20F-MATH168-1 HW 3

PRATYUSHA MAJUMDER

TOTAL POINTS

59 / 61

QUESTION 1

1 Reading 1 / 1

✓ + 1 pts Complete

QUESTION 2

2 Problem 6.7 10 / 10

✓ + 3 pts Part a

✓ + 3 pts Part b

✓ + 4 pts Part c

+ 2 pts Part b partial credit (correct answer, no work)

+ 2 pts Part c partial credit

QUESTION 3

3 Problem 6.11 10 / 10

✓ + 2 pts Part a

✓ + 2 pts Part b

✓ + 2 pts Part c

✓ + 2 pts Part d

✓ + 2 pts Part e

+ 1 pts Part e partial credit

QUESTION 4

4 Problem 10.2 8 / 10

✓ + 2 pts part a

✓ + 2 pts part b

✓ + 2 pts part c

✓ + 2 pts part d

+ 2 pts part e

+ 1 pts part a partial credit

+ 1 pts part b partial credit

+ 1 pts part c partial credit

+ 1 pts part e partial credit

Part e- What is the meaning of the "unphysical" values? you can get W> 1 when k = log(P)/log(a) < 1, but this is "unphysical" because k is a degree, hence an integer.

QUESTION 5

5 Clustering coefficients 10 / 10

✓ + 10 pts Complete (everyone has a different network so I am grading on completion)

QUESTION 6

6 ER graphs 10 / 10

✓ + 5 pts Part a

✓ + 5 pts Part b

+ 3 pts Part a partial credit

QUESTION 7

7 Watts-Strogatz networks 10 / 10

✓ + 5 pts Part a

✓ + 5 pts Part b

+ 0 pts Missing

Math 168, Networks, UCLA, Fall 2020
Problem Sheet 3

(submit to CCLE (Gradescope) by 27 October 2020 at 6:00 pm)

1. *Reading.* Read Chapter 10 and Sections 13.1–13.2 of Newman's book.
2. Do Problem 6.7 of Newman's book.
3. Do Problem 6.11 of Newman's book.
4. Do Problem 10.2 in Newman's book.
5. *Clustering coefficients.* Draw a small network in which the global clustering coefficient and mean local clustering coefficient have different values. Write down the adjacency matrix for this network.
6. *More on ER graphs.*
   (a) Calculate the global clustering coefficient C of an ER graph (in expectation over the ensemble).
   (b) Is the ER network ensemble a good model for a social network? Why or why not? How can it be adjusted to produce a better model?
7. *Simulating Watts–Strogatz networks or a variant of it.*
   (a) Using numerical simulations of either Watts–Strogatz networks or Newman–Watts networks (or both) for many values of p, illustrate via a plot like the one I showed in lectures the transition between large worlds and small worlds. Your vertical axis should include both a clustering coefficient and mean geodesic path length. Try this for different numbers of nodes N. What do you observe as N becomes larger? For a given value of N, what does the plot look like if you show the results for a single realization of a network for each value of p? What does it look like if the vertical axis has sample means over some reasonably large number of realizations of a WS and/or NW ensemble for each value of p?

   Note: There is numerous code online for generating WS networks or variants thereof. Here is one of them: `http://uk.mathworks.com/help/matlab/math/build-watts-strogatz-small-world-graph-model.html`. The Brain Connectivity Toolbox is another one that you may wish to use. If you use existing code, explicitly cite the one that you use.

   (b) Is the Watts–Strogatz model a good model for a social network? Why or why not?

1 Reading 1 / 1

✓ **+ 1 pts** **Complete**

Q1. Ex 6.7

6.7 Consider an acyclic directed network of n nodes, labeled $i = 1 \ldots n$, and suppose that the labels are assigned in the manner of Fig. 6.3 on page 111, such that all edges run from nodes with higher labels to nodes with lower.

a) Write down an expression for the total number of ingoing edges at nodes $1 \ldots r$ and another for the total number outgoing outgoing at nodes $1 \ldots r$, in terms of the in- and out-degrees k_i^{in} and k_i^{out} of the nodes.

b) Hence find an expression for the total number of edges running to nodes $1 \ldots r$ from nodes $r+1 \ldots n$.

c) Hence or otherwise show that in any acyclic network the in- and out-degrees must satisfy

$$k_r^{\text{in}} \leq \sum_{r+1}^{n} (k_i^{\text{out}} - k_i^{\text{in}}), \qquad k_{r+1}^{\text{out}} \leq \sum_{i=1}^{r} (k_i^{\text{in}} - k_i^{\text{out}}),$$

for all r.

a) • Let k_i^{in} be the total # of incoming edges for node i.

• Let k_i^{out} be the total # of outgoing edges for node i.

∴ Total # of incoming edges for nodes 1, 2, ..., r :

$$\sum_{i=1}^{r} k_i^{(in)}$$

Total # of outgoing edges from nodes 1, 2, ..., r:

$$\sum_{i=1}^{r} k_i^{(out)}$$

b) Total # of edges running from nodes 1, ..., r from nodes r+1, ..., n.

is the total number of outgoing edges from nodes r+1, ..., n that are incoming edges for nodes 1, ..., r.

For a graph with nodes 1, ..., n, and looking at nodes 1, ..., r, the incoming edges for nodes 1, ..., r would be outgoing from either nodes (1, ..., r) or nodes (r+1, ..., n).

∴ excluding outgoing edges from nodes (1, ..., r) would give you incoming edges for nodes (1, ..., r) that are outgoing from nodes (r+1, ..., n)

$$\therefore \sum_{i=1}^{r} (k_i^{in} - k_i^{out})$$

(c) · Let r be a node in the directed acyclic network such that $r = 1, \dots, n$.

· Let k_r^{in} be the # incoming edges into node r, from nodes $\{1, \dots, n\}$

· Let k_r^{out} be the # outgoing edges from node r, to nodes $\{1, \dots, n\}$

· $\sum_{i=r+1}^{n} (k_i^{out} - k_i^{in})$ is the total # of outgoing edges from nodes $\{r+1, \dots, n\}$ that connect to nodes $\{1, \dots, r\}$.

· By hypothesis, graph is acyclic, directed and has edges running from a higher label node to a lower label node.

Proof: (by contradiction)

· Suppose $k_r^{in} > \sum_{i=r+1}^{n} (k_i^{out} - k_i^{in})$ for a node r such that $r \in [1, \dots, n]$

· $\Rightarrow \exists$ atleast one incoming edge into node r that is not coming from nodes $\{r+1, \dots, n\}$.

$\Rightarrow \exists$ atleast one incoming edge into node r that is coming from nodes $\{1, \dots, r\}$

Case 1:

· Suppose this edge comes into node r from node r

$\Rightarrow$ a cycle is created, which is a contradiction to the assumption that the graph is acyclic.

Case 2: Suppose this edge comes from nodes $\{1, \dots, r-1\}$

$\Rightarrow$ an incoming edge to node r comes from a node with label lower than r and this is a contradiction to the assumption that in the graph, edges run from nodes with a higher label to nodes with lower label.

$\therefore$ existence of atleast one incoming edge into node r from nodes $\{1, \dots, r\}$ is a contradiction to the supposition

$$k_r^{in} > \sum_{i=r+1}^{n} (k_i^{out} - k_i^{in})$$

$$\therefore \boxed{k_r^{in} \leq \sum_{i=r+1}^{n} (k_i^{out} - k_i^{in})} \quad \forall r \in [1, \dots, n]$$

$\square$

Proof (2) by contradiction

- k_{r+1}^{out} is the # of outgoing edges from node r+1 , $\forall r \in [1, \dots, n-1]$
- $\sum_{i=1}^{r} (k_i^{in} - k_i^{out})$ is the total # of incoming edges into $\{1, \dots, r\}$ that come from $\{r+1, \dots, n\}$
- Suppose $k_{r+1}^{out} > \sum_{i=1}^{r} (k_i^{in} - k_i^{out})$, $\forall r \in [1, \dots, n]$

$\Rightarrow$ $\exists$ atleast one outgoing edge from node (r+1) that is not incoming into nodes $\{1, \dots, r\}$

$\Rightarrow$ $\exists$ atleast one outgoing edge from node (r+1) that is incoming into nodes $\{r+1, \dots, n\}$

Case 1: that edge is incoming into node $(r+1) \in \{r+1, \dots, n\}$
$\Rightarrow$ a cycle is formed starting at node (r+1) and this is a contradiction to the hypothesis that the graph is acyclic.

Case 2: That edge is incoming into nodes $\{r+2, \dots, n\}$
$\Rightarrow$ the edge is coming from a lower label node, (r+1), to a higher label node, in $\{r+2, \dots, n\}$ and this is a contradiction to the hypothesis that the graph only has edges going from nodes with a higher label to nodes with a lower label.
↯

$\therefore$ existence of atleast one outgoing edge from node(r+1) to nodes $\{1, \dots, r\}$ is a contradiction to the supposition

$$k_{r+1}^{out} > \sum_{i=1}^{r} (k_i^{in} - k_i^{out})$$

$$\therefore \boxed{k_{r+1}^{out} \leq \sum_{i=1}^{r} (k_i^{in} - k_i^{out})} \quad \forall r \in [1, \dots, n]$$

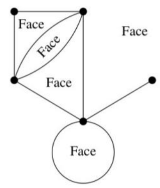
$\square$

2 Problem 6.7 **10 / 10**

- ✓ **+ 3 pts** **Part a**
- ✓ **+ 3 pts** **Part b**
- ✓ **+ 4 pts** **Part c**
- **+ 2 pts** Part b partial credit (correct answer, no work)
- **+ 2 pts** Part c partial credit

Q2, Ex 6.11

6.11 Consider a connected planar network with n nodes and m edges. Let f be the number of "faces" of the network, i.e., areas bounded by edges when the network is drawn in planar form. The "outside" of the network, the area extending to infinity on all sides, is also considered a face. The network can have multiedges and self-edges:



a) Write down the values of n, m, and f for a network with a single node and no edges.
b) How do n, m, and f change when we add a single node to the network along with a single edge attaching it to another node?
c) How do n, m, and f change when we add a single edge between two extant nodes (or a self-edge attached to just one node), in such a way as to maintain the planarity of the network?
d) Hence by induction prove a general relation between n, m, and f for all connected planar networks.
e) Now suppose that our network is simple (i.e., it contains no multiedges or self-edges). Show that the mean degree c of such a network is strictly less than six.

a)

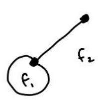
$n = 1$
$m = 0$
$f = 1$

b)

$n = 2$
$m = 1$
$f = 1$

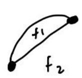
(nodes increase by 1, edges increase by 1, faces stay the same)

c)



$n = 2$
$m = 2$
$f = 2$

(nodes remain same, edges increase by 1 faces increase by 1)



$n = 2$
$m = 2$
$f = 2$

d) Euler's formula says: for any connected planar graph, $n - m + f = 2$.

Proof (by induction)

For $m = 0$: (base case)

- A connected graph with $m = 0$ will have $n = 1$ and $f = 1$
- $n - m + f = 1 - 0 + 1 = 2$ ✓

$\therefore$ $n - m + f = 2$ true for $m = 0$

For $m = 1$:

- A connected graph implies $n = 2$ and $f = 1$
- $n - m + f = 2 - 1 + 1 = 2$ ✓

$\therefore$ $n - m + f = 2$ true for $m = 1$

Inductive hypothesis: Suppose $n - m + f = 2$ true for some $m = k$, for a connected planar graph. for a graph with n nodes and f faces.

$k \Rightarrow k+1$

For k+1 edges, you either add the edge and another node to keep the graph connected, or you add the edge between two extant nodes.

If (k+1) th edge added alongside another node:

$$\left.\begin{aligned} f' &= f \\ m' &= k+1 \\ n' &= n+1 \end{aligned}\right\} \Rightarrow \begin{aligned} & n' - m' + f' \\ &= (n+1) - (k+1) + f \\ &= n + \cancel{1} - k - \cancel{1} + f \\ &= n - k + f \\ &= 2 \end{aligned}$$

$\therefore$ adding another edge and node still satisfies $n-k+f=2$.

If (k+1)th edge added between extant nodes:

Since the graph is connected $\Rightarrow$ extant nodes have a path between them.

$\Rightarrow$ adding an edge between them will form a cycle.

A formation of a new cycle creates a face.

So, we have:

$$\left.\begin{aligned} n' &= n \\ m' &= k+1 \\ f' &= f+1 \end{aligned}\right\} \begin{aligned} & n' - m' + f' \\ &= n - (k+1) + (f+1) \\ &= n - k + f \cancel{-1} + \cancel{1} \\ &= 2 \end{aligned}$$

Thus, adding a (k+1)th edge between two existing nodes still preserves the $n-m+f=2$ relation.

Thus, by principle of mathematical Induction, for a connected graph with n nodes, m edges and k faces,

$$n - m + f = 2 \qquad \forall m. \qquad \square$$

e) Suppose we have a simple connected graph G with n nodes, m edges and f faces.

For $m=0$:

$n=1$, $f=1$

mean degree $= \frac{0}{1} = 0 < 6$ ✓

For $m=1$: $n=2$, $f=1$, mean degree $= \frac{1}{2}$ < 6 ✓

For $m=2$: $n=3$, $f=1$, mean degree $= \frac{2}{3}$ < 6 ✓

(since no self-edges & multi-edges)

For $m \geq 3$:

Case 1:

Case 2:

By Euler's: $n-m+f=2$

For $m \geq 3$, each face is bounded by 3 or more edges

each edge surrounds up to 2 faces

$\Rightarrow$ $f \leq 3m$ and $m \geq 2f$

$\Rightarrow$ $3f \leq 9m$ $\qquad \Rightarrow 2m \geq 4f \geq 3f$

$\Rightarrow$ $3f \leq 2m$

$\Rightarrow$ $f \leq \frac{2}{3}m$

By Euler: $2 = n-m+f \leq n-m+\frac{2}{3}m$

$2 \leq n-m+\frac{2}{3}m$

$6 \leq 3n-3m+2m = 3n-m$

$m \leq 3n-6$

$\Rightarrow 2m \leq 6n-12$

$\therefore \frac{2m}{n} \leq \frac{6n-12}{n} = 6-\frac{12}{n} < 6$

$\therefore$ mean degree $c < 6$

3 Problem 6.11 **10 / 10**

- ✓ **+ 2 pts** Part a
- ✓ **+ 2 pts** Part b
- ✓ **+ 2 pts** Part c
- ✓ **+ 2 pts** Part d
- ✓ **+ 2 pts** Part e
- **+ 1 pts** Part e partial credit

Q3, Ex 10.2

10.2 Suppose that a network has a degree distribution that follows the exponential (or geometric) form $p_k = Ca^k$, where C and a are positive constants and $a < 1$.

a) Assuming the distribution is properly normalized, find C as a function of a.

b) Calculate the fraction P of nodes that have degree k or greater.

c) Calculate the fraction W of ends of edges that are attached to nodes of degree k or greater.

d) Hence show that the Lorenz curve—the equivalent of Eq. (10.24) for this degree distribution—is given by

$$W = P - \frac{1 - 1/a}{\log a} P \log P.$$

e) Show that the value of W is greater than one for some values of P in the range $0 \le P \le 1$. What is the meaning of these "unphysical" values?

Normalization assumes: $\sum_{k=0}^{\infty} p_k = 1$.

Then, for

$$p_k = C a^k$$

$$\sum_{k=0}^{\infty} p_k = \sum_{k=0}^{\infty} C \cdot a^k = 1$$

$$= C \cdot \underbrace{\sum_{k=0}^{\infty} a^k}_{\substack{\text{geometric series,} \\ \text{since } |a|<1, \\ \text{this converges to } \frac{1}{(1-a)}}} = 1$$

$$\therefore \Rightarrow C \cdot \left(\frac{1}{1-a}\right) = 1$$

$$\Rightarrow \boxed{C = (1-a)}$$

b) P (nodes that have degree k or larger)

P_k = prob. that nodes have degree k. or larger

$$\therefore P = \underbrace{\sum_{i=k}^{\infty} p_i}_{\substack{\text{prob that nodes w/} \\ \text{degree k or} \\ \text{larger}}} = 1 - \underbrace{\sum_{i=0}^{k-1} p_i}_{\substack{\text{prob. that} \\ \text{nodes have} \\ \text{degree less} \\ \text{than k}}} = 1 - \left(\sum_{i=0}^{k-1} C a^i\right)$$

$$= 1 - \left(C \sum_{i=0}^{k-1} a^i\right)$$

$$= 1 - (1-a) \underbrace{\sum_{i=0}^{k-1} a^i}_{\substack{\text{geometric series,} \\ |a|<1 \ \therefore \\ \text{sum is } \left(\frac{1-a^k}{1-a}\right)}}$$

$$= 1 - (1-\cancel{a}) \cdot \left(\frac{1-a^k}{\cancel{1-a}}\right) = 1 - (1-a^k)$$

$$= a^k$$

$$\therefore \boxed{P = a^k}$$

c)
$$W = \frac{\sum_{i=k}^{\infty} (n \cdot p_i) \times i}{\underbrace{\sum_{i=0}^{\infty} \underbrace{(n \cdot p_i)}_{\text{\# nodes with degree } i} \times i}_{\text{Total \# of edge ends}}}$$

$$= \frac{\sum_{i=k}^{\infty} (p_i) \cdot i}{\sum_{i=0}^{\infty} (p_i) \cdot i} = \frac{\sum_{i=k}^{\infty} (C a^i) \cdot i}{\sum_{i=k}^{\infty} (C \cdot a^i) \cdot i} = \frac{\cancel{C} \sum_{i=k}^{\infty} a^i \cdot i}{\cancel{C} \cdot \sum_{i=0}^{\infty} a^i \cdot i}$$

DENOMINATOR
Since $|a|<1$, by sum of geometric series,

$\sum_{i=0}^{\infty} a^i \cdot i$ is given by $S = a + 2a^2 + 3a^3 + 4a^4 + \dots$

$S \cdot a = a^2 + 2a^3 + 3a^4 + 4a^5 + \dots$

$S - Sa : (a-0) + (2a^2 - a^2) + (3a^3 - 2a^3) + (4a^4 - 3a^4) + (5a^5 - 4a^5) + \dots$

$\Rightarrow S(1-a) = a + a^2 + a^3 + a^4 + \dots$

$= 1 + a + a^2 + a^3 + \dots]$

$\Rightarrow S(1-a) = a\left(\frac{1}{1-a}\right)$ by sum of geometric series

$\Rightarrow S = a\left(\frac{1}{1-a}\right) \cdot \left(\frac{1}{1-a}\right)$

$S = \frac{a}{(1-a)^2}$

$\therefore \boxed{\text{Denominator} = \frac{a}{(1-a)^2}}$

Numerator:

$\sum_{i=k}^{\infty} a^i \cdot i = = ka^k + (k+1)a^{k+1} + \dots$

let $S = ka^k + (k+1)a^{k+1} + (k+2)a^{k+2} + \dots$

$\Rightarrow (S \cdot a) = ka^{k+1} + (k+1)a^{k+2} + (k+2)a^{k+3} + \dots$

$S - (S \cdot a) = ka^k + a^{k+1} + a^{k+2} + a^{k+3} + \dots$

$= a^k(k + a + a^2 + a^3 + \dots)$

$= ka^k + a^k(a + a^2 + a^3 + \dots)$

$S(1-a) = ka^k + a^{k+1}(1 + a + a^2 + a^3 + \dots)$

$\Rightarrow S = \left[ka^k + a^{k+1} \cdot \frac{1}{(1-a)}\right] \cdot \frac{1}{(1-a)}$

$= \frac{ka^k}{(1-a)} + \frac{a^{k+1}}{(1-a)^2}$

$\therefore$ Numerator $= \frac{ka^k}{(1-a)} + \frac{a^{k+1}}{(1-a)^2}$

We know that $\sum_{i=0}^{\infty} a^i = \frac{1}{(1-a)}$

$\therefore \sum_{i=k}^{\infty} a^i = \sum_{i=0}^{\infty} a^i - \sum_{i=0}^{k-1} a^i$

$= \frac{1}{(1-a)} - \frac{(1-a^k)}{(1-a)}$

$\sum_{i=k}^{\infty} a^i \cdot i = \frac{a^k}{(1-a)}$

$\therefore$

$\therefore W = \frac{\frac{ka^k}{(1-a)} + \frac{a^{k+1}}{(1-a)^2}}{\frac{a}{(1-a)^2}} = \frac{\frac{ka^k}{\cancel{(1-a)}} \cdot \cancel{(1-a)^2} + \frac{a^{k+1}}{\cancel{(1-a)^2}} \cdot \cancel{(1-a)^2}}{a} = \frac{ka^k(1-a) + a^{k+1}}{a}$

10(a)

$$W = \frac{k a^k (1-a) + a^{k+1}}{a}$$

$$W = P - \frac{1-1/a}{\log a} P \log P.$$

$$= \frac{k a^k - k a^{k+1} + a^{k+1}}{a}$$

$$= a^k - \frac{1-\frac{1}{a}}{\log a} a^k . \log P$$

$$= a^k - \left(1-\frac{1}{a}\right) P \, k$$

$$= a^k - \left(1-\frac{1}{a}\right) a^k . k =$$

$$= k a^{k-1} - k a^k + a^k$$

$$= \frac{k a^k}{a} - k a^k + a^k$$

$$= a^k - \left(k a^k - \frac{k a^k}{a}\right)$$

$$= a^k - \left(k a^k - \frac{k a^k}{a}\right)$$

$$= a^k - \left(k a^k - \frac{k a^k}{a}\right)$$

$$= a^k - k a^k + \frac{k a^k}{a}$$

$$= a^k - k . a^k \left(1 - \frac{1}{a}\right)$$

$$= \boxed{a^k - k a^k + k a^{k-1}}$$

a

$a^k = P$

$k \log a = \log P$

Sub $p = a^k$.

$$= P - k . P \left(1 - \frac{1}{a}\right)$$

Sub $\frac{\log P}{\log a} = k$

$$= P - \frac{\log P}{\log a} . P \left(1 - \frac{1}{a}\right)$$

$$= \boxed{P - \frac{\left(1-\frac{1}{a}\right)}{\log a} . P \log P}$$

$\square$

10e)

For $W > 1$,

$$P - \frac{\left(1-\frac{1}{a}\right)}{\log a} . P \log P > 1$$

$$\Rightarrow P - 1 > \frac{\left(1-\frac{1}{a}\right)}{\log a} P \log P$$

$$0 \geq P-1 > \left(1-\frac{1}{a}\right) . \underbrace{P}_{>0} \underbrace{\frac{\log P}{\log a}}_{>0}$$

· For $0 \leq P \leq 1$, $\log P < 0$

· since $k \geq 0$ and $k = \frac{\log P}{\log a}$

$\Rightarrow \log(a) < 0$

$\Rightarrow 0 \leq a \leq 1$

for RHS ≤ 0, $\left(1 - \frac{1}{\alpha}\right) < 0$

$\Rightarrow 1 < \frac{1}{\alpha} \Rightarrow \alpha < 1$ and $\frac{1}{\alpha} > 1$ so $0 \leq \alpha < 1$

Thus, for values of α such that $0 \leq \alpha \leq 1$, $W > 1$

4 Problem 10.2 **8 / 10**

- ✓ **+ 2 pts** part a
- ✓ **+ 2 pts** part b
- ✓ **+ 2 pts** part c
- ✓ **+ 2 pts** part d
- **+ 2 pts** part e
- **+ 1 pts** part a partial credit
- **+ 1 pts** part b partial credit
- **+ 1 pts** part c partial credit
- **+ 1 pts** part e partial credit

Part e- What is the meaning of the "unphysical" values? you can get W> 1 when k = log(P)/log(a) < 1, but this is "unphysical" because k is a degree, hence an integer.

These 'unphysical' values mean that the fraction of ends of edges attached to nodes of degree > 1

→ More than <u>all</u> the ends of edges are attached to nodes of degree > 1

Q5.

5. *Clustering coefficients.* Draw a small network in which the global clustering coefficient and mean local clustering coefficient have different values. Write down the adjacency matrix for this network.

Global clustering coefficient formula:

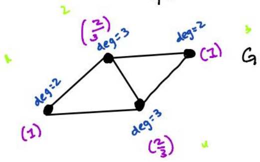
$$\frac{\#\text{ closed triplets}}{\text{total }\#\text{ triplets}} \Rightarrow \frac{3\times \#\ \Delta\text{'s}}{\#\text{ connected triplets}}$$

Local clustering coefficient formula:

$$\frac{\left(\begin{array}{c}\#\text{ neighbours of node } i\\ \text{connected to each other}\end{array}\right)}{\left(\begin{array}{c}\text{total }\#\text{ pairs of neighbours}\\ \text{of } i\end{array}\right)}$$

$$\begin{array}{c|cccc} & 1 & 2 & 3 & 4 \\ \hline 1 & & & & \\ 2 & & & & \\ 3 & & & & \\ 4 & & & & \end{array}$$

$$\begin{array}{cc} & \begin{array}{cccc}1 & 2 & 3 & 4\end{array} \\ \begin{array}{c}1\\2\\3\\4\end{array} & \begin{bmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix} \end{array}$$



$$GCC = \frac{2\times 3}{8} = \frac{6}{8} = \frac{3}{4}$$

m = local clustering coefficient

$\therefore$ mean local clustering coefficient $= \frac{1}{n}\sum_{i=1}^{n} C_i =$

$$= \frac{(1) + (1) + \frac{2}{3} + \frac{2}{3}}{4} = \frac{5}{6}$$

Q6

6. *More on ER graphs.*

(a) Calculate the global clustering coefficient C of an ER graph (in expectation over the ensemble).

(b) Is the ER network ensemble a good model for a social network? Why or why not? How can it be adjusted to produce a better model?

(a)

The global clustering coefficient, C, of an ER graph is as follows:

$$= \frac{(\#\text{ of triangles} \times 3)}{\#\text{ connected triplets}}$$

For an ER-graph, $G(n,p)$,

$$\mathbb{E}[\#\text{ triangles}] = 3\binom{n}{3}\cdot p^3$$

3 ways of choosing each triplet of node

choosing 3 out of n nodes to form Δ

probability of 3-edges existing between those nodes

$$\mathbb{E}[\#\text{ connected triplets}] = 3\binom{n}{3} p^3 + 3\binom{n}{3} \underbrace{p^2(1-p)}$$

choosing 3 out of n nodes to form triplet

the triplet of nodes is connected, forming Δ

↳ only 2 out of 3 edges between triplet of nodes

5 Clustering coefficients 10 / 10

✓ + 10 pts **Complete (everyone has a different network so I am grading on completion)**

These 'unphysical' values mean that the fraction of ends of edges attached to nodes of degree > 1

→ More than <u>all</u> the ends of edges are attached to nodes of degree > 1

Q5.

5. *Clustering coefficients.* Draw a small network in which the global clustering coefficient and mean local clustering coefficient have different values. Write down the adjacency matrix for this network.

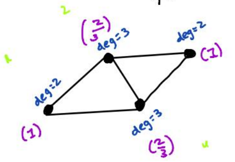
Global clustering coefficient formula:

$$\frac{\#\text{ closed triplets}}{\text{total }\#\text{ triplets}} \Rightarrow \frac{3\times \#\ \Delta\text{'s}}{\#\text{ connected triplets}}$$

$$\begin{array}{c|cccc} & 1 & 2 & 3 & 4 \\ \hline 1 & 0 & 1 & 0 & 1 \\ 2 & 1 & 0 & 1 & 1 \\ 3 & 0 & 1 & 0 & 1 \\ 4 & 1 & 1 & 1 & 0 \end{array}$$

Local clustering coefficient formula:

$$\frac{(\#\text{ neighbours of node } i \text{ connected to each other})}{(\text{total }\#\text{ pairs of neighbours of } i)}$$



$$GCC = \frac{2\times 3}{8} = \frac{6}{8} = \frac{3}{4}$$

m = local clustering coefficient

∴ mean local clustering coefficient $= \frac{1}{n}\sum_{i=1}^{n} C_i =$

$$= \frac{(1)+(1)+\frac{2}{3}+\frac{2}{3}}{4} = \frac{5}{6}$$

<u>Q6</u>

6. *More on ER graphs.*

(a) Calculate the global clustering coefficient C of an ER graph (in expectation over the ensemble).

(b) Is the ER network ensemble a good model for a social network? Why or why not? How can it be adjusted to produce a better model?

(a)

The global clustering coefficient, C, of an ER graph is as follows:

$$= \frac{(\#\text{ of triangles} \times 3)}{\#\text{ connected triplets}}$$

For an ER-graph, $G(n,p)$,

$$\mathbb{E}[\#\text{ triangles}] = 3\binom{n}{3}\cdot p^3$$

3 ways of choosing each triplet of node

choosing 3 out of n nodes to form Δ

probability of 3-edges existing between those nodes

$$\mathbb{E}[\#\text{ connected triplets}] = 3\binom{n}{3}p^3 + 3\binom{n}{3}\underbrace{p^2(1-p)}$$

choosing 3 out of n nodes to form triplet

the triplet of nodes is connected, forming Δ

↳ only 2 out of 3 edges between triplet of nodes

$$\therefore\ C = \frac{\cancel{3}\cancel{\binom{n}{3}}p^3}{\cancel{3}\cancel{\binom{n}{3}}\left[p^3 + p^2(1-p)\right]}$$

$$= \frac{p^3}{p^3 + p^2 - p^3} = \frac{p^3}{p^2} = \boxed{p}$$

b) No, ER model is not a good model for social networks because it models local clustering coefficient poorly because it assumes that $p \to 0$ as $N \to \infty$ which gives a very small local clustering coefficient and does not represent the "my friends' friends are more likely to be my friends" phenomenon.

ER graphs don't model the small-world effect well.

To improve, there needs to be a way to add shortcuts to ER-graphs to model the 'small-world' aspect of social networks, and there should also be a way to model hubs by looking at the degrees of the nodes, which are often common in social networks

6 ER graphs 10 / 10

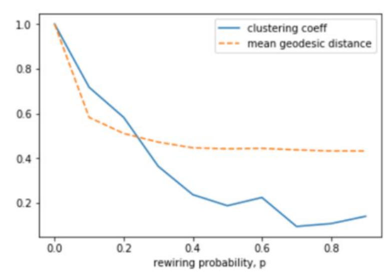
- ✓ **+ 5 pts** **Part a**
- ✓ **+ 5 pts** **Part b**
- **+ 3 pts** Part a partial credit

```python
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
import networkx as nx
```

```python
#create sample of Watts-Strogatz graphs
# for each graph in sample, calculate mean geodesic distance and clustering coeff

def sample_WS_for_p(num_graphs, num_nodes_G, k=4):
    clustering_coeff = []    #list of avg. clustering coeffs for each G in sample
    mean_geodesic_dist = []  #list of mean_geodesic distance for each G in sample
    p_list = np.arange(0, 1, 0.1)
    for p in p_list:
        G = nx.watts_strogatz_graph(num_nodes_G, k, p)
        clustering_coeff.append(nx.average_clustering(G))
        mean_geodesic_dist.append(nx.average_shortest_path_length(G))


    #Normalizing each value of clustering coeff and mean geodesic distance
    clustering_coeff_norm = []
    mean_geodesic_dist_norm = []

    clustering_coeff_norm = np.divide(clustering_coeff , max(clustering_coeff))
    mean_geodesic_dist_norm = np.divide(mean_geodesic_dist , max(mean_geodesic_dist))

    #Graphing mean geodesic distance and clustering coefficient v/s rewiring probability, p

    data = pd.concat([pd.Series(clustering_coeff_norm), pd.Series(mean_geodesic_dist_norm)], axis = 1)

    print(data)
    plt.plot(p_list, clustering_coeff_norm, label = 'clustering coeff')
    plt.plot(p_list, mean_geodesic_dist_norm, ls='--', label = 'mean geodesic distance')
    plt.legend(loc='best')
    plt.xlabel('rewiring probability, p')
    plt.show()
```

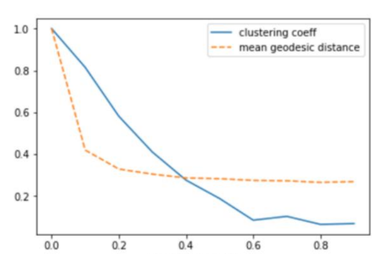
In [25]:
```python
sample_WS_for_p(100, num_nodes_G = 50)
```

	0	1
0	1.000000	1.000000
1	0.717333	0.582769
2	0.582857	0.511015
3	0.364000	0.472123
4	0.237333	0.446769
5	0.187905	0.442462
6	0.224667	0.444554
7	0.095238	0.437538
8	0.108000	0.432738
9	0.140667	0.432492



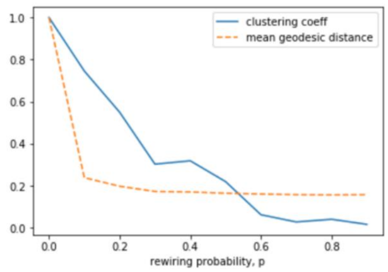
In [26]:
```python
sample_WS_for_p(100, num_nodes_G = 100)
```

	0	1
0	1.000000	1.000000
1	0.814667	0.419671
2	0.581333	0.329051
3	0.410000	0.304659
4	0.276000	0.286525
5	0.188238	0.282792
6	0.084857	0.274886
7	0.103032	0.272894
8	0.064762	0.265286
9	0.068429	0.269082



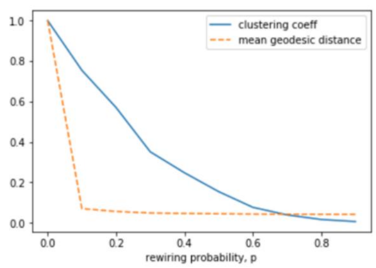
```
In [27]: sample_WS_for_p(100, num_nodes_G = 200)
```

	0	1
0	1.000000	1.000000
1	0.746000	0.239580
2	0.550333	0.198655
3	0.303619	0.173844
4	0.319556	0.171483
5	0.220571	0.165614
6	0.063381	0.161620
7	0.029000	0.158657
8	0.041714	0.157826
9	0.017690	0.158711




```
In [28]: sample_WS_for_p(100, num_nodes_G = 1000)
```

	0	1
0	1.000000	1.000000
1	0.754400	0.070519
2	0.570238	0.056527
3	0.351290	0.048938
4	0.247733	0.046665
5	0.154483	0.044962
6	0.076844	0.043577
7	0.039314	0.042764
8	0.016798	0.042577
9	0.006602	0.042408

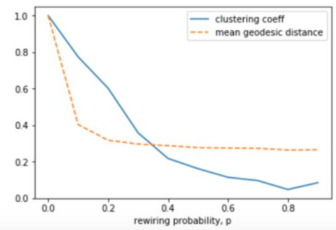


Overall, both the clustering coefficient and the mean geodesic distance decrease as the rewiring probability increases.

In general, the clustering coefficient takes a higher value than the mean geodesic distance, for a certain probability p. As the number of nodes, N, increases, the clustering coefficient remains greater than the mean geodesic distance, for larger values of p than when N is smaller. This can be seen from the graphs, as the clustering coefficient is greater than mean geodesic distance from p = 0 to p = 0.34 approximately, for N = 50 and from p = 0 to p = 0.65 for N = 1000.
The curve for the mean geodesic distance has a consistent pattern of a sharp drop and then levelling out, for all the values of N. However, for higher values of N, the mean geodesic distance decreases more, over the lower values of p, than it does for the lower values of N. This results in the mean geodesic distance levelling out at a much lower value over p, for higher N.

```
In [31]: #Modelling clustering coefficient and mean geodesic distance over rewiring probability for
         #fixed num of nodes, sample with single graph

         sample_WS_for_p(num_graphs = 1, num_nodes_G = 100)
```

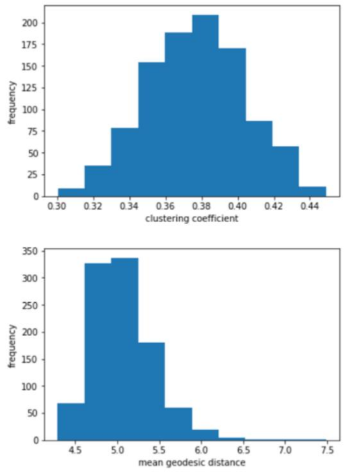
	0	1
0	1.000000	1.000000
1	0.774000	0.403075
2	0.602000	0.317176
3	0.358000	0.296298
4	0.217333	0.287467
5	0.161333	0.276314
6	0.114286	0.274698
7	0.096063	0.272894
8	0.047302	0.263498
9	0.084635	0.265522



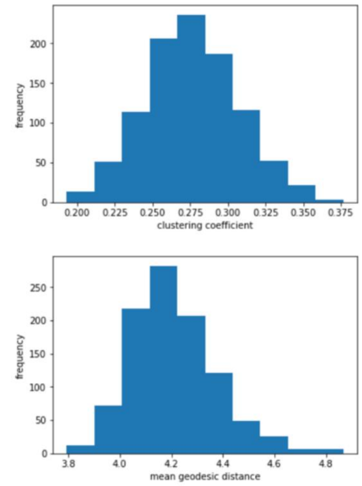
For a sample with a single realization of network, and for a fixed value of N=100, the graph looks like the one above. One can see that the patterns remain the same as mentioned before, with the clustering coefficient and the mean geodesic distance decreasing as the rewiring probability increases, and the clustering coefficient remaining higher than the mean geodesic distance until a certain p, and the pattern of the mean geodesic distance sharply decreasing over a small range of p , before levelling off.

```
#Finding the frequencies of the sample mean clustering coefficient and sample mean geodesic distance

def WS_over_sample_means(num_graphs, num_nodes_G, k=4, p=1):
    clustering_coeff = []     #list of avg. clustering coeffs for each G in sample
    mean_geodesic_dist = []   #list of mean_geodesic distance for each G in sample

    for i in range(0, num_graphs):   #
        G = nx.watts_strogatz_graph(num_nodes_G, k, p)
        clustering_coeff.append(nx.average_clustering(G))
        mean_geodesic_dist.append(nx.average_shortest_path_length(G))

    #Graphing sample means versus clustering coeff, and sample means versus mean geodesic distance

    plt.hist(clustering_coeff)
    plt.xlabel('clustering coefficient')
    plt.ylabel('frequency')
    plt.show()

    plt.hist(mean_geodesic_dist)
    plt.xlabel('mean geodesic distance')
    plt.ylabel('frequency')
    plt.show()
```

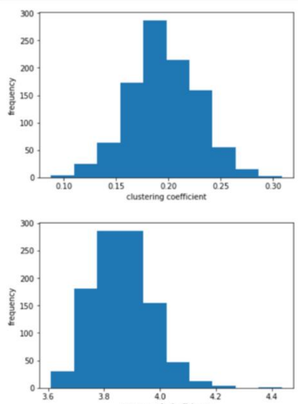
```
In [66]: WS_over_sample_means(num_graphs = 1000, num_nodes_G = 100, p = 0.1)
```



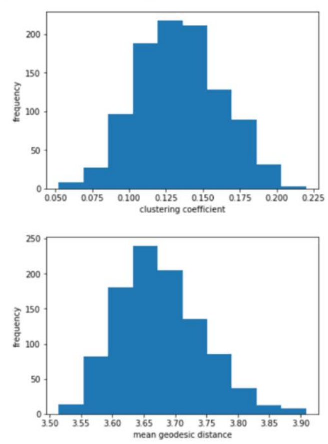

```
In [67]: WS_over_sample_means(num_graphs = 1000, num_nodes_G = 100, p = 0.2)
```



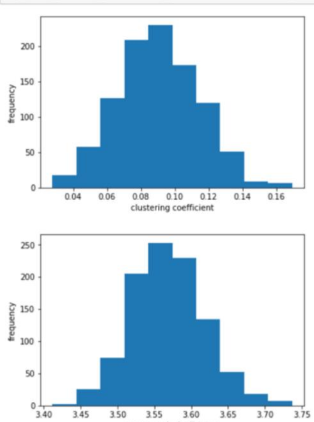

```
In [68]: WS_over_sample_means(num_graphs = 1000, num_nodes_G = 100, p = 0.3)
```



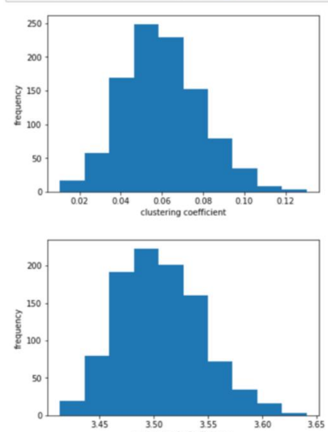

```
In [69]: WS_over_sample_means(num_graphs = 1000, num_nodes_G = 100, p = 0.4)
```



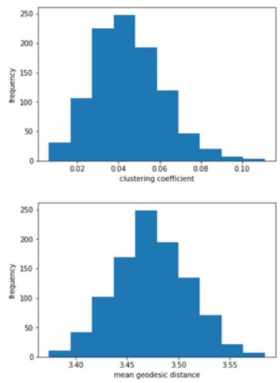

```
In [70]: WS_over_sample_means(num_graphs = 1000, num_nodes_G = 100, p = 0.5)
```



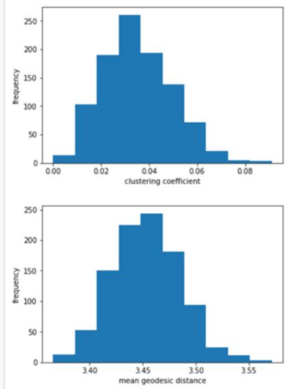

```
In [71]: WS_over_sample_means(num_graphs = 1000, num_nodes_G = 100, p = 0.6)
```



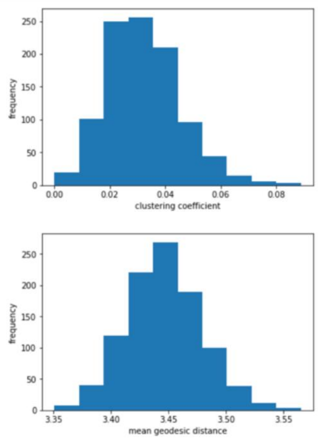
```
In [72]: WS_over_sample_means(num_graphs = 1000, num_nodes_G = 100, p = 0.7)
```




```
In [73]: WS_over_sample_means(num_graphs = 1000, num_nodes_G = 100, p = 0.8)
```




```
In [74]: WS_over_sample_means(num_graphs = 1000, num_nodes_G = 100, p = 0.9)
```



From the graphs above, you can see that for lower values of p, the sample mean geodesic distance is right-skewed. As rewiring probability increases, both these values become less skewed and their distribution moves towards a normal distribution. By p=0.9, the sample mean geodesic distance has an approximately normal distribution.

For the sample mean clustering coefficient, for lower values of p, the distribution is approximately normal (i.e. from p=0.1 to p = 0.5). As p increases above 0.5, the sample mean distribution of the clustering coefficient becomes increasingly right-skewed.

The Watts-Strogatz model is a good model for social networks because it is based on high local clustering and short mean geodesic distance, giving the model a small-world property. This is one of the findings that we have seen from Milgram's experiment on social networks, and therefore can be used to model other social networks very well as it represents the phenomenon of "my friends' friends tend to be my friends" well.

7 Watts-Strogatz networks **10 / 10**

- ✓ **+ 5 pts** Part a
- ✓ **+ 5 pts** Part b
- **+ 0 pts** Missing