CS 5854 : Networks, Crowds, and Markets
Homework 1

Instructor: Rafael Pass TAs: Cody Freitag, Drishti Wali

Assigned: September 5, 2019 Due: September 24, 2019, 11:59 pm

Ananya Shivaditya (aks298)

September 24, 2019

Collaborators: I worked with Alexander Popeil, Nathan Cinnamond, and Shobhna Jayaraman.
Outside resources: I uses the following outside resources:

1. Wikipedia

Late days: I have used 2 late days on this assignment.

Part 0: Slack

Join the slack channel for the course, via the following link: SLACK

Part 1: Game Theory

1. For each of the following three two-player games, find (i) all strictly *dominant* strategies, (ii) the action profiles which survive iterative removal of strictly *dominated* strategies, and (iii) all pure-strategy Nash equilibria. Give a brief justification for each part.

(a)

	$(*, L)$	$(*, R)$
$(U, *)$	(5, 4)	(4, 5)
$(D, *)$	(4, 4)	(0, 0)

(b)

	$(*, L)$	$(*, R)$
$(U, *)$	(2, 2)	(2, 1)
$(D, *)$	(3, 2)	(0, 3)

(c)

	$(*, L)$	$(*, R)$
$(U, *)$	(6, 5)	(4, 5)
$(D, *)$	(5, 4)	(2, 2)

Solution:

(a)

	$(*, L)$	$(*, R)$
$(U, *)$	(5, 4)	(4, 5)
$(D, *)$	(4, 4)	(0, 0)

i. All strictly *dominant* strategies

A. Column: Doesn't have a strictly dominant strategy. If Row plays U, Column's best response would be R to get 5 (5 better than 4 if he plays L). But if Row plays D, Column's best response would be L to get 4(4 better than 0 if he plays R). So his best response is dependent on Row's move, and is therefore not dominant.

B. Row: The strictly dominant strategy of Row is (U,*), independent of Column's move, he gets 5 (5 better than 4 in D) and 4 (4 better than 0 in D) if Column plays L and R respectively.

ii. Action profiles which survive iterative removal of strictly *dominated* strategies:
Removing (D,*) from Row's strategy:

	$(*, L)$	$(*, R)$
$(U, *)$	(5, 4)	(4, 5)

Now removing (*,L) for Row:

	$(*, R)$
$(U, *)$	(4, 5)

(U,R) = (4,5) is the only action profile remaining.

iii. All pure-strategy Nash equilibria (U,R) = (4,5) is the only PNE since no player would unilaterally deviate from this state.

(b)

	$(*, L)$	$(*, R)$
$(U, *)$	(2, 2)	(2, 1)
$(D, *)$	(3, 2)	(0, 3)

i. All strictly *dominant* strategies

A. Column: Doesn't have a strictly dominant strategy. If Row plays U, Column's best response would be L to get 2 (2 better than 1 if he plays L). But if Row plays D, Column's best response would be R to get 3(3 better than 2 if he plays R). So his best response is dependent on Row's move, and so is not dominant.

B. Row: Doesn't have a strictly dominant strategy. If Column plays L, Row's best response would be D to get 3 (3 better than 2 if he plays U). But if Column plays R, Row's best response would be U to get 2(2 better than 0 if he plays R). So his best response is dependent on Column's move, and so is not dominant.

ii. All action profiles will survive iterative removal of strictly *dominated* strategies since there are no strictly dominated strategies.
The surviving action profiles are:

	$(*, L)$	$(*, R)$
$(U, *)$	(2, 2)	(2, 1)
$(D, *)$	(3, 2)	(0, 3)

iii. All pure-strategy Nash equilibria: There is a cycle which spans all states, and so no state is PNE. BRD doesn't converge for the following case:
(U,L) -> (D,L) -> (D,R) -> (U,R) -> (U,L) ...
(2, 2) -> (3, 2) -> (0, 3) -> (2, 1) -> (2,2) ...

(c)

	$(*, L)$	$(*, R)$
$(U, *)$	(6, 5)	(4, 5)
$(D, *)$	(5, 4)	(2, 2)

i. All strictly *dominant* strategies

A. Column: Doesn't have a strictly dominant strategy. It has a weakly dominant strategy (*,L). If Row plays U, Column's best response could be either L or R to get 5 either way. But if Row plays D, Column's best response would be L to get 4(4 better than 2 if he plays R). So his best response is not dependent on Row's move, but can be chosen ambiguously in one case, and so is not *strictly* dominant, but is weakly dominant.

B. Row: The strictly dominant strategy of Row is (U,*), independent of Column's move, he gets 6 (6 better than 5 in D) and 4 (4 better than 2 in D) if Column plays L and R respectively.

ii. Action profiles which survive iterative removal of strictly *dominated* strategies Removing (D,*) from Row's strategy:

	$(*, L)$	$(*, R)$
$(U, *)$	(6, 5)	(4, 5)

iii. All pure-strategy Nash equilibria
(U,L) = (6,5) and (U,R) = (4,5) are a PNEs since no player would unilaterally deviate from these states.

□

2. Consider the two-player game given by the following payoff matrix:

	$(*, L)$	$(*, M)$	$(*, R)$
$(t, *)$	(-1, 2)	(5, 1)	(0, 0)
$(m, *)$	(1, 2)	(-1, 0)	(6, 2)
$(b, *)$	(4, 1)	(3, 1)	(2, 0)

(a) Does either player have a strictly dominant strategy? If so, which player, what strategy, and why? If not, what is the smallest number of entries in the payoff matrix which would need to be changed so that some player did have a strictly dominant strategy? Justify why this is the minimum, i.e. there is no smaller value that works.

(b) What are player 1's and player 2's best-response sets given the action profile (m, L)?

(c) Find all pure-strategy Nash equilibria for this game. (Argue why all that you wrote are PNEs and why there are no others.) Describe how best-response dynamics might converge to each pure-strategy Nash equilibria.

Solution:

(a) No player has strictly dominant strategy. Furthermore: At least 2 states must be altered for any player to get a dominant strategy. This is explained for each player below.

   i. Column: Has a weakly dominant strategy (*,L).

      A. If Row plays t, Column's BR: L to get 2 (2 > 1 in M and 2 > 0 in R).

      B. If Row plays m, Column's BR: L or R to get 2 either way (both > 0 in M).

      C. If Row plays b, Column's BR: L or M to get 1 either way (both > 0 in R).

      So his best response is not dependent on Row's move, but can be chosen ambiguously in two cases (B and C above), and is therefore not *strictly* dominant, but is weakly dominant.

   To change this to strict, we would need to make one of L or R or M strictly greater than the others in each of the 3 rows. This means 2 rows would need altering to match the third, and any less changes would not work.

   ii. Row: Doesn't have a strictly dominant strategy.

   - If Column plays L, Row's BR : choose b to get 4 (4 > 1 if m and 4 > -1 if t).
   - If Column plays M, Row's BR : choose t to get 5 (5 > -1 if m and 5 > 3 if b).
   - If Column plays R, Row's BR : choose m to get 6 (6 > 2 if b and 6 > 0 if t).

      So his best response is dependent on Column's move, and so is not dominant.

   To change this to strict dominance, we would need to make one of m or t or b strictly greater than the others in each of the 3 columns. This means 2 columns would need altering to match the third, and any less changes would not work.

(b) i. Column: This is a state of max utility for Column for given choice of Row. He has no incentive to move. Best response would be to stay where he is or he might choose to switch to a state with equally large utility at (m,R) = 2

ii. Row: BR would be to increase utility to 4 by moving to (b,L)

(c) (6,2) and (4,1) are the two PNEs. All other states do give incentive for either Column or Row to switch states to increase utility without the other player switching.
BRD might converge to each PNE becuase:

i. BRD takes a player to a new state only if there is increase in utility for him, and so PNE might be landed on during BRD.

ii. Once landing on a PNE, BRD will converge, since there is no incentive for movement for either player at PNE.

iii. BRD may not converge, if cycle in BRD states exist, and PNE lies outside the cycle.

□

3. (a) Prove the following: If player 1 in a two-person game has a dominant strategy s_1, then there is a pure-strategy Nash equilibrium in which player 1 plays s_1 and player 2 plays a best response to s_1.

   (b) Is the equilibrium from part (a) necessarily a *unique* pure-strategy Nash equilibrium? Justify your answer.

   (c) In particular, can there also exist a pure-strategy Nash equilibrium where player 1 does not play s_1? Justify your answer.

   (d) If s_1 is instead a *strictly dominant* strategy for player 1, how do the answers to (a)-(c) change? Provide proper justifications for each part.

Solution:

(a)
- Player 1 having a strictly or a weakly dominant strategy means one of his choices of state maximizes his utility over his other choices, irrespective of which move his opponent chooses. SO in the playoff matrix, he would likely choose this state's column s1 and stay put.
- PNE is a state where neither player would deviate from unilaterally. So if player 1 plays his dominant strategy's column, player two would choose the row value that gives him maximum utility in that column, i.e. the best response to s1 and that would converge BRD. A PNE would exist BR(s1),s1 in the playoff matrix.

(b) The PNE from (a) would not necessarily be unique. This is because player 2's best response to s1 (i.e the maximum value in that column) could be in two rows of the column s1, equal to each other and higher than all other utility values in that column. So there could be multiple PNEs, states which no player would have incentive to deviate from.

(c) Yes, if s1 is a weakly dominant strategy. This means in one or more rows, s1 would give as high a utility as a different column. If, in that row, player 1 chooses to play the other column, and that choice of (row,column) also has the highest utility across rows, then that will be a PNE.

(d) In case s1 is a strictly dominant strategy:

   i. (a) would not change. There would still be a PNE as explained above. One of player1's choices of state maximizes his utility over his other choices, irrespective of which move his opponent chooses. SO in the playoff matrix, he would likely choose this state's column s1 and stay put, and player2 would maximize his utility in that column and land on a PNE. A PNE would exist BR(s1),s1 in the playoff matrix.

(e) (b) would not change. The PNE would not necessarily be unique, because player 2's best response to s1 (i.e the maximum value in that column) could be in two rows of the column s1, equal to each other and higher than all other utility values in that column. So there could be multiple PNEs, states which no player would have incentive to deviate from.

(f) (c) No. In case of s1 being a strictly dominant strategy, there would be no PNE in any of the columns where player1 doesn't play s1. Player1 would try to increase his utility to s1 and switch columns. So the PNE from (a) might not be unique, but would have to exist in the column of s1.

□

4. Formulate a normal-form game (as a payoff matrix) that has a unique pure-strategy Nash equilibrium, but for which best-response dynamics does not always converge (i.e. there are possible starting states for which BRD will not converge). Justify your answer. (*Hint:* Rock-paper- scissors has no equilibrium, and thus BRD will not converge. Can you combine this with a game that does have an equilibrium?)

Solution:

	(*,R)	(*,P)	(*,S)
(R,*)	(0, 0)	(0, 3)	(3, 0)
(P,*)	(3, 0)	(0, 0)	(0, 3)
(S,*)	(0, 3)	(3, 0)	(3, 3)

This playoff matrix represents Rock Paper Scissors, but with the state of (S, S) modified to be of utility (3,3) instead of (0,0). This state is a PNE since neither player would want to deviate away from this state as it is equal to the highest utility they can achieve. So there is no gain in utility if the other player's state is the same. But for this case, BRD may or may not converge, as there exists a cycle:
(R,P) -> (S,P) -> (S,R) -> (P,R) -> (P,S) -> (R,S) -> (R,P)...
(0, 3) -> (3, 0) -> (0, 3) -> (3, 0) -> (0,3) -> (3, 0) -> (0,3) ...

So this cycle exists, and the PNE (S,S) = (3,3) isn't included in it, and may never be reached. So BRD may not converge to the PNE, and the PNE does exist. □

Part 2: Graph Theory

Note: Unless stated otherwise, please assume for any problem involving graphs that we refer to *undirected* and *unweighted* graphs.

5. Given a graph, we call a node x in this graph *pivotal* for some pair of nodes y and z if x (not equal to y or z) lies on every shortest path between y and z.

   (a) Give an example of a graph in which every node is pivotal for at least one pair of nodes. Explain your answer.

   (b) For any integer $c \geq 1$, construct a graph where every node is pivotal for at least c different pairs of nodes. That is, if I give you any value for $c \geq 1$, you should be able to give me such a graph. Explain your answer.

   (c) Give an example of a graph having at least four nodes in which there is a single node x which is pivotal for every pair of nodes not including x. Explain your answer.

Solution:

(a) Regular polygon with 5 nodes and 5 edges. Each node is pivotal for its immediate neighbours. (Figure 1 below.) This means: 1 is pivotal for (5,2), 2 is pivotal for (1,3), 3 is pivotal for (2,4), 4 is pivotal for (3,5), 5 is pivotal for (4,1).

(b) (Figure 2 given on next page.) Regular polygon for n nodes where n = 2c+3. This works because each node will lie in the shortest path for at least (n-3)/2 pairs of the neighbors it has on it's half of the polygon. n-3 because in the case n is even, there will be two shortest paths for each pair that's diametrically opposite, reducing the possible pair that can be counted for each node to be pivotal by 2. In case n is odd, this solution works well, so n becomes 2c+3. This solution is not the minimal solution but it does satisfy all requirements. Also, there must be many other graphs that would satisfy the requirements.

(c) (Figure 3 given on next page.) Node 3 is pivotal for all pairs shortest paths (1,2),(2,4),(1,4)

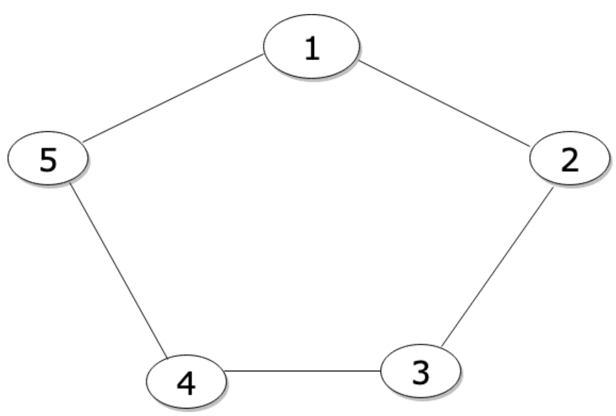


Figure 1: (a) c = 1

□

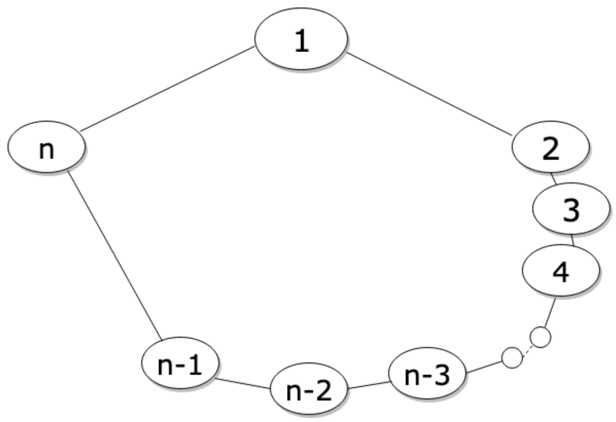


Figure 2: (b) n = 2c+3

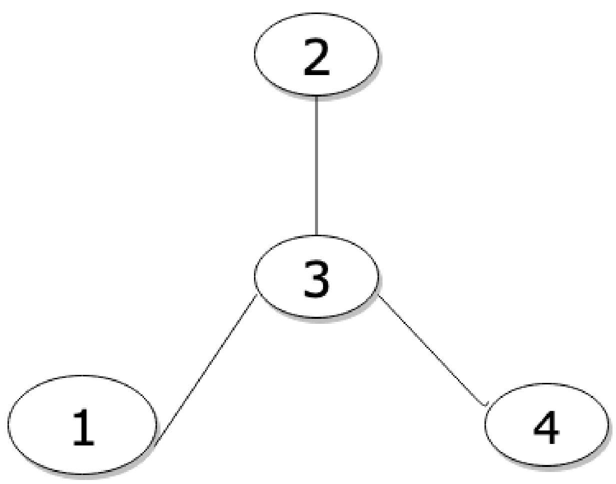


Figure 3: (c) 4 node graph with Node 3 as pivot for all shortest paths (1,2),(2,4),(1,4)

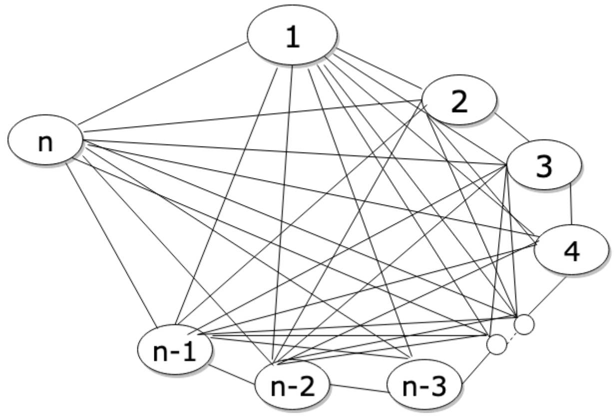


Figure 4: Fully connected graph, each node connected to all other nodes with an edge. Number of nodes can be any arbitrary n

6. Given some connected graph, let the *diameter* of a graph be the maximum distance (i.e. shortest path length) between any two nodes. Let the *average distance* be the expected shortest path length between a randomly selected pair of distinct nodes.

   (a) Let G be a graph with average distance A. What is the smallest diameter possible for such a graph? Provide a graph G that attains this minimum and prove that any smaller is impossible.

   (b) Give a graph G with diameter at least $3 \cdot A$.

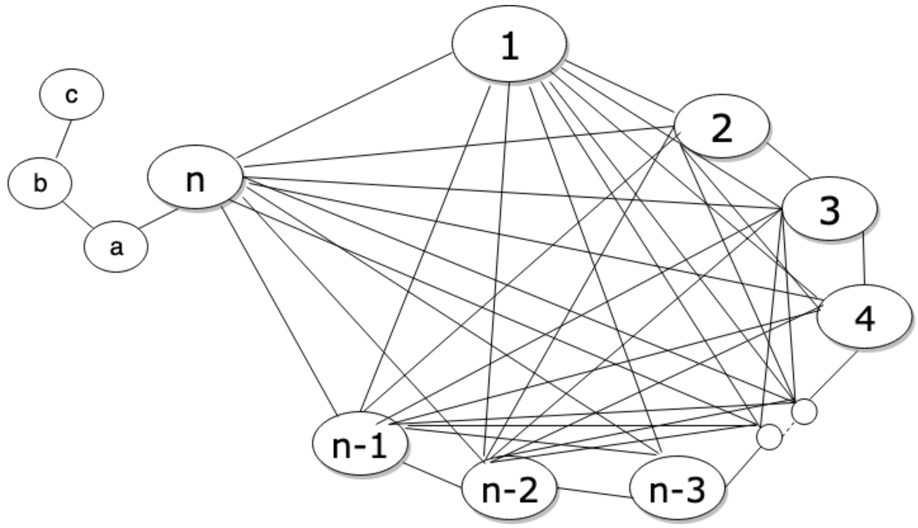
   (c) Repeat (b) for a diameter of at least $100 \cdot A$. (You don't need to draw the graph, just describe it and briefly justify why the diameter is at least 100 times larger than than the average distance.) Describe how you could extend this to an arbitrarily large factor $C \cdot A$.

   (d) Discuss what the diameter and average distance of a social network (given as a graph) might represent. What might it mean if the diameter is very similar to the average distance? What might it mean if the diameter is much greater?

Solution:

(a) Smallest possible diameter = A. This is because diameter is the *longest* shortest path. A is the average shortest path length, it means there must exist a shortest path which is as high as A. The smallest possible diameter therefore that could exist in the graph would be A. Any shortest path smaller than A would not qualify as the diameter of the graph, since there would definitely be a longer shortest path to bring the average higher up to A. Figure 4 (above) shows a graph which has A = 1 and D = 1 which is the smallest it can be.

(b) Here we need diameter to be atleast 3 times A. Consider a fully connected graph, each node connected to all other nodes with an edge and number of nodes = n. Now add a 3 node tail to it, i.e. connect one node to the fully connected component, and another



$$\text{Average Distance} = \frac{\frac{n(n-1)}{2}+1+2+3+4*(n-1)+1+2+3*(n-1)+1+2*(n-1)}{\frac{(n+3)(n+2)}{2}} = \frac{n^2+17n+2}{n^2+5n+6}$$

to that, and another to that. So, the diameter = 4, and we get a formula for average distance in terms of n.

Solving for n after creating the inequality A <= 4/3 (Since D is 4 and we want it to be greater than 3*A) we get n >= 31. So if n >=31, A will be <= 4/3, and the diameter will be more than three times as large as the average distance.

(c) Similar to the structure of a graph in (b) of n nodes extended by a single path with c nodes on it, we derive:

$$A = \frac{\frac{n(n-1)}{2} + \frac{c(c+3)}{2}(n-1) + \frac{c(c+1)(c+2)}{6}}{\frac{(n+c)(n+c-1)}{2}}$$

$$D = c + 1$$

And we can alter the values of c and determine n by:

$$\frac{D}{A} \geq c$$

And this can be arbitrarily extended to c = 100 or more.

(d) Diameter of a social network would represent the largest number of connections that a person might have to hop to reach a specific person. It is unlikely that he'd actually have to make as many hops as the diameter though, because a social netwrok's structure is much like a large strongly connected component, and several smaller strongly connected components attached to it. So it is far mor likely that he would only have to hop the average distance of the social network, in order to reach a specific person.

If D is nearly equal to A, it would mean a more dense graph closer to completely connected graph. If D >> A, it would mean a sparse graph, with less cycles and closed connections.

□

7. Consider a graph G on n nodes.

   (a) What is the fewest number of edges such that G is connected? Give an example with that many edges, and argue why any fewer edges must result in a graph G which is disconnected.

   (b) What is the fewest number of edges such that any two nodes in G have a shortest path length of 1? Again, prove that this is the minimum by arguing that no fewer is possible and that the number you give is attainable.

   (c) Repeat part (b) for a shortest path length of at most 2.

Solution:

(a) Proved by induction:

   i for n=1, no edges are required, since graph is a single node. It is connected. So:

$$E[G(n)] = n - 1$$

   ii for some G(n) which is a connected graph, in order to add one more node to the graph in a manner that it results in a connected graph, we only to add one edge between the new node and any one existing node. So:

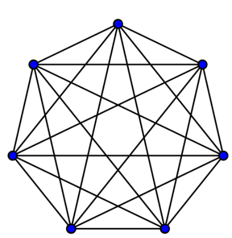
$$E[G(n+1)] = E[G(n)] + 1$$

   Following i and ii above, we get:

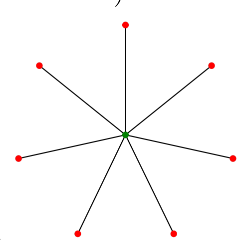
$$E[G(n)] = n - 1 \ldots\ldots\ldots\ldots\ldots\ldots \forall n$$

   Any less than (n - 1) would result in a disconnected graph, because some nodes would still be left disconnected in a graph with n nodes.

(b) Only a fully connected graph would give shortest path length 1 to every node pair. This means the fewest number of edges such that any two nodes in G have a shortest path length of 1 would be equal to number of nodes, with is n. Any fewer would result in the shortest path between some pair of nodes to be greater than 1. (Figure shown below)



<— (b)Complete Graph and Star graph (c) —>



(c) The fewest number of edges such that any two nodes in G have a shortest path length of 2 is n - 1 in a star graph where all nodes are connected to one node directly. No fewer edges than this is possible, for the same reason that it would result in a disconnected graph. (Figure shown above)

□

Part 3: Coding: Shortest Paths

8. Submit the following:

   (a) In graph.py, implement (and turn in) a function create_graph(n,p) that produces an undirected graph with n nodes where each pair of nodes is connected by an edge with probability p.

   (b) Implement a general shortest-path algorithm for graphs, as described in lecture, that works on your graph. In graph.py, include a function shortest_path(G,i,j) that outputs the length of the shortest path from node i to j in your graph G. Make sure to handle the case where the graph is disconnected (i.e. no shortest path exists) by outputting "infinity".

   (c) Construct a graph for $n = 1000$ and $p = 0.1$. Estimate the average shortest path between a random pair of two (connected) nodes in the graph. For accuracy, repeat for 1000 random pairs of nodes in your graph. Output an execution trace in avg_shortest_path.txt containing all path lengths written as (i, j, length).

   (d) For $n = 1000$, run the shortest-path algorithm on data sets for many values of p (for instance, 0.01 to 0.04 using .01 increments, and then 0.05 to 0.5 using .05 increments). Turn in your numerical data as varying_p.txt, and plot the average shortest path as a function of p and submit as an image file varying_p.(image extension) or include in your main .pdf file.

   Note: For $p = 0.01$ there is actually a small but reasonable chance (around 4%) to produce a disconnected graph. If this occurs, resample and produce a connected graph for the purposes of gathering data.

   (e) Intuitively explain the behavior of the data you found; specifically, as p increases (in particular, look at the larger values, e.g. 0.3 and above), what function does the average shortest path length seem to asymptotically approach and why?

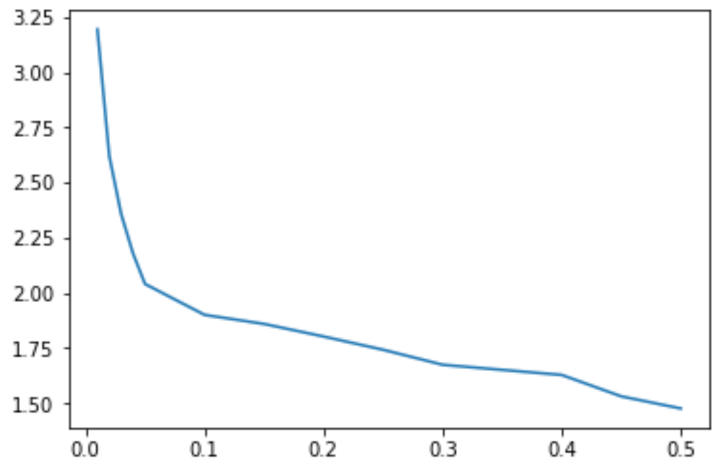
Solution:

(a) (Implementation included in graph.py)

(b) (Implementation included in graph.py)

(c) (Implementation included in graph.py and execution trace in avg_shortest_path.txt)
Final average length: 1.902

(d) (Implementation included in graph.py, data in varying_p.txt, and graph in varying_p.png and also shown on next page.)



x-axis = p
y-axis = A

(e) As p increases, the function asymptotically approaches f(x) = 1. This is intuitively explained as the larger the probability of any two nodes having an edge, the shorter their shortest path, and in turn, the shorter the average shortest path of the entire graph. This function also resembles:

$$A = e^{\frac{1}{p}}$$

or

$$A = 1 + \frac{1}{p}$$

if estimated to 1st position in Taylor series.

$\square$

9. Now run your code on the Facebook social network data available at:
http://snap.stanford.edu/ data/egonets-Facebook.html

(In particular, please refer to the file "facebook_combined.txt.gz"; the data is formatted as a list of undirected edges between 4,039 nodes, numbered 0 through 4038. You will need to parse this data as part of your code; knowing how to do this will be useful for subsequent assignments!)

(a) Repeat the same analysis as in part 8(c) (i.e. run your algorithm on 1000 random pairs of nodes and determine the average shortest path length). Include your code in graph.py and include an execution trace in fb_shortest_path.txt

(b) For the Facebook data, estimate the probability p that two random nodes are connected by an edge. Explain how you computed p.

(c) Is the average shortest path length of the Facebook data greater than, equal to, or less than you would expect it to be if it were a random graph with the same number of nodes and value of p? (To answer this, you may wish to run your code from question (8c) using the p you determined in part (9b) and 4039 nodes.) Explain why you think this is the case.

Solution:

(a) (Implementation included in graph.py and execution trace in fb_shortest_path.txt.)
Final average length: 3.931

(b) From 8(e) above, using A as a function of p as:

$$A = e^{\frac{1}{p}}$$

We get:

$$p = \frac{1}{ln(A)}$$

Taking A from 9(a), we get p = 0.7305

(c) Intuitively, one would expect A to be higher than the real facebook data. This is because life and society are the only things that defy the second law of thermodynamics that entropy always increases. Intuition would suggest there is less randomness in the network than what we computed, and more order. So, real facebook network data must have organised itself in tighter graphs, more dense than the randomly generated graph, causing A(real) to be less that A(computed). This however, can be falsified by the actual computation of A over a randomly generated graph with the same number of nodes and similar p. Running the code from 8(c) using p from 9(b), A' was computed 10 times and averaged.
mean A' was found to be 1.266 $<$ 3.931 [real A from 9(a)] This must mean that along with a dense component, there are many tails of the real world data increasing the average distance of the network by a large coefficient.

□