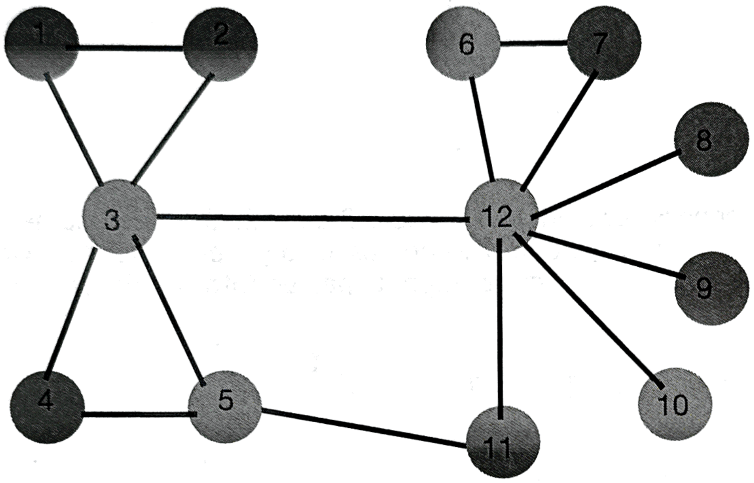
Name(s): Michael Belmear

Homework 3: CSCI 347: Data Mining

Show your work. Include any code snippets you used to generate an answer, using comments in the code to clearly indicate which problem corresponds to which code.

Consider the following graph:



1. [3 points] Without using networkx or other graph analysis packages (though you may use them to check your answer), find the closeness centrality of vertices 3 and 12.

$$cc(x_3) = \frac{1}{\sum_{j=1}^{n} d(x_3, x_j)}$$

$$= \frac{1}{1+1+0+1+1+2+2+2+2+2+2+1}$$

$$= \frac{1}{17} = 0.05882...$$

$$cc(x_{12}) = \frac{1}{\sum_{j=1}^{n} d(x_{12}, x_j)}$$

$$= \frac{1}{2+2+1+2+2+1+1+1+1+1+1+0}$$

$$= \frac{1}{15} = 0.06666...$$

2. [3 points] Without using networkx or other graph analysis packages (though you may use them to check your answer), find the eccentricity of vertices 3, 12, and 11.

$$e(x_i) = \max_{i} \{d(x_i, x_j)\}$$

$$e(x_3) = 2$$

$$e(x_{12}) = 2$$

$$e(x_{11}) = 3$$

3. [3 points] Find the betweenness centrality of vertices 3 and 12. You may use networkx or other graph analysis packages, but include the code used to generate your answer in your submission (either as code in D2L, or as code snippets pasted into the document turned into Gradescope).

* see code turned in *

The betweeness centrality of vertices 3 and 12 is 27 and 40.5.

4. [3 points] **Using networkx**, find the prestige/eigenvector centrality of vertices 3 and 12. Include the code used to generate the answer.

see code turned in

The prestige centrality of vertices 3 and 12 is 0,4653 and 0.5310.

5. [3 points] Find μ_L, the average length of the shortest path between two vertices in this graph.

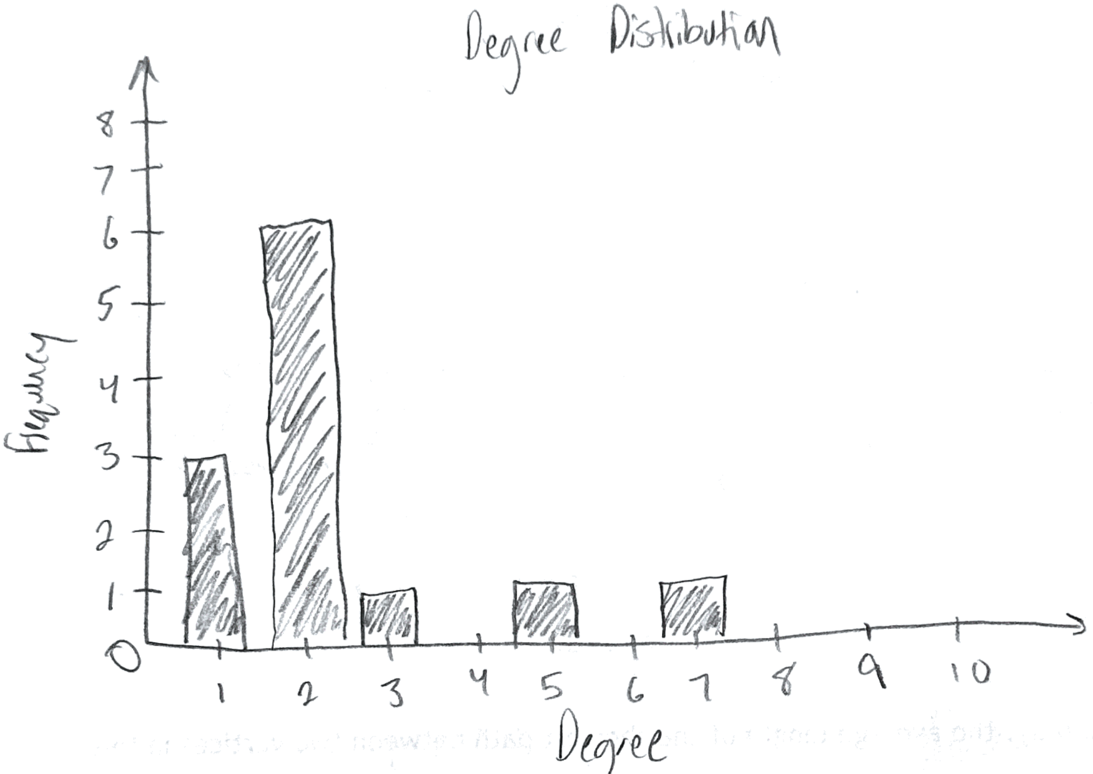
see code turned in

The average length of the shortest path in the graph is 2.106...

6. [3 points] Use Python to create a plot of the degree distribution of this graph. Include the code used to generate the plot as well as the plot in your submission.

run python code turned in for betts view

* see turned in screen shot



7. [3 points] Without using networkx or other graph analysis packages (though you may use them to check your answer), find the clustering coefficient of vertex 3.

$$\frac{m_i}{\binom{n_i}{2}} = \frac{\text{\# of edges among neighbors}}{\text{\# of possible edges among neighbors of } x_i}$$

$$\frac{m_3}{\binom{n_3}{2}} = \frac{2}{10}$$

$$= \frac{1}{5} \text{ or } 0.2$$

8. [3 points] Without using networkx or other graph analysis packages (though you may use them to check your answer), find the clustering coefficient of the graph.

x_i	$\frac{M_i}{\binom{n_i}{2}}$
x_1	1
x_2	1
x_3	$\frac{1}{5}$
x_4	1
x_5	$\frac{1}{3}$
x_6	1
x_7	1
x_8	0
x_9	0
x_{10}	0
x_{11}	0
x_{12}	$\frac{1}{21}$

$$CC = \frac{1}{n} \sum_{i=1}^{n} x_i$$

$$= \frac{1}{12}\left(1+1+\frac{1}{5}+1+\frac{1}{3}+1+1+0+0+0+0+\frac{1}{21}\right)$$

$$= \frac{1}{12} \times \frac{586}{105}$$

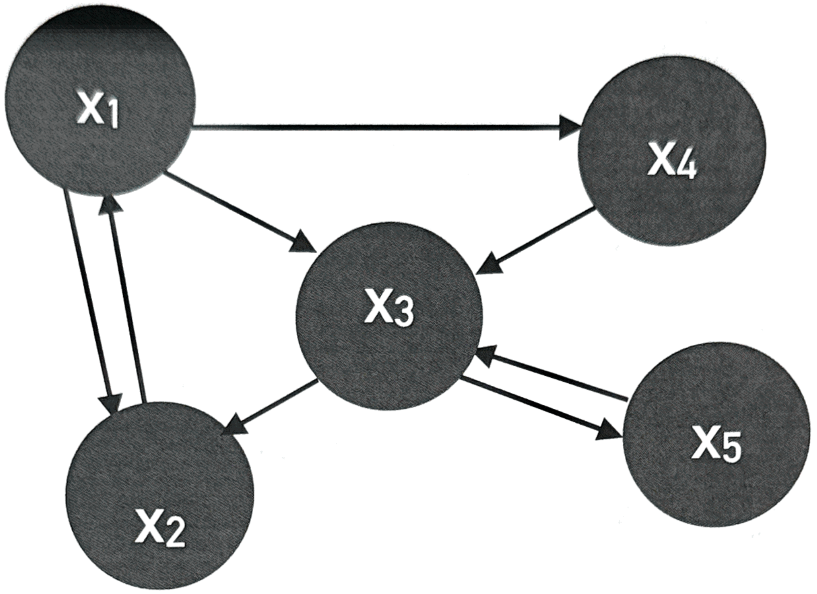
$$= \frac{293}{630}$$

$$= 0.46507$$

9. [3 points] Use networkx to create an (undirected) Erdos-Renyi random graph with parameters n=200 and p=0.1. Create a visualization of the graph, with vertex sizes dependent on betweenness centrality (the higher the betweenness centrality, the greater the size) and node color dependent on degree. Include both the code to generate the plots, as well as the plots themselves, in your submission.

*see code and turned in screen shot

10. [5 points Extra Credit]: Consider the graph below. When using Power Iteration to find the prestige (aka eigenvector) centralities of each node in the graph, assuming that the initial prestige vector is the vector consisting of all 1's, show the the prestige vector after each of the first 3 iterations. Also show the ranking of the nodes in this graph based on their prestige values, from largest to smallest (using the converged prestige vector). If there are any ties, write down which nodes are tied for which positions in the ranking.


X1
X4
X3
X5
X2