



تمرین دوم



داده کاوی - پاییز ۱۳۹۹

گزارش کار

دانشکده علوم مهندسی

نام و نام خانوادگی:
مینو احمدی

استاد:
دکتر علی فهیم

مقدمه:



این تمرین شامل دو قسمت می باشد:

بخش اول در فایل `Ecoli1.py` پیاده سازی شده است.
در این بخش به محاسبه ۳ چیز پرداختیم:

Diameter,
Degree distribution
, Clustering coefficient

بخش دوم در فایل `Ecoli2.py` پیاده سازی شده است.
محاسبه بردار ویژه با روش power iteration

قسمت اول:

نحوه اجرای برنامه

```
python Ecoli1.py Ecoli.txt
```

خروجی برنامه:

```
plot_degree_distribution(G)
```

```
plot_degree_probability(G)
```

```
print(clustering_coeff(G))
```

:Diameter

با توجه به تعریف قطر گراف که داریم: فاصله بین دو رأس را طول کوتاه‌ترین مسیر ممکن بین آن دو تعریف می‌کنیم. حال به ازای هر رأسی اگر اسم بیشترین فاصله آن تا سایر رأس‌ها را خروج از مرکز آن رأس بگیریم، قطر گراف برابر بیشترین خروج از مرکز است. توجه کنید که قطر همیشه برابر طولانی‌ترین مسیر نیست.

چون استفاده از تابع `single_source_shortest_path_length` بلامانع بود (در متن پروژه ذکر شده) با این تابع طول کوتاه‌ترین مسیرها به هر رأس را پیدا کرده سپس بین این فاصله‌ها ماکزیمم را پیدا کرده و برگردانیم که همان قطر گراف است.

:Degree distribution

هر رأس در یک گراف درجه خاص خود را دارد. این درجات توزیع خاصی دارند این توزیع میتواند از مدل‌های خاصی پیروی کند یا رندوم باشد و هماهنگی ویژه‌ای نداشته باشد.

در تابع `plot_degree_distribution`

درجات هر رأس را محاسبه کرده سپس تعداد ریوسی که از هر درجه داشتیم را حساب کردیم و نمودار تعداد دفعات تکرار درجات را بر حسب عدد درجه رأس رسم میکنیم.

در تابع `plot_degree_probability`

درجات و احتمال حضور هر درجه را محاسبه کرده سپس از هر دو لگاریتم گرفته و بعد نمودار این دو بر حسب هم را رسم میکنیم و سپس سعی میکنیم تا خطی را با رگرسیون به شکل برازش دهیم .

در واقع این خط نمایان گر رابطه بین این دو متغیر است و برای دیتا های بعدی میتواند نشان دهد که نقاط بعدی حدودا در کجا می افتد.

برای برازش بهتر داده ها میتوان داده های پرت را حذف کرد یا داده ها را هرس کرد.

Clustering coefficient یا ضریب خوشه بندی:

ضریب خوشه بندی معیاری است که درجه گره ها در یک گراف تمایل به ایجاد خوشه باهم دارند را اندازه گیری میکند .

با استفاده از تابع `adj` میتوان رابطه بین دو راس و یال های کشیده شده از یک راس به راس دیگر را پیدا کرد . مثلا برای یک راس خاص چه راس هایی همسایه ان هستند و با چه یال هایی به ان متصل هستند.
با استفاده از این تابع و تعریف اصلی خوشه بندی عدد خوشه را محاسبه میکنیم .

قسمت دوم :

ورودی:

`python Ecoli2.py Ecoli-directed.txt`

خروجی:

`Print(first_ten)`

در واقع ۱۰ ژن برتر با مقادیر ویژه ی ان ها را چاپ میکند.

در این بخش می خواهیم با روش `power iteration` بردار و مقدار ویژه را محاسبه کنیم .

$$M = (1-d)*M + d*np.ones((n,n))/n$$

$$V = np.ones(n)/n$$

را تشکیل داده سپس به صورت هر بار M را در V ضرب میکنیم و اختلاف ان را با V اولیه محاسبه میکنیم هر گاه این اختلاف از میزان مورد نظر ما کوچک تر بود در واقع ما به جواب رسیده ایم.