Examination: IDS 564 (Social Media and Network Analytics) Spring 2020
Instructor: Prof. Ali Tafti

You must submit the answer spreadsheet by the deadline indicated on the Blackboard page for this exam. The spreadsheet to fill is also provided for you on the Blackboard page for this exam. This exam has **51 questions.**

Instructions:

Please select the best answer. In some cases, you may find that more than one answer can be justified. In those cases, you need to select what you believe is the ***best answer***. Please indicate the correct answer as a capital letter A, B, C, D or E, in the second column of the row next to the corresponding question number on the provided spreadsheet. **You must submit the spreadsheet file by the deadline.** Please do not alter the formatting of the spreadsheet or enter any information other than the answers (a single capital letter) for all questions.

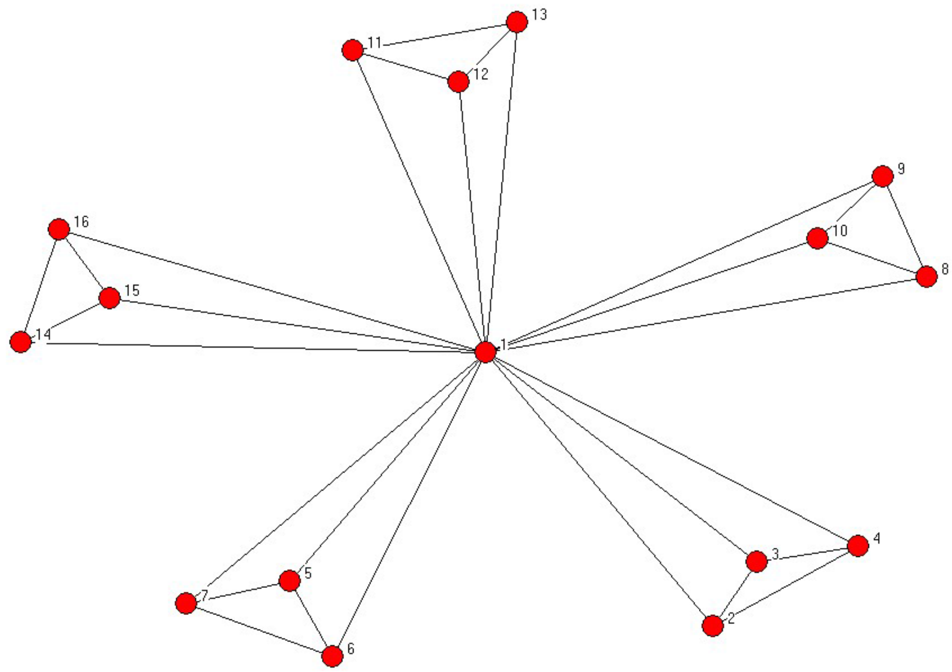
Statement: (Please acknowledge that you have agreed to this statement when you submit the exam answers.)

I acknowledge this is an *open book, open notes and open computer exam*. I will not communicate about the exam with any other persons except for the instructor of the course. *The instructor will not assist in interpreting or clarifying the questions of this exam.*

I understand that using the R igraph library may produce misleading results for some of the graphs in this exam (perhaps due to default settings), and so I will calculate metrics such as PageRank, hub scores, and authority scores manually or with a spreadsheet. Note that some of the questions may appear to be repetitive but may (or may not) in fact be different. Please read each question carefully, and do not assume that it is the same as a previous one that appears similar!

In submitting this exam, I attest to be fully compliant with academic integrity policies of the UIC College of Business. I have not intentionally used or attempted to use unauthorized information, people, or study aids in any academic exercise during the exam. I have not and will not provide or receive from another person, any kind of unauthorized assistance on this examination.

1) Which of the following statements is true?
    A. Every cycle is a walk.
    B. Every walk is a cycle.
    C. Every walk is a path.
    D. Every cycle is a path.
    E. A & C only.



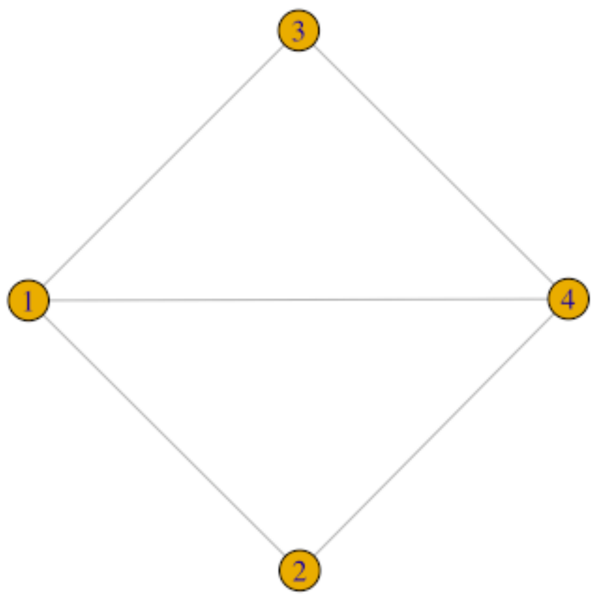
2) Consider the network diagram above that has a star configuration. Which of the following statements is true of a network with this configuration, as the number of branches changes (but the number of nodes in each branch remain the same)?
    A. The overall (global) clustering coefficient will increase as the number of branches increases.
    B. The overall (global) clustering coefficient will decrease as the number of branches increases.
    C. The aggregate average of local clustering coefficients will increase as the number of branches increases.
    D. The aggregate average of local clustering coefficients will decrease as the number of branches increases.
    E. Both the aggregate average of local clustering and the overall (global) clustering coefficient will increase as the number of branches increases.

3) Consider the network diagram above that has a star configuration. Which of the following statements is true of a network with this configuration, as the number of nodes in each branch changes (but the number of branches remains the same)?
   A. The overall (global) clustering coefficient will increase as the number of nodes in each branch increases.
   B. The overall (global) clustering coefficient will decrease as the number of nodes in each branch increases.
   C. The aggregate average of local clustering coefficients will increase as the number of nodes in each branch increases.
   D. The aggregate average of local clustering coefficients will decrease as the number of nodes in each branch increases.
   E. Neither the aggregate average of local clustering nor the overall (global) clustering coefficient will increase as the number of nodes in each branch increases.

4) Which of the following statements is true about cliques within a connected graph?
   A. An edge can belong to two different cliques.
   B. A node can belong to two different cliques.
   C. A graph can have more cliques than nodes.
   D. A & B only.
   E. A & C only.

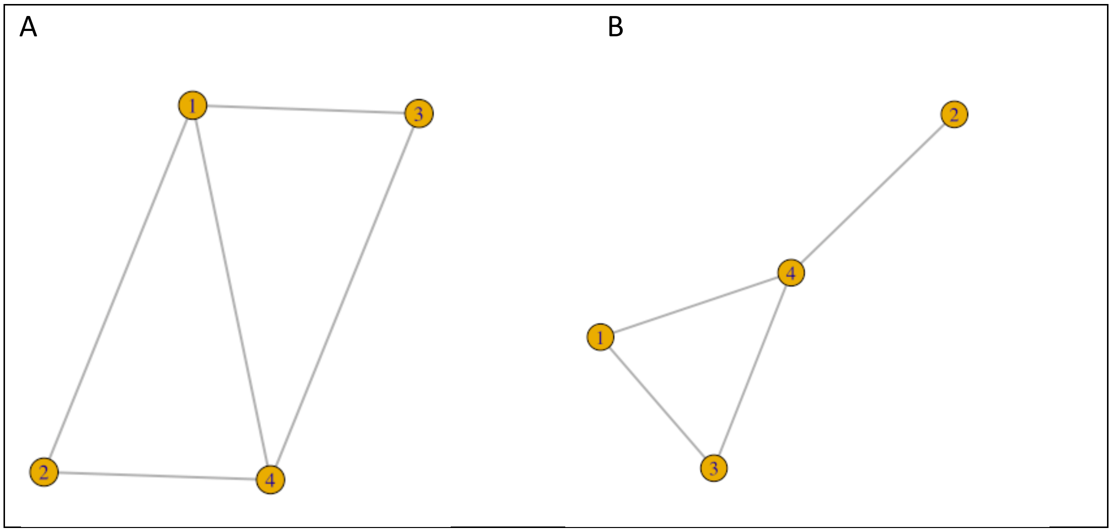
5) Consider nodes A, B and C, among any number of other nodes in a larger set. Which of the following is an advantage of using partial correlations rather than bivariate correlations to infer how nodes are connected in a graph?
   A. Partial correlations help to better distinguish a causal influence relationship between A and B, from an influence due to a third node C that jointly affects both A and B.
   B. Partial correlations help us to better infer the direction of causation between two nodes; for example, to distinguish whether A affects B or whether B affects A.
   C. Partial correlations are more parsimonious because they require data on only a partial number of nodes.
   D. A & B only.
   E. All of the above.

6) Which of the following describe a situation in which the basic network cascade model (as covered in Ch. 19 of the Easley-Kleinberg textbook) would be realistic?
    A. Investment in a technology A or B is costly and difficult to reverse.
    B. Agents need to commit to investment in either A or B, because investing in a hybrid combination of A and B would be infeasible.
    C. The value of choosing A increases with the number of friends using A, while the value of choosing B increases with the number of friends using B.
    D. All of the above.
    E. A & C only.



7) In the graph above consisting of four nodes, which of the following is true of the non-normalized betweenness values?
    A. Nodes 1 and 4 have betweenness centrality values of 0.
    B. Nodes 2 and 3 have betweenness centrality values of 0.
    C. Nodes 1 and 4 have betweenness centrality values of 0.25.
    D. Nodes 2 and 3 have betweenness centrality values of 0.25.
    E. B & C only.

8) In the same graph above consisting of four nodes, which of the following is true of the non-normalized betweenness values?
    A. Nodes 1 and 4 have betweenness centrality values of 0.25.
    B. Nodes 2 and 3 have betweenness centrality values of 0.25.
    C. Nodes 1 and 4 have betweenness centrality values of 0.5.
    D. Nodes 2 and 3 have betweenness centrality values of 0.5.
    E. B & C only.



9) For the left-hand side graph A above, which of the following is true of the edge connecting nodes 3 and 4?
   A. Its neighborhood overlap is 0.
   B. Its neighborhood overlap is ¼ (or 0.25).
   C. Its neighborhood overlap is 1/3.
   D. Its neighborhood overlap is ½ (or 0.5).
   E. Its neighborhood overlap is 1.

10) For the left-hand side graph A above, which of the following is true of the edge connecting nodes 1 and 2?
   A. Its neighborhood overlap is 0.
   B. Its neighborhood overlap is ¼ (or 0.25).
   C. Its neighborhood overlap is 1/3.
   D. Its neighborhood overlap is ½ (or 0.5).
   E. Its neighborhood overlap is 1.

11) For the right-hand side graph B above, which of the following is true of the edge connecting nodes 3 and 4?
   A. Its neighborhood overlap is 0.
   B. Its neighborhood overlap is ¼ (or 0.25).
   C. Its neighborhood overlap is 1/3.
   D. Its neighborhood overlap is ½ (or 0.5).
   E. Its neighborhood overlap is 1.

12) For the right-hand side graph B above, which of the following is true of the edge connecting nodes 2 and 4?

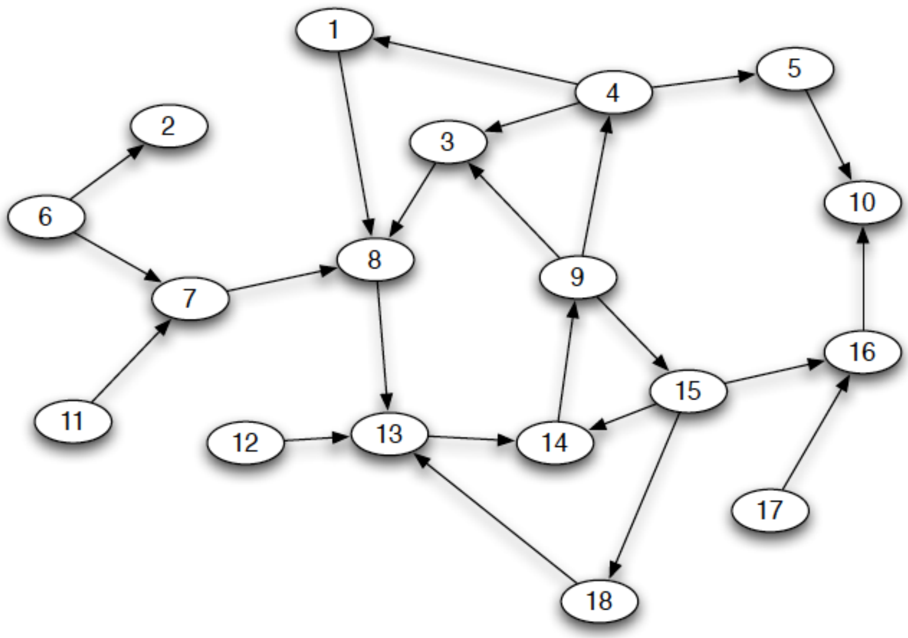
A. Its neighborhood overlap is 0.
B. Its neighborhood overlap is ¼ (or 0.25).
C. Its neighborhood overlap is 1/3.
D. Its neighborhood overlap is ½ (or 0.5).
E. Its neighborhood overlap is 1.

13) In the network below, which of the following is a list of nodes within the strongly connected component?

A. 1, 4, 9, and 14
B. 1, 4, 8, and 16
C. 3, 4, 9 and 17
D. 1, 4, 7, and 8
E. 1, 3, 12, and 18

14) For the network below, which of the following explains why node 10 is in the *out* component rather than in the strongly connected component? Note that A, B and C are all true statements, but they may not all be relevant to answering the question.

A. There are no directed paths from node 10 to nodes in the strongly connected component.
B. There exist directed paths from nodes in the strongly connected component to node 10.
C. Node 10 does not have any outgoing edges.
D. All of the above.
E. A and B only.



15) Which of the following statements is true of networks that are generated through a preferential attachment process, where higher degree nodes are more likely to attract subsequent connections?

A. They exhibit a degree distribution with exponential decay.
B. They exhibit long tails, with a greater density of both very low-degree and very high-degree nodes than networks generated randomly.
C. They exhibit a scale-free degree distribution.
D. All of the above.
E. B and C only.

16) Consider a model of network growth where, in each round, a new node joins the network by first choosing a node randomly among the existing set of nodes in the network. Rather than forming a link with that node, the new node instead picks randomly among the neighbors linked to that node. What kind of network does this process generate?

A. A preferential attachment network, exhibiting a scale-free degree distribution.
B. A preferential attachment network, exhibiting an exponentially decreasing degree distribution.
C. A random growth network, exhibiting an exponentially decreasing degree distribution.
D. A random growth network, exhibiting a Poisson degree distribution.
E. A random growth network, exhibiting a scale-free degree distribution.

17) Consider a model of network growth where, in each round, a new node joins the network by choosing a node randomly among the existing set of nodes in the network, and forming a link with that node. What kind of network does this process generate?

A. A preferential attachment network, exhibiting an exponentially decreasing degree distribution.
B. A preferential attachment network, exhibiting a scale-free degree distribution.
C. A random growth network, exhibiting an exponentially decreasing degree distribution.
D. A random growth network, exhibiting a scale-free degree distribution.
E. None of the above.

18) Consider a model of network growth where, in each round, a new node joins the network by choosing *m* nodes randomly among the *t* nodes that already exist in the beginning of the round. Thus, each pre-existing node has a probability of *m/t* of getting a new link in each round. Which of the following represents the expected degree at time *t* of a node that was born in round *i*:

A. m/(i+t)
B. m/(i-1) + m/(i - 2) + ... + m /(i – t)
C. m(log(i/t)+1))
D. m(1-log(i/t))
E. None of the above

19) Which of the following statements is correct about the random network growth model described in the prior question, at time *t=50* and with *m=1*?

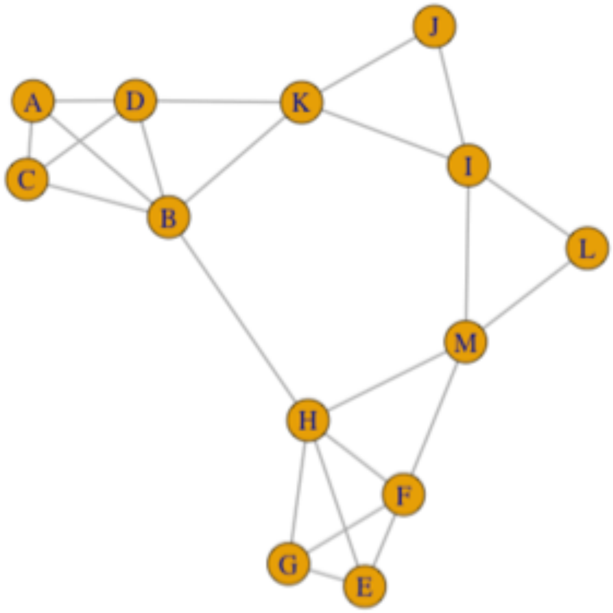
A. A node born at time period *i* has an expected degree of 1 + 1/(i+1) + 1/(i+2) + ... + 1/50.
B. The number *i* equals the number of nodes with expected degree greater than or equal to 1+log(50/*i*).
C. The fraction of nodes with expected degree less than some value *d* is 50 e^{1-d}.
D. All of the above.
E. A and B only.

20) The discovery of the Western hemisphere and subsequent arrival by European explorers led to a sudden, dramatic, and ultimately catastrophic cataclysm experienced by native civilizations in the Western hemisphere. Which of the following explain why the technology and diseases of one civilization quickly and disastrously overwhelmed the other?

A. The global social network transformed from a weakly connected component to a strongly connected component.
B. The global social network contained two components that merged into one.
C. The global social network rapidly turned into a small-world network.
D. The global social network experienced a sudden increase in homiphily.
E. All of the above.

21) Consider two separate networks, the first with *n* nodes and the second with *m* nodes. Suppose that the probability of link formation after a period of time between any specific node in the first network with any specific node in the second network is *p.* What is the probability that the two networks remain separated from each other as components after this period of time?

A. $p^{(m)(n)}$
B. $1 – p^{m} – p^{n}$
C. $1 – (1 – p)^{(m)(n)}$
D. $1 – p^{(m)(n)}$
E. $(1 – p)^{(m)(n)}$



22) Which of the following is true in the graph above?

A. ~~Node M has a clustering coefficient of 1/4.~~

B. Nodes L and J each have a clustering coefficient of 2/3.

C. ~~The edge connecting H and B is a local bridge of span 1.~~

D. The edge connecting H and B has a higher betweenness centrality than the edge connecting B and C.

E. ~~All of the above~~.

23) Which of the following describes a scenario in which the graph above violates the strong triadic closure property?

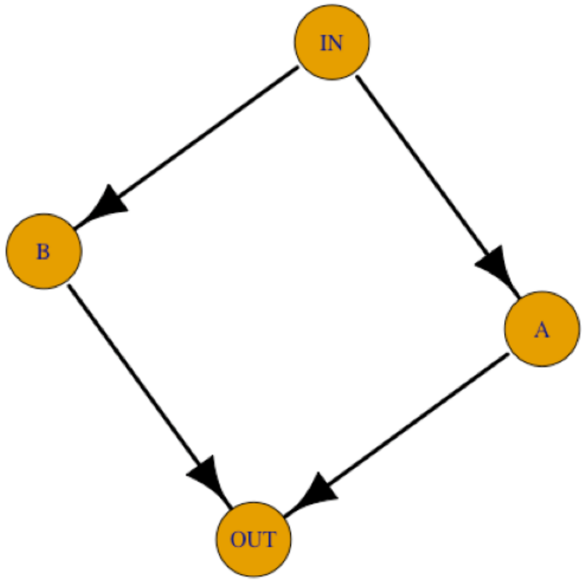
A. Node M maintains weak ties with all of its neighbors.

B. H and B maintain only weak ties with all of their neighbors.

C. Except for a weak tie that connects nodes H and B, both nodes maintain strong ties with all of their other neighbors.

D. All of the above

E. None of the above.



24) For the network above, which statement correctly describes the *normalized* values of hub and authority scores? PLEASE NOTE THE FOLLOWING HINTS: ***Hint 1)*** Normalizing hub scores is done by dividing down each hub score by the sum of all hub scores, and normalizing authority scores is done by dividing down each authority score by the sum of all authority scores. ***Hint 2)*** Your calculations will converge to their equilibrium values after just ***two rounds*** of calculation. Also, please ***do not*** rely on the igraph R package, as the default settings produce misleading results for this size of graph. You are better off calculating the answers manually or with a spreadsheet.

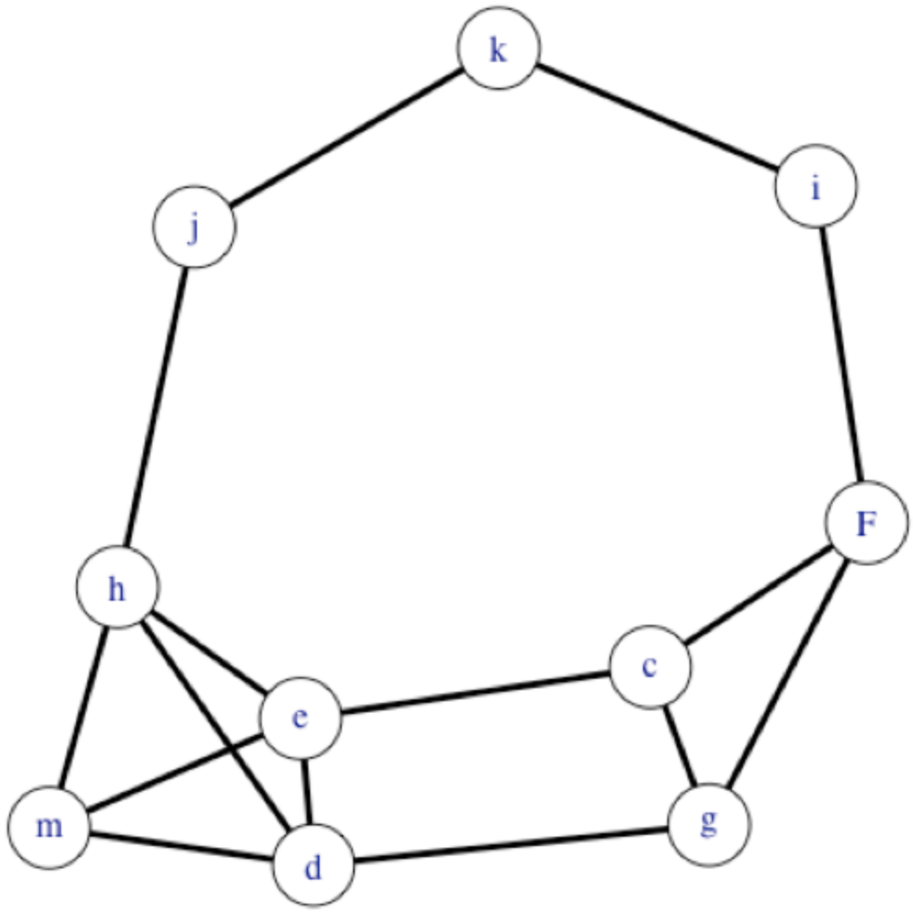
A. Hub scores for IN, A, B and OUT are 0, 1/3, 1/3 and 1/3, respectively. Authority scores for IN, A, B and OUT are 0, 0.5, 0.5 and 1, respectively.
B. Hub scores for IN, A, B and OUT are 1, 0.5, 0.5 and 0, respectively. Authority scores for IN, A, B and OUT are 0, 0.5, 0.5 and 1, respectively.
C. Hub scores for IN, A, B and OUT are 1/3, 1/3, 1/3 and 0, respectively. Authority scores for IN, A, B and OUT are 0, 0.25, 0.25 and 0.5, respectively.
D. Hub scores for IN, A, B and OUT are 0.50, 0.25, 0.25 and 0, respectively. Authority scores for IN, A, B and OUT are 0, 0.25, 0.25 and 0.5, respectively.
E. Hub scores for IN, A, B and OUT are 0, 0.25, 0.25 and 0.50, respectively. Authority scores for IN, A, B and OUT are 0, 0.25, 0.25 and 0.50, respectively.

25) For the network above, which statement correctly describes the *PageRank* scores with scaling factor of 1? ***Hint:*** **PageRank** scores will converge to their equilibrium values after **four rounds of calculation.** Also, please ***do not*** rely on the igraph R package, as the default settings produce misleading results for this size of graph. You are better off calculating the answers manually or with a spreadsheet.

A. PageRank scores for IN, A, B and OUT are 0, 0, 0, and 0, respectively when the scaling factor is 1.
B. PageRank scores for IN, A, B and OUT are 0, 0.25, 0.25, and 0.50 respectively, when the scaling factor is 1.
C. PageRank scores for IN, A, B and OUT are 0, 1/3, 1/3, and 1/3 respectively, when the scaling factor is 1.
D. PageRank scores for IN, A, B and OUT are 0, 0.25, 0.25, and 1 respectively, when the scaling factor is 1.
E. PageRank scores for IN, A, B and OUT are 0, 0, 0, and 1 respectively, when the scaling factor is 1.

26) For the network above, which statement correctly describes the equilibrium values for *PageRank* scores with **scaling factor of 0.8?**

A. PageRank scores for IN, A, B and OUT are 0, 0.15, 0.15, and 0.8 respectively, when the scaling factor is 0.8.
B. PageRank scores for IN, A, B and OUT are 0, 0, 0, and 0.8 respectively, when the scaling factor is 0.8.
C. PageRank scores for IN, A, B and OUT are 0, 0.13, 0.13 and 0.740 respectively, when the scaling factor is 0.8.
D. PageRank scores for IN, A, B and OUT are 1/6, 1/6, 1/6 and 0.131 respectively, when the scaling factor is 0.8.
E. PageRank scores for IN, A, B and OUT are 0.05, 0.07, 0.07, and 0.162 respectively, when the scaling factor is 0.8.

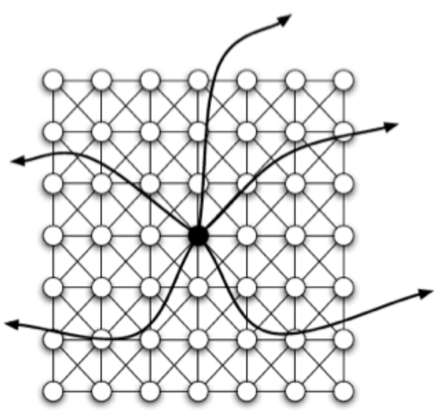
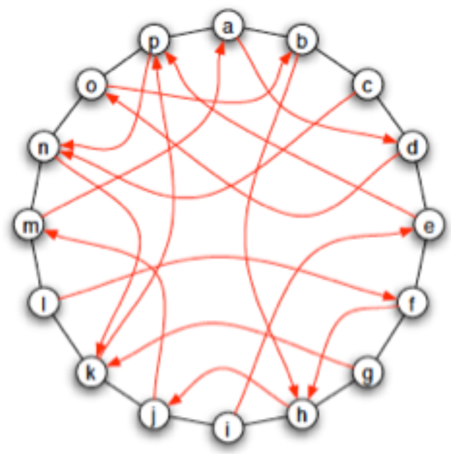


27) Consider the above diagram. Suppose that each node requires that at least one-third of its neighbors switch to behavior B, in order for it to also switch to behavior B (i.e. switching threshold of q = 1/3). Now, suppose that node *F* is the initial adopter of behavior B, while all other nodes start with behavior A. Which of the following nodes is in the set of nodes that eventually adopts B?

A. j
B. e
C. h
D. d
E. None of the above

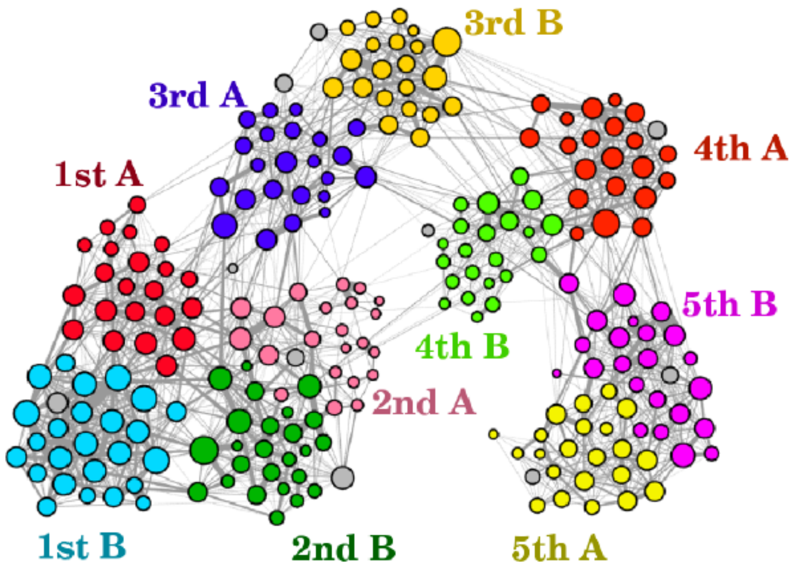
28) Consider a two-dimensional small worlds network as pictured on the left of the diagram below (network A). Nodes are connected to their immediate spatial neighbors in two dimensions. In addition, random connections are made with a certain probability. Which of the following allocation of random edges results in the optimal network for efficient decentralized search?

A. Random edges in Network A are made with probability inversely proportional to the distance.
B. Random edges in Network A are made with probability inversely proportional to the square of distance.
C. Random edges in Network A are made with probability directly proportional to the distance.
D. Random edges in Network A are made with probability directly proportional to the square of distance.
E. None of the above is optimal.

A	B
	

29) Consider a one-dimensional small worlds network as pictured on the right of the diagram above (network B). Nodes are connected to their immediate spatial neighbors in one dimension. In addition, random connections are made with a certain probability. Which of the following allocation of random edges results in the optimal network for efficient decentralized search?

A. Random edges in Network B are made with probability inversely proportional to the distance.
B. Random edges in Network B are made with probability inversely proportional to the square of distance.
C. Random edges in Network B are made with probability directly proportional to the distance.
D. Random edges in Network B are made with probability directly proportional to the square of distance.
E. None of the above is optimal.



30) Consider the network of school children derived by tracking their interactions over the course of a day, as shown in the above figure. Suppose a virus has infected one of the children. After a virus has infected the first child, that child can spread the virus for a period of eight days, and every other child is susceptible to infection. After eight days, the first infected child is no longer spreading the virus, and also enters a stage of immunity for 60 days. Consider a probability of transmission **p** for each link between an infected and susceptible node. Which of the following statements is likely to be true **of the eight days in which the first child has been infected**?

A. The virus spreads more quickly to classrooms with a low ratio of edges in proportion to the number of nodes.
B. Once the virus has spread to a classroom, the rate of infection within the classroom is proportional to the cluster density of that classroom.
C. The virus is unlikely to spread beyond the first child's classroom.
D. The cluster density of each classroom is directly proportional to p.
E. High cluster density serves to inoculate a classroom from the spread of infection.

31) Consider the same network of primary school children above; except now assume that the edges in this network **represent ties of friendship**. On the first day of the academic year in September 1, the school announces that there will be an optional summer field trip in June when classes are finished. All students can view the list of students signed up to attend the trip (the sign-up list), and on September 1 only one student has signed up. That student is in classroom **1B.** Suppose that on each day, other students will add their name to the list **if and only if** at least a fraction **p** of

their friends appear on that list. The sign-up list is updated once each day, until the deadline on May 1. ***Assume the network structure does not change over that time.*** Which of the following statements is most accurate?

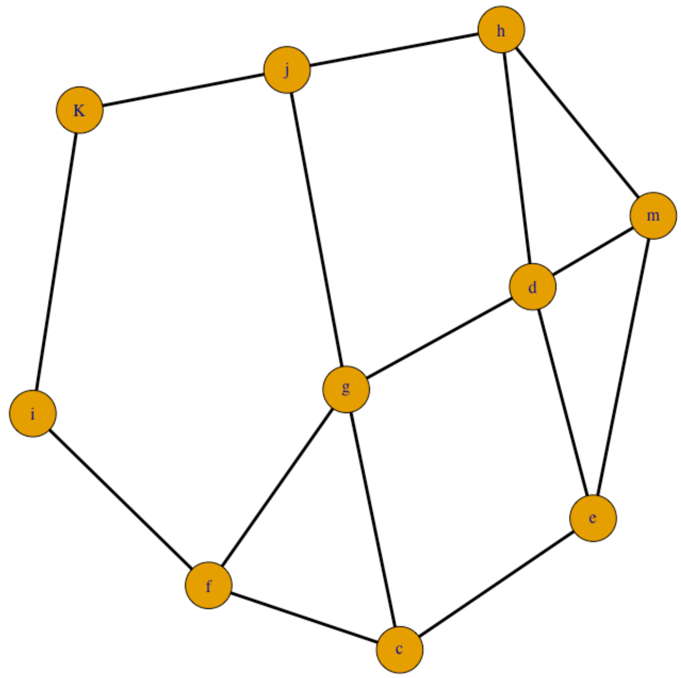
A. By the end of the school year, the number of children within any classroom *other than 1B* signing up to attend is directly proportional to the cluster density of their own classroom.
B. By the end of the school year, the probability that any given child in classroom 5A attends is likely to be greater than the probability that any given child in classroom 1A attends, assuming that classrooms 1A and 5A have the same cluster density.
C. By the end of the school year, the number of children within any classroom *other than 1B* signing up to attend is *inversely* proportional to the cluster density of their own classroom.
D. A & C only.
E. B & C only.

32) For the same scenario as in the prior question, which of the following tasks would be enabled by a standard network community detection algorithm?

A. Identify subsets of children who are likely to make the same choices in a network cascade model, in which students follow what the majority of their friends are doing.
B. Classify children by their classroom designations and not by their positioning within the network.
C. Identify nodes with the highest influence.
D. All of the above.
E. A & C only.

33) Suppose that the operators of a news site are considering changing the way that links are sorted on the front web page of the site. Which of the following ways of sorting links on the front page will result in the popularity distribution of articles to follow a power-law distribution?

A. Sort the links to articles in alphabetical order by the author's last name.
B. Sort the links to articles in order of decreasing frequency of being shared, with the most-widely shared articles listed first.
C. Sort the links to articles in order with the most recent articles appearing first.
D. All of the above.
E. B & C only.



34) Consider the above diagram. Suppose that each node requires that at least half of its neighbors switch to behavior B, in order for it to also switch to behavior B (i.e. switching threshold of q = 1/2). Now, suppose that *K* is the initial adopter of behavior B, while all other nodes start with behavior A. Which of the following nodes is in the set of nodes that eventually adopts B?

A. f
B. i
C. j
D. All of the above
E. None of the above

35) Suppose that the threshold for switching is changed to one-third. Which of the following is true about nodes that switch from behavior A to B when the threshold for switching is changed to one-third (requiring a portion 1/3 of neighbors to switch to B in order to follow suit), with the same initial adopter *K* as in the previous question?

A. These nodes would not cause a complete cascade if and only if the nodes in the remaining network contain a cluster of density greater than 1/3.
B. These nodes would not cause a complete cascade if and only if the nodes in the remaining network contain a cluster of density greater than 2/3.
C. A complete cascade forms with the new threshold.
D. B and C only.
E. None of the above.

36) In one of the lab assignments you predicted links between nodes based on statistically significant correlations in certain activity among the nodes (e.g. gene stimulation, Twitter activity). Which of the following statements is true about the process of link prediction?

A. The p-values for statistical significance need to be adjusted upwards to enable false discovery detection, which can be done using the Benjamini-Hochberg method.
B. We calculate partial correlation coefficients in order to measure the correlation that remains between nodes after adjusting for their correlation with every other node in the graph.
C. Fischer’s transformation is used to obtain confidence intervals for the partial correlations.
D. All of the above.
E. B and C only.

37) Together with some anthropologists, you're studying a sparsely populated region of a rain forest, where 50 farmers live along a 50-mile-long stretch of river. Each farmer lives on a tract of land that occupies a 1-mile stretch of the river bank, so their tracts exactly divide up the 50 miles of river bank that they collectively cover. The farmers all know each other, and after interviewing them, you've discovered that each farmer is friends with all the other farmers that live at most 20 miles from him or her, and is enemies with all the farmers that live more than 20 miles from him or her. Based on this information, which of the following could you conclude about a signed complete graph corresponding to this social network?

A. This network would violate the *weak* form of structural balance.
B. This network would violate the *strong* form of structural balance.
C. This network would violate the principle of *strong triadic closure*.
D. A and B only.
E. All of the above.

38) Consider a network, drawn from the United States population, in which each person has a directed edge to anyone they know on a first-name basis. Which of the following describes the myopic search process that was described in the Travers and Milgram (1969) chain mail experiment and later modeled algorithmically as a *small-world network search process*? [1]

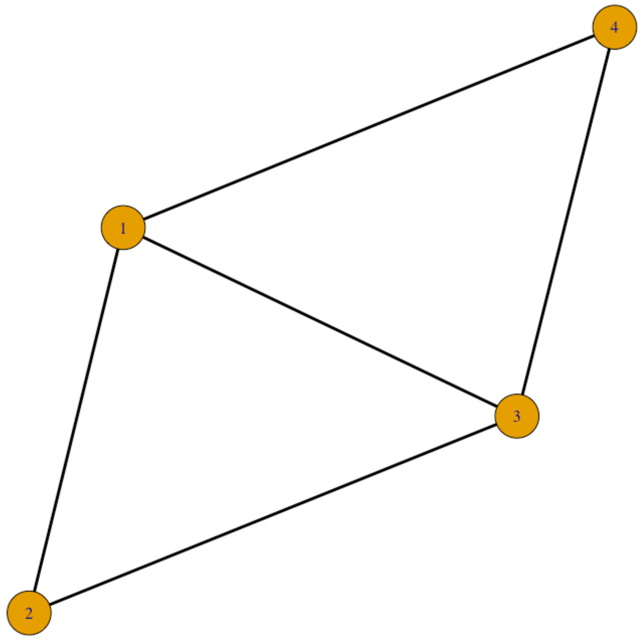
Travers, Jeffrey, and Stanley Milgram. "An experimental study of the small world problem." *Sociometry* (1969): 425-443.

[1] Travers, Jeffrey, and Stanley Milgram. "An experimental study of the small world problem." *Sociometry* (1969): 425-443.

A. An intended target is reached over an average of six degrees because some individuals are able to infer a map of the overall social network.
B. Individuals can send messages only within a specified radius in terms of geographical distance.
C. The model assumes that individuals act upon *complete information* to infer who among their direct neighbors has a smaller shortest-path (i.e. geodesic) distance to the target recipient.
D. The model assumes that individuals act upon *incomplete information* to infer who among their direct neighbors has a smaller shortest-path (i.e. geodesic) distance to the target recipient.
E. All except for D are correct.

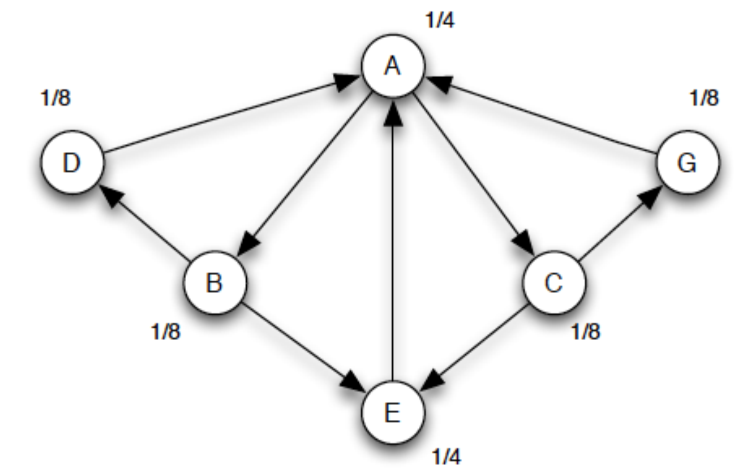
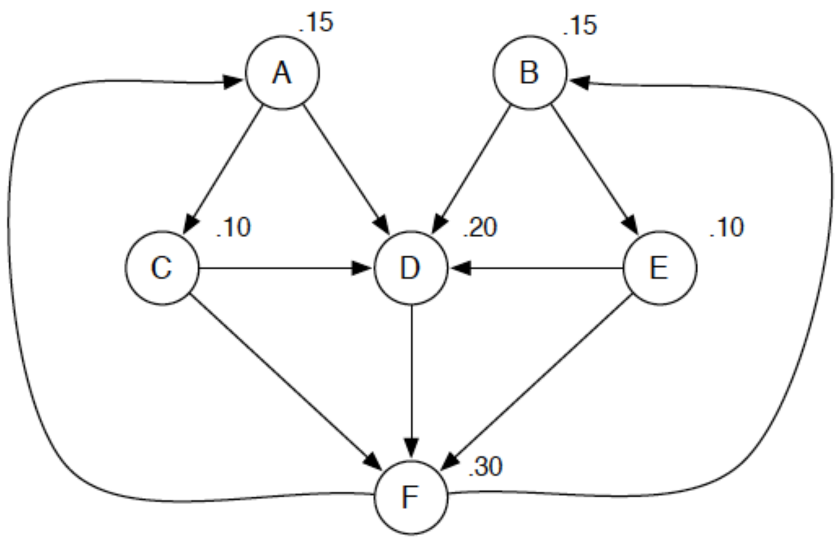
39) What can you conclude about this network, based upon the conclusions of the Travers and Milgram (1969) chain mail experiment?

A. For any randomly selected pair of people, it is very unlikely that there would be a path of at most six edges between them.
B. For any randomly selected pair of people, it is likely that there would be a path of at most six edges between them.
C. This network would exhibit the *strong* form of structural balance.
D. All of the above.
E. B and C only.



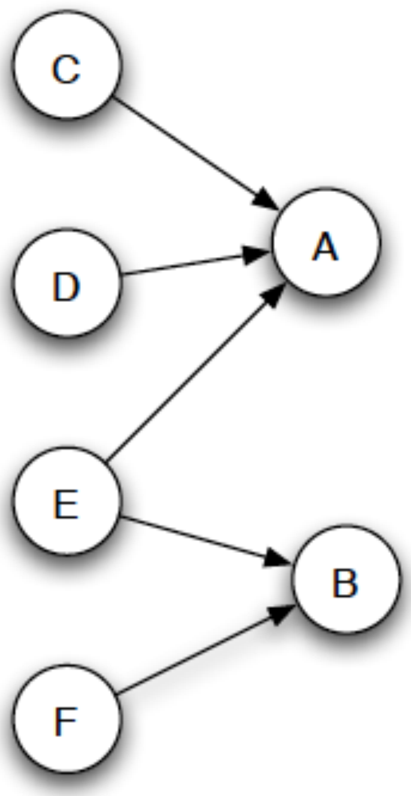
40) Which of the following is true of the above graph?

A. The clustering coefficient for nodes 1 and 3 is equal to 2/3.
B. The clustering coefficient for nodes 2 and 4 is equal to 1/3.
C. The clustering coefficient for nodes 2 and 4 is equal to 2/3.
D. A and B only.
E. None of the above.

A	
B	

41) Assuming a scaling factor of 1, which of the following statements is true about the assignment of PageRank values to nodes in the two network diagrams A and B above?

A. Figure A shows correct equilibrium values for the Basic PageRank Update Rule, while Figure B does not.
B. Figure B shows correct equilibrium values for the Basic PageRank Update Rule, while Figure A does not.
C. Neither Figure A nor Figure B show the correct equilibrium values for the Basic PageRank Update Rule.
D. Both Figure A and Figure B show the correct equilibrium values for the Basic PageRank Update Rule.
E. None of the above.



42) For the network above, which statement correctly describes the values after running **three** rounds of computing hub and authority values, before the final normalizing step? NOTE: We define a round as first computing authority scores, then computing the corresponding hub scores.

A. Hub scores for the left-hand nodes C, D, E and F are 3, 3, 5 and 2, respectively. Authority scores for the right-hand side nodes A and B are 3 and 2, respectively.
B. Hub scores for the left-hand nodes C, D, E and F are 4, 4, 6 and 2, respectively. Authority scores for the right-hand side nodes A and B are 14 and 8, respectively.
C. Hub scores for the left-hand nodes C, D, E and F are 11, 11, 18 and 7, respectively. Authority scores for the right-hand side nodes A and B are 11 and 7, respectively.
D. Hub scores for the left-hand nodes C, D, E and F are 10, 10, 16 and 22, respectively. Authority scores for the right-hand side nodes A and B are 36 and 38, respectively.
E. Hub scores for the left-hand nodes C, D, E and F are 40, 40, 65 and 25, respectively. Authority scores for the right-hand side nodes A and B are 40 and 25, respectively.

43) The research study by Onnela, et al (2007) (citation below), which examines communication patterns among millions of mobile phone users, finds that strength of a link is inversely correlated with its betweeness centrality. Strength of a link connecting two nodes represents the amount of phone conversation between them.

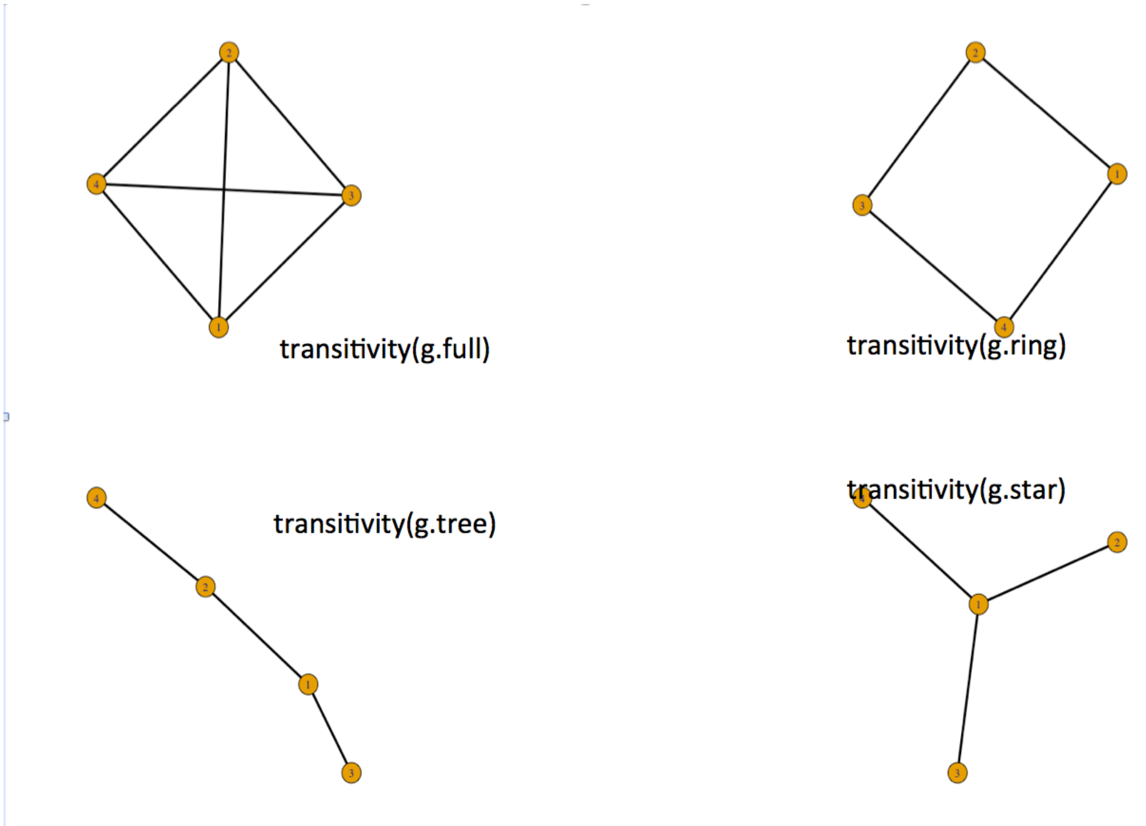
Onnela, J-P., Jari Saramäki, Jorkki Hyvönen, György Szabó, David Lazer, Kimmo Kaski, János Kertész, and A-L. Barabási. "Structure and tie strengths in mobile communication networks." *Proceedings of the National Academy of Sciences* 104, no. 18 (2007): 7332-7336.

Which of the following would explain such a pattern?

A. Complex networks organize themselves according to a global efficiency principle, meaning that the tie strengths are optimized to maximize the overall flow of information in the network.
B. The strength of a particular tie depends on the nature of the relationship between two individuals.
C. The strength of a tie between two nodes increases with the overlap of their friendship circles, resulting in the importance of weak ties in connecting communities.
D. All of the above.
E. A and C only.

44) Which of the following is a correct statement about the *weak* form of structural balance in networks?

A. Any triangles must have two negative edges and one positive edge.
B. Any triangles must have exactly one or three negative edges.
C. No cycle in the network can have an odd number of negative edges.
D. Nodes in the network can be separated into no more than two different factions, where each pair of nodes within the same faction are connected by friendship (with a positive edge), and any pair of nodes across different factions are enemies (with a negative edge).
E. None of the above.



45) The R function transitivity() returns the overall global clustering coefficient for a graph. Consider the 4-node graphs depicted above with their corresponding names: g.full, g.ring, g.tree, and g.star. Which of the following statements is true for the transitivity() calculation?

A. Its value is 1 for g.full.
B. It returns no defined value (or NaN) for g.tree and g.star.
C. Its value is 0 for g.ring, g.star and g.tree.
D. A and B only.
E. A and C only.

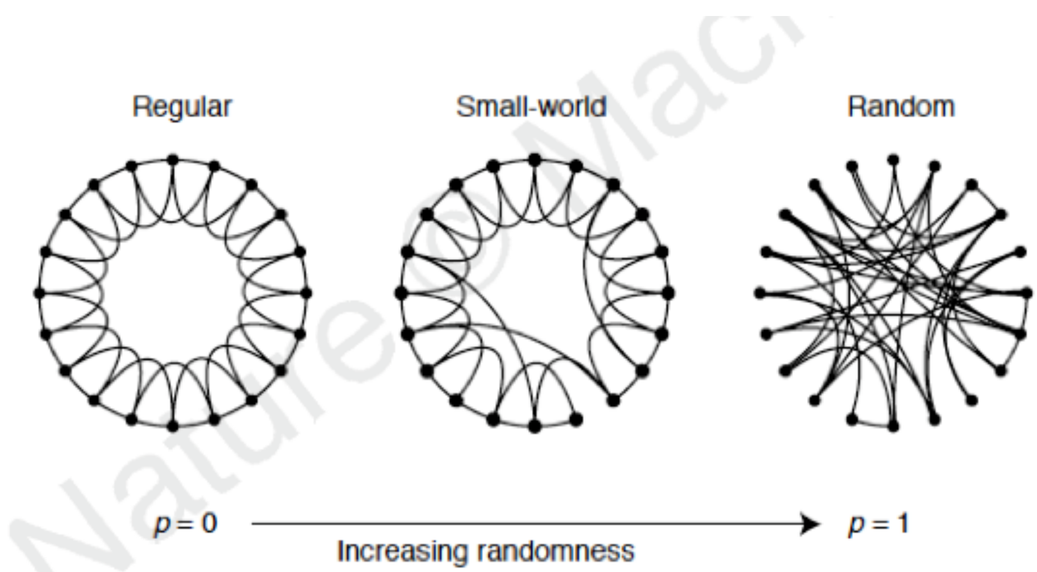
46) To obtain the mean value of the clustering coefficients for all nodes in a graph, you can run the transitivity() function with the type=”local” option, and then obtain the mean; as in the following for a graph g:

```
mean(transitivity(g, type="local"))
```

Please note that if any undefined (NaN) values are part of the list of inputs to the mean() function, the result is undefined. For the same diagram set in the previous question: Which of the following is true for this calculation?

A. Its value is 1 for g.full.
B. It returns no defined value (NaN) for g.tree and g.star.
C. Its value is 0 for g.ring, g.star and g.tree.
D. A and B only.
E. A and C only.

Consider the set of three graphs below, from Watts and Strogatz (1998). The left-most graph is a ring of *n* vertices, each connected to four of its nearest neighbors by undirected edges. With some probability *p,* the node of one side of each edge is reconnected to a different vertex chosen at random over the entire ring.



Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of ‘small world’networks. *Nature*, *393*(6684), 440.

47) Which of the following is true of the small-world graph depicted at intermediate values of *p*, as compared to the regular lattice depicted at *p*=0*.*

A. The small-world network features *slightly lower* average clustering coefficients and much *lower* characteristic path lengths than the regular lattice.
B. The small-world network features much *higher* average clustering coefficients and *slightly lower* characteristic path lengths than the regular lattice.
C. The small-world network features much *lower* average clustering coefficients and *slightly higher* characteristic path lengths than the regular lattice.
D. The small-world network features much *higher* average clustering coefficients and *slightly higher* characteristic path lengths than the regular lattice.
E. None of the above.

48) As compared to the random network depicted in the right-most diagram above, what happens to a network rewired in the scheme described above, as the parameter of p is gradually decreased from one (in the direction of decreasing randomness) as the value of p enters the region of the small-world network?

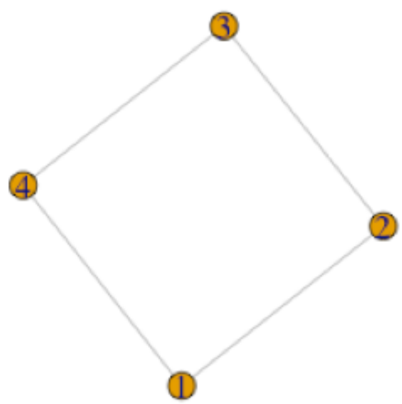
A. Characteristic path length and average clustering coefficient both decline *rapidly*, in comparison to the random network.
B. Characteristic path length declines *rapidly* while average clustering coefficient declines *gradually*, in comparison to the random network.
C. Characteristic path length increases *gradually* while average clustering coefficient increases *rapidly*, in comparison to the random network.
D. Characteristic path length increases *rapidly* while average clustering coefficient *increases gradually*, in comparison to the random network.
E. None of the above.

49) Which of the following is true of the small-world graph depicted at intermediate values of p, as compared to the random graph depicted at p=1.

A. The small-world network features much *lower* average clustering coefficients and *slightly lower* characteristic path lengths than the random network.
B. The small-world network features much *higher* average clustering coefficients and *slightly lower* characteristic path lengths than the random network.
C. The small-world network features much *lower* average clustering coefficients and *slightly higher* characteristic path lengths than the random network.
D. The small-world network features much *higher* average clustering coefficients and *slightly higher* characteristic path lengths than the random network.
E. None of the above.

50) As compared to the two-dimensional ring lattice depicted in the left-most diagram above, what happens to a network rewired in the scheme described above, as the parameter of *p* is gradually increased from zero at its low and intermediate values?

A. Characteristic path length and average clustering coefficient both decline *rapidly*, in comparison to the ring lattice.
B. Characteristic path length declines *rapidly* while average clustering coefficient declines *gradually*, in comparison to the ring lattice.
C. Characteristic path length declines *gradually* while average clustering coefficient declines *rapidly*, in comparison to the ring lattice.
D. Characteristic path length declines *rapidly* while average clustering coefficient *increases gradually*, in comparison to the ring lattice.
E. None of the above.



51) What is the *diameter* for the above graph, a ring with four nodes?

A. 0
B. 1/2
C. 1
D. 2
E. None of the above.