Maria Kagkeli
Maria Regina Lily
Mihai Verzan

1 Degree Correlation Coefficient

For all the computations, please see `Problem 6-1.ipynb`

1.1 Degree Correlation Matrix

$$E = \begin{bmatrix} 0 & 0 & 1/8 \\ 0 & 1/4 & 1/4 \\ 1/8 & 1/4 & 0 \end{bmatrix}$$

1.2 Probabilities q_k

$$\begin{aligned} q_1 &= 0 + 0 + 1/8 = 1/8 \\ q_2 &= 0 + 1/4 + 1/4 = 1/2 \\ q_3 &= 1/8 + 1/4 + 0 = 3/8 \end{aligned}$$

1.3 Degree Correlation Coefficient

$r = -0.7142857142857143$

Based on r, the given network is disassortative ($r < 0$)

2 Degree Correlations in Random Graphs

2.1

$$p(a_{ij}=1)=\frac{L}{\frac{N(N-1)}{2}}=\frac{2L}{N(N-1)}$$

2.2

If link a_{xy} exists, then $p(a_{ij})(i,j \neq x,y)$ goes down, since we know that one link is already present and the final number of links is fixed. Similarly, if the link doesn't exist, the probability of all other links goes up.

$$\begin{aligned} p(a_{ij}=1|a_{xy}=1) &= \frac{L-1}{\frac{N(N-1)}{2}-1} \\ &= 2\frac{L-1}{N(N-1)-2} \\ &= \frac{2L-2}{N^2-N-2} \end{aligned}$$

$$\begin{aligned} p(a_{ij}=1|a_{xy}=0) &= \frac{L}{\frac{N(N-1)}{2}-1} \\ &= 2\frac{L}{N(N-1)-2} \\ &= \frac{2L}{N^2-N-2} \end{aligned}$$

2.3

$$\begin{aligned} \frac{P\left(a_{ij}=1|a_{xy}=1\right)}{P\left(a_{ij}=0\right)} &= \frac{L-1}{\frac{N(N-1)}{2}-1}\cdot\frac{\frac{N(N-1)}{2}}{L} \\ &= \frac{L-1}{\frac{N(N-1)-2}{2}}\cdot\frac{N(N-1)}{2L} \\ &= \frac{N(L-1)(N-1)}{\frac{2LN(N-1)-2}{2}} \\ &= \frac{N(L-1)(N-1)}{LN(N-1)-2} \end{aligned}$$

$$\lim_{N\to\infty}\frac{N(L-1)(N-1)}{LN(N-1)-2}=\frac{L-1}{L}$$

$$\begin{aligned}
\frac{P\left(a_{ij}=1 | a_{xy}=0\right)}{P\left(a_{ij}=0\right)} &= \frac{L}{\frac{N(N-1)}{2}-1} \cdot \frac{\frac{N(N-1)}{2}}{L} \\
&= \frac{2}{N(N-1)-2} \cdot \frac{N(N-1)}{2} \\
&= \frac{N(N-1)}{N(N-1)-2}
\end{aligned}$$

$$\lim_{N \to \infty} \frac{N(N-1)}{N(N-1)-2} = 1$$

2.4

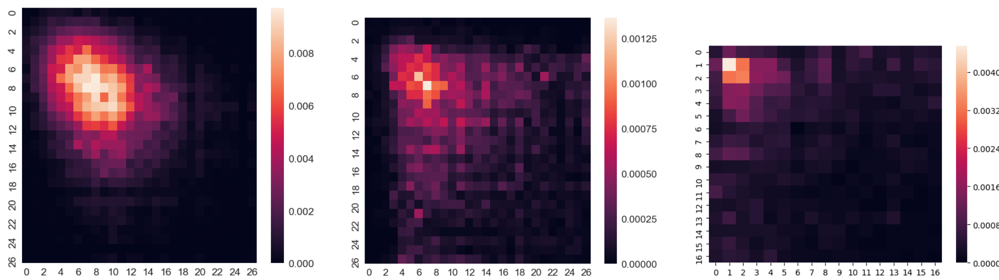
For $G(N,p)$ model, $r'_0 = r'_1 = 1$, in other words, the conditional probabilities don't change and are always the same as $p(a_{ij}=1)$. This is the because the probability p that an edge exists in the $G(N,p)$ model is fixed, unlike in the $G(N,L)$ model.

2.5

3 Degree Correlations and Assortativity

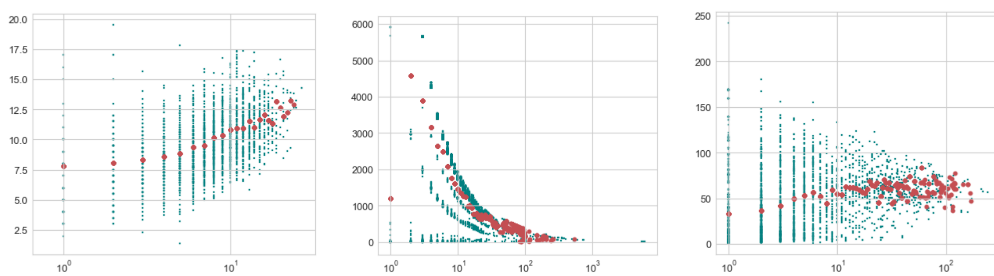
For the source code, please see `Problem 6-3.ipynb`

	# Nodes	# Edges	Degree Correlation Coefficient
Adolescent heatlh	2530	10026	0.2522525430884104
JDK dependency	6424	53587	-0.2229706960177308
OpenFlights	2911	15593	0.04686705925649979



Adolescent health JDK dependency (zoomed in) OpenFlights (zoomed in)

Figure 1: Degree correlation heatmaps



Adolescent health JDK dependency OpenFlights

Figure 2: Scatter plots

The degree correlation coefficient gives a clear definition of a network's assortativity, and the scatter plots give a nice visualization of that. Both the degree correlation and the scatter plots lead to the same conclusion regarding assortativity (coefficient $\approx$ slope of the scatter plot).

We found that the heatmaps are the most difficult to use. It is good for visualizing the degree correlation matrix, but it's hard to tell the assortativity from a heatmap. Theoretically, should the network be assortative, then the highest values of the matrix, or the lightest parts of the heatmap, would be at the diagonal. Practically, this might be hard to see unless the assortativity coefficient of the network is equal to or very close to 1. Moreover, most nodes in a large network has a rather small degree, which means that the only interesting part of the heatmap is only the small part on the upper left.