادیر گاما مختلف تولید کنیم که برای این کار از گراف Bianconi-barabasi کرده ایم. به این شکل که:

۱- ابتدا یک گراف کامل که درجه گره ها در آن برابر با حداقل درجه گراف است تولید میکنیم.

۲- سپس با توجه به تعداد گره مدنظر به تعداد گره باقی مانده یک حلقه اجرا میکنیم که:

۱.۱- لیست گره های موجود در گراف را محاسبه میکنیم

۱.۲- لیست درجه گره های موجود در گراف را محاسبه میکنیم

۱.۳- با توجه به مقادیر eta و ورودی مقادیر fitness را محاسبه میکنیم

۱.۴- با ضرب مقادیر fitness در توزیع درجه ها و تقسیم آن به جمع کل آن ها مقدار احتمال انتخاب هر گره را به دست می آوریم.

۱.۵- بر اساس احتمال به دست آمده به تعداد حداقل درجه از گره های موجود به شکل تصادفی گره انتخاب میکنیم.

۱.۶- از گره تازه اضافه شده به گره های انتخاب شده یال اضافه میکنیم.

(البته برای پیاده سازی به دلیل کند بود اضافه کردن گره در حلقه از ابتدا تمامی گره ها در گراف اضافه شدند و صرفاً با اندیس دهی در هر مرحله یال جدید به گراف اضافه میکنیم)

روش محاسبه مقادیر Eta در Fitness Function:

برای محاسبه مقادیر eta جهت تغییر گاما از روش زیر استفاده میکنیم:

۱. برای گاما های حدود ۲، به جز ۲ گره اول مقدار eta متناظر مابقی گره ها را صفر قرار داده و برای ۲ گره اول مقداری مخالف صفر، با این روش مقدار گاما چیزی حدود ۲.۱۱ خواهد شد، دلیل اینکار این است که ما برای به دست آوردن گاما برابر با ۲ به گرافی با توپولوژی شبیه به ستاره نیاز داریم، اما برای اینکه متوسط درجه گراف با بقیه گراف ها برابر باشد نمی توانیم حداقل درجه را ۱ بگذاریم (متوسط درجه در مابقی گراف ها عددی حدود ۳.۹ است و متوسط درجه با حداقل درجه ۱ حدود ۲ است) بنابراین با اضافه کردن یک گره دیگر به عنوان hub گراف میتوان علاوه بر بدست آوردن گامایی حدود ۲.۱۱، متوسط درجه ۳.۹ را به دست آورد.

۲. برای گاما های بین ۲ و ۳، به جز یک گره مقدار eta تمام گره ها را مقداری برابر و مخالف صفر قرار می دهیم، و برای یک گره مقدار eta را بیشتر از تمامی گره ها قرار میدهیم.

۳. برای گامای برابر با ۳، مقدار تمام eta ها را برابر با یکدیگر قرار می دهیم.

۴. برای گاما های بیشتر از ۳، به شکل صعودی مقادیر eta را به نسبت زمان اضافه شدن به گراف افزایش می دهیم.

برای اینکه مقدار متوسط درجه را گراف ها ثابت نگه داریم، حداقل درجه گره ها را در گراف برابر با ۲ در نظر می گیریم، و از هر نوع گراف ۲۰ نمونه می سازیم (برای گراف با تعداد گره ۱۰۰۰۰۰ به دلیل زمان زیاد تولید ناچار به تولید تنها یک عدد هستیم)

روش محاسبه پارامتر های گراف:

۱- گاما:

پس از ساخت گراف برای محاسبه گاما از فرمول زیر استفاده میکنیم:

$$K_{max} = K_{min} N^{\frac{1}{\gamma-1}}$$

بدین صورت که با بدست آوردن دنباله درجات گراف و محاسبه ماکزیمم و مینیمم گراف میتوان مقدار گاما را به شکل زیر بدست آورد:

$$\gamma = 1 + \frac{\log N}{\log \frac{K_{max}}{K_{min}}}$$

۲- متوسط فاصله:

برای بدست آوردن متوسط فاصله به اندازه یک درصد مشخص به نام `sampling_ratio` از تعداد گره‌های گراف، به شکل تصادفی زوج گره انتخاب میکنیم و سپس فاصله آن‌ها را به دست آورده و میانگین میگیریم. اندازه سمپل گرفته شده را نیز حدود ۵ درصد تعداد کل گره‌ها در نظر میگیریم.

۳- متوسط درجه:

با میانگین گرفتن از دنباله درجات گره‌ها میتوان متوسط درجه گراف را بدست آورد.

این سه فرمول در سه تابع `calc_gamma`، `calc_average_distance` و `calc_average_degree` تعریف شده اند.

برای سهولت در ساخت گراف‌ها یک کلاس به نام `graph` تعریف شده است که در آن با مشخص کردن تعداد گره، مقدار گاما (به شکل حدودی) حداقل درجه یک گراف تعریف می‌شود و با اجرای متد `run` گراف ساخته شده و مقادیر مربوط به آن نیز محاسبه می‌گردد، همچنین برای جلوگیری از ساخت گراف‌های تکراری، میتوان گراف با تعداد گره کمتر را به عنوان ورودی به کلاس داده و با اضافه کردن گره به آن (با استفاده از پارامترهای `init_graph` و `G`) تعداد گره‌ها را افزایش داد.

در نهایت با استفاده از کتابخانه `joblib` کد به حالت موازی در آمده تا ۲۰ نمونه از گراف را به شکل موازی و همزمان تولید نماید پس از ساخت هر نوع گراف نتایج برای بررسی و تحلیل در هارد دیسک ذخیره میشود.

تحلیل نتایج

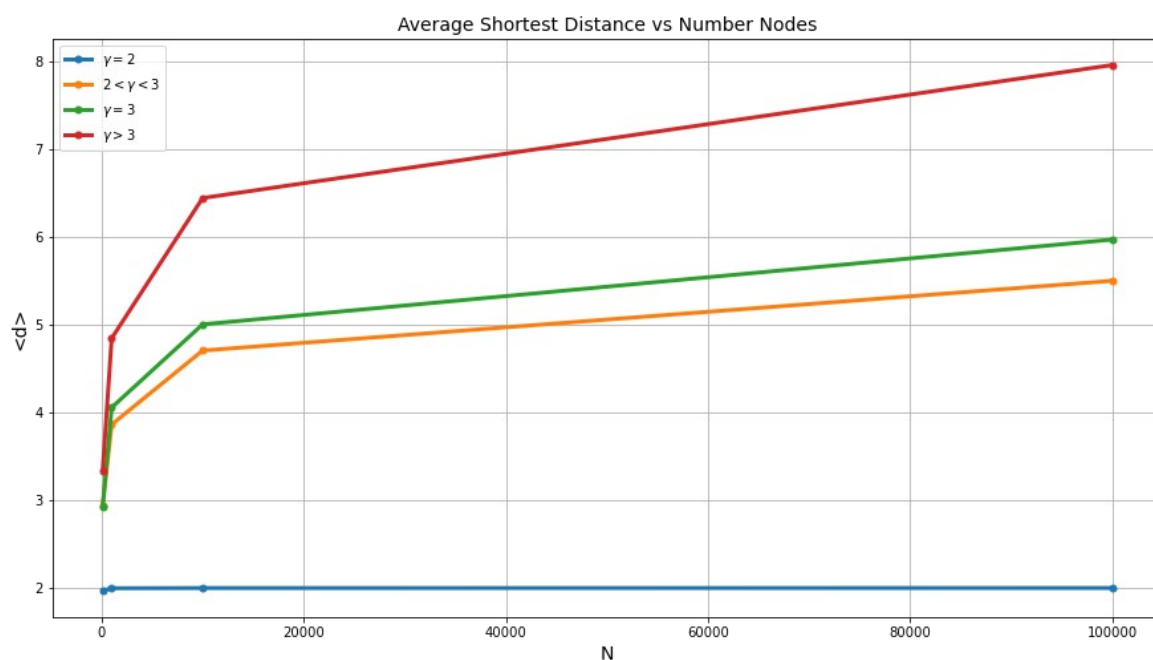
در ابتدا به تحلیل گراف های ساخته شده می پردازیم:

Graph Type	Calculated Gamma	Average Degree	Average Distance
$G=2, N=100$	2.18	3.93	1.97
$G=2, N=1000$	2.11	3.99	1.99
$G=2, N=10000$	2.08	3.99	1.99
$G=2, N=100000$	2.06	3.99	2.0
$G=2.5, N=100$	2.63	3.93	2.93
$G=2.5, N=1000$	2.56	3.99	3.86
$G=2.5, N=10000$	2.51	3.99	4.70
$G=2.5, N=100000$	2.46	3.99	5.50
$G=3, N=100$	2.81	3.93	2.92
$G=3, N=1000$	2.88	3.99	4.05
$G=3, N=10000$	2.9	3.99	5.00
$G=3, N=100000$	3.08	3.99	5.97
$G>3, N=100$	3.45	3.93	3.33
$G>3, N=1000$	3.91	3.99	4.84
$G>3, N=10000$	4.49	3.99	6.44
$G>3, N=100000$	4.98	3.99	7.96

جدول ۱-۱: مشخصات گراف های تولید شده با الگوریتم bianconi-barabasi

همانطور که از مقادیر جدول قابل تشخیص است مقدار متوسط درجه در تمامی گراف ها حدود ۴ است و گاما گراف تولید شده نیز حدوداً با مقدار گاما مدنظر در زمان ساخت برابر است.

نمودار متوسط فاصله گره ها بر اساس اندازه گراف نیز به شکل زیر می باشد:



شکل ۱-۱: نمودار متوسط فاصله بین گره ها بر حسب تعداد نود

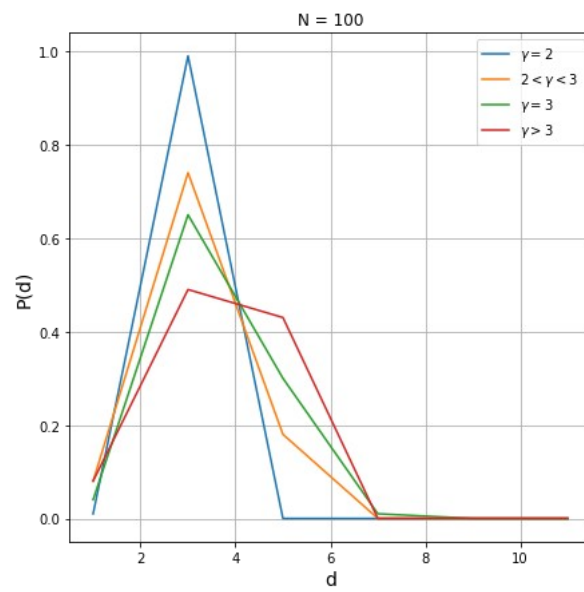
شکل نمودار از لحاظ ظاهری حدوداً شبیه به شکل نمودار داده شده در صورت سوال است حال برای بررسی دقیق تر رفتار گراف و بررسی مقادیر بدست آمده و مقادیر بدست آمده توسط فرمول نسبت این دو مقدار را در اندازه های مختلف گراف بدست می آوریم:

	$N=100$	$N=1000$	$N=10000$	$N=100000$
$G=2$	۱.۰۱	۱.۰۰	۱.۰۰	۱.۰
$G=2.5$	۱.۰۵	۱.۱۱	۱.۱۳	۱.۱۵
$G=3$	۱.۰۳	۰.۸۸	۰.۸۲	۰.۷۸
$G>3$	۱.۳۸	۱.۴۲	۱.۴۲	۱.۴۴

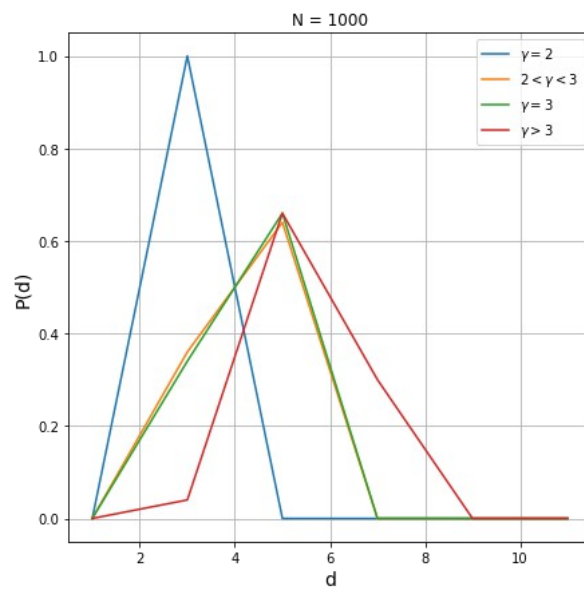
جدول ۱-۲: نسبت مقادیر پیش بینی شده توسط فرمول بر مقادیر به دست آمده از گراف ساخته شده

با توجه به مقادیر به دست آمده، مقدار واقعی متوسط فاصله با مقدار پیش بینی شده با بزرگتر شدن گراف حدوداً نسبتی حدوداً برابر با یک دارد، دلیل اختلاف اندک این دو مقدار را میتوان به ۱- کوچک بودن اندازه گراف ها به ۲- نمونه برداری از گراف جهت محاسبه متوسط فاصله نسبت داد.

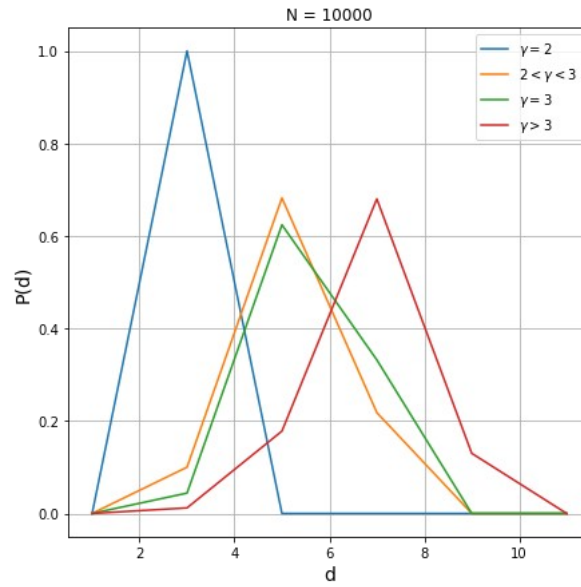
در نهایت به بررسی توزیع احتمالی فواصل گراف ها می پردازیم:



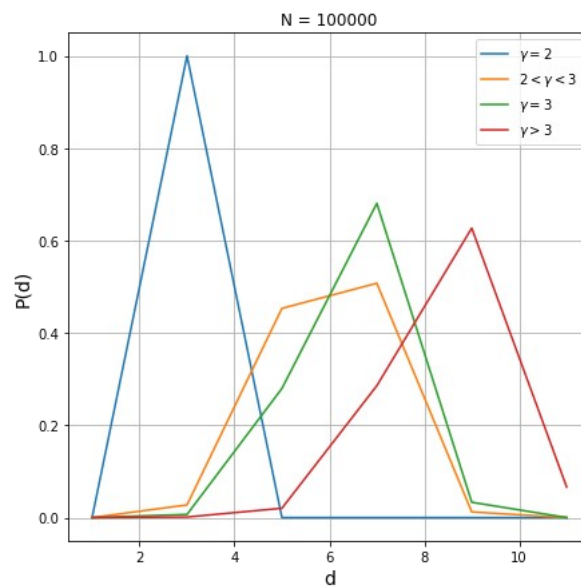
شکل ۲-۱: توزیع احتمالی گراف ها با اندازه ۱۰۰ گره



شکل ۲-۳: توزیع احتمالی گراف ها با اندازه ۱۰۰۰ گره



شکل ۴-۱: توزیع احتمالی گراف ها با اندازه ۱۰۰۰۰ گره



شکل ۵-۱: توزیع احتمالی گراف ها با اندازه ۱۰۰۰۰۰ گره

با توجه به شکل توزیع ها مشخص است که مانند شکل صورت سوال با بزرگتر شدن اندازه گراف ها میانگین توزیع احتمالی آنها نیز از یکدیگر فاصله میگیرد.