hw01

Charles Valentine

October 7, 2016 6PM

Question #1

Let $\mathbf{A}$ be the $N \times N$ adjacency matrix of an undirected unweighted network, without self-loops, of size N. Let $\mathbf{1}$ be a column vector of N elements all equal to 1.

$$\mathbf{A} = \begin{bmatrix} x_{ij} & \dots & x_i \\ \vdots & \ddots & \vdots \\ x_{Nj} & \dots & x_{NN} \end{bmatrix} \text{where} \quad \text{diag}(\mathbf{A}) = [0_{ij}, ..., 0_{NN}] \tag{1}$$

$$\mathbf{1} = \begin{bmatrix} 1_{ij} \\ \vdots \\ 1_{NN} \end{bmatrix} \tag{2}$$

a) The vector $\vec{k}$ whose elements are the degrees k_i of the nodes $i = 1, 2, ..., N$ is defined as:

$$\vec{k} = \mathbf{1}^T \bullet \mathbf{A} \tag{3}$$

b) The total number of L links in the network is defined as:

$$L = \frac{\vec{k} \bullet \mathbf{1}}{2} \tag{4}$$

c) The matrix $\mathbf{N}$ whose elements $\mathbf{N}_{ij}$ is equal to the number of common neighbors of nodes i and j is defined as:

$$\mathbf{N} = \mathbf{A} \bullet \mathbf{A} \tag{5}$$

d) The number of triangles T in the network is defined as:

$$T = \frac{\text{trace}(\mathbf{A}^3)}{6} \tag{6}$$

e) If we square our adjacency matrix $\mathbf{A}$ N times then any row i will tell us the nodes that are connected to that node i. If there exists a 0 in any row of $\mathbf{A}^k$ where $k \geq N$ then the network is not connected. Formally, a network is connected if this condition is true:

$$\mathbf{A}^k_{i,j} \neq 0 \quad \text{for} \quad i = 1, ..., N \quad \text{and} \quad j = 1, ..., N \quad \text{where} \quad k \geq N \tag{7}$$

Question #2

In a random graph with $N = 5000$ nodes, the linking probability $p = 10^{-4}$.

a) The expected total number of links is defined as:

$$\langle L \rangle = p\frac{N(N-1)}{2} = 10^{-4}\frac{5000(5000-1)}{2} \approx 1250 \tag{8}$$

b) To categorize this network into the regimes of network evolution we must first define $\langle k \rangle$ as:

$$\langle k \rangle = \frac{2\langle L \rangle}{N} = p(N-1) \approx 0.5 \tag{9}$$

The average node degree $\langle k \rangle < 1$ so this network is in the subcritical regime. This network has no giant component and the clusters are represented as trees and branches.

c) We must solve for p_c in the following equation to find the linking probability p_c for a network with 5000 nodes at the critical point.

$$p_c(5000-1) = 1, \quad p_c \approx 2 \cdot 10^{-4} \tag{10}$$

d) We must solve for N' in the following equation to find the minimum number of nodes N' that, with very high probability, a random network will have only one component. We will then choose a value for N' for subsequent calculation.

$$p(N'-1) > 1 \quad \text{where} \quad p = 10^{-4}, \quad N' \geq 10,002 \quad \text{let} \quad N' = 70,000 \tag{11}$$

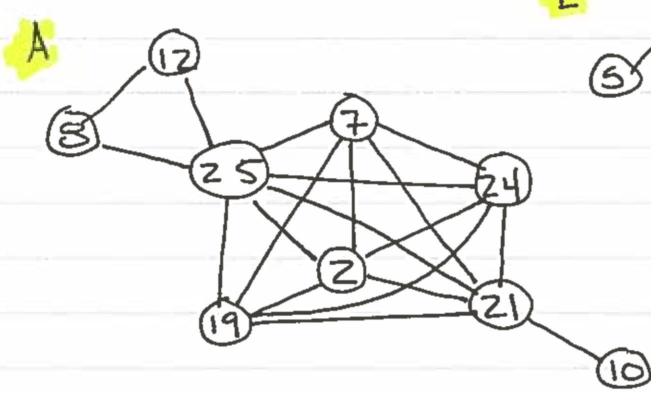
e) The average degree $\langle k' \rangle$ and average distance $\langle d \rangle$ of the network in 2d is defined as:

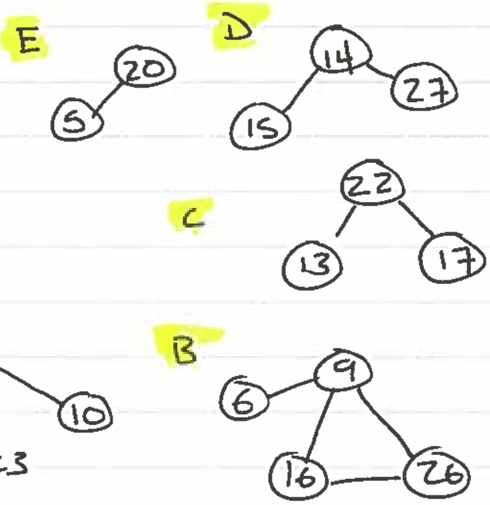
$$\langle k' \rangle = 10^{-4}(70,000-1) \approx 7 \tag{12}$$

$$\langle d \rangle \approx \frac{\ln(70,000)}{\ln(7)} \approx 5.73 \tag{13}$$

f) The formula for the degree distribution $P(k)$ of this network is approximated using the Poisson distribution:

$$P(k) = e^{-70,000}\frac{70,000^k}{k!} \tag{14}$$

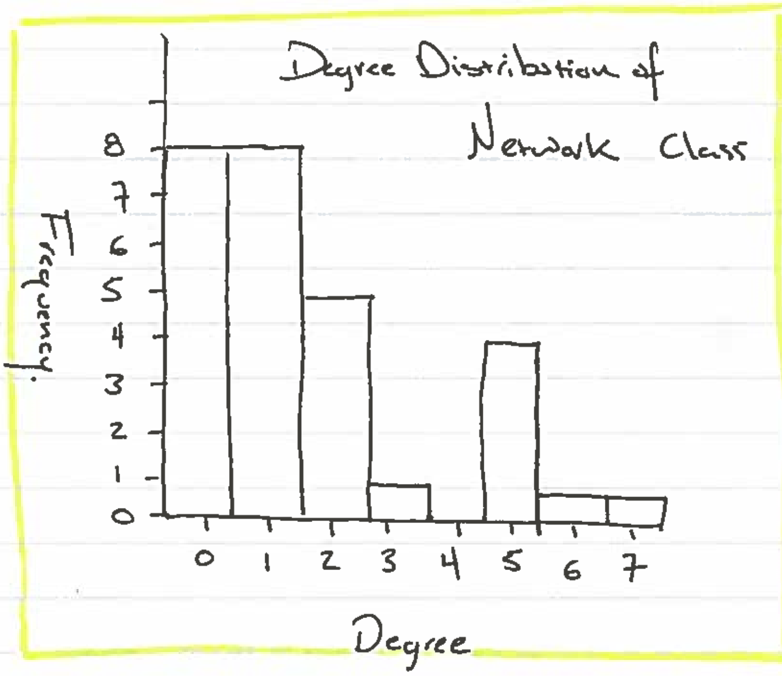




Node	k
2	5
5	1
6	1
7	5
9	3
10	1
12	2
13	1
14	2
15	1
16	2
17	1
19	5
20	1
21	6
22	2
24	5
25	7
26	2
27	1

nodes: 0, 1, 3, 4, 8, 11, 18, 23 have K = 0

K	freq.
0	8
1	8
2	5
3	1
4	∅
5	4
6	1
7	1



node$_i$	C_i
9	1/3
14	0
19	1
21	2/3
25	11/21

← using the formula .... $C_i = \dfrac{2L_i}{k_i(k_i-1)}$

node$_i$	L_i	k_i	C_i
9	1	3	$1/3 = 0.\overline{3}$
14	0	2	0
19	10	5	1
21	10	6	$2/3 = 0.\overline{6}$
25	11	7	$11/21 = 0.5238$

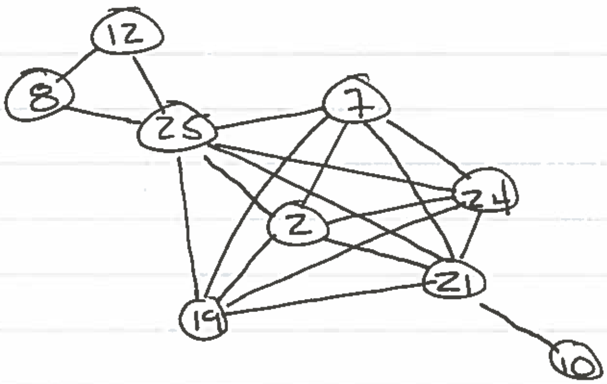
C. The network has five components $(N>1)$. There sizes are:

Component$_i$	N_i
A	9
B	4
C	3
D	3
E	2

There are also 8 more components (not drawn) of $N=1$ for a total of

12 components

d. A single giant component is defined under the evolution of a random network. A giant componet generally emerges in a random network a $\langle K \rangle > 1$. We cannot, however claim to have a giant component in our network class graph because the process by which students ~~in~~ enroll in the course with their peers ~~in~~ is not a random process. We see evidence of this by having a $\langle K \rangle >> 1$ at $\langle K \rangle = 1.93$ while observing many components of similar size that are NOT connected by definition.



Notes: Shortest Distance (path length) from node to node.

	2	7	8	10	12	19	21	24	25
2	0	1	2	2	2	1	1	1	1
7	1	0	2	2	2	1	1	1	1
8	2	2	0	3	1	2	2	2	1
10	2	2	3	0	3	2	1	2	2
12	2	2	1	3	0	2	2	2	1
19	1	1	2	2	2	0	1	1	1
21	1	1	2	1	2	1	0	1	1
24	1	1	2	2	2	1	1	0	1
25	1	1	1	2	1	1	1	1	0

Diameter is the longest shortest path at 3

Question #4

a) The component size distribution $P(s)$ for a random network with average degree $\langle k \rangle_{int} < 1$ can be modeled using the Poisson distribution:

$$P(s) \approx e^{-\langle k \rangle_{int} s} \frac{(s\langle k \rangle_{int})^{s-1}}{s!} \tag{15}$$

b) To determine the minimal fraction of extroverts $f = {}^{N_{ext}}/_{N}$ needed to connect the network we must first consider what condition needs to be met for a network to be connected. If all nodes have an average degree greater than $\ln N$ then there is a high probability that all nodes are connected to a giant component. This condition marks the point in which a random network enters into a connected regime composed of only the the giant component ($N_G = N$).

$$\langle k \rangle = \frac{2(L_{int} + L_{ext})}{N} > \ln N \quad \text{where} \quad L_{int} = \frac{N_{int}k_{int}}{2} \quad \text{and} \quad L_{ext} = \frac{N_{ext}k_{ext}}{2} \tag{16}$$

$$\frac{2(\frac{N_{int}k_{int}}{2} + \frac{N_{ext}k_{ext}}{2})}{N} = \frac{N_{int}k_{int} + N_{ext}k_{ext}}{N} > 1 \tag{17}$$

$$\frac{(N - N_{ext})k_{int} + N_{ext}k_{ext}}{N} > \ln N \quad \text{where} \quad N_{int} = N - N_{ext} \tag{18}$$

$$\frac{Nk_{int} - N_{ext}k_{int} + N_{ext}k_{ext}}{N} > \ln N \tag{19}$$

$$k_{int} + f(k_{ext} - k_{int}) > \ln N \tag{20}$$

$$f > \frac{\ln N - k_{int}}{k_{ext} - k_{int}} \tag{21}$$

c) The distance between two introvert nodes $d_{i,i}$ can approximately be found by finding the diamater of the network as shown in equation (23). The average distance between an extrovert and an introvert node $d_{i,e}$ can be modeled by finding the average distance of the network as a whole as shown in equation (24). This network does show the Small World property as $\langle k \rangle > \ln N > 1$.

$$N(d_{max}) \approx \frac{\langle k \rangle^{d_{max}+1} - 1}{\langle k \rangle - 1} \approx N \tag{22}$$

$$d_{max} \approx \frac{\ln N}{\ln \langle k \rangle} \approx d_{i,i} \quad \text{where} \quad \langle k \rangle >> 1 \tag{23}$$

$$\langle d \rangle = \frac{1}{N(N-1)} \times \sum_{\substack{i,j=1,N \\ i \neq j}} d_{i,j} \tag{24}$$

Question #5

In 2.12.5:

a) The adjacency matrix $\mathbf{A}$ for Figure 2.21 is defined below. This matrix is in block-diagonal form because no node in set U can link with any other node in set U and likewise for set V. This corresponds with the definition of a bipartite graph with two projections.

$$\mathbf{A} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{25}$$

b) The adjacency matrices $\mathbf{A}_U$ and $\mathbf{A}_V$ for the projects U and V are described as:

$$\mathbf{A}_U = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix} \tag{26}$$

$$\mathbf{A}_V = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix} \tag{27}$$

c) The average degree for the purple nodes $\langle k \rangle_{purple}$ in the bipartite network is $1.\bar{6}$. The average degree for the green nodes in the bipartite network $\langle k \rangle_{green}$ is 2.

d) The average degree for the projection U $\langle k \rangle_U$ is $2.\bar{3}$. The average degree for projection V $\langle k \rangle_V$ is 2. It is not surprising that these values are different than in 5c. To assert the average degree of a node in the bipartite graph claims nothing about it's average degree in it's own projection. An example of this can be seen where a node in one projection with a single link to a node in a second projection would have a degree of 1 whereas in the first node's own projection it could be connected to hundreds of other nodes.

In 2.12.6:

d) An expression connecting N_1, N_2 and the average degrees for two sets in a bipartite network, $\langle k \rangle_1$, $\langle k \rangle_2$ is defined as:

$$\frac{L}{N_1} = \langle k \rangle_1 \quad \text{and} \quad \frac{L}{N_2} = \langle k \rangle_2 \quad \text{so} \quad N_1 \langle k \rangle_1 = N_2 \langle k \rangle_2 \tag{28}$$