LevineMIDTERM EXAMINATION
Networked Life (NETS 112)
October 23, 2018

Problem 1 (10 points). Next to each lettered item on the left, write the number of the item on the right that is the best match. It might help to do the ones that are most obvious to you first.

1. 7 - a
2. 3 - b
3. 6 - c
4. 1 - d
5. 10 - e
6. 4 - f
7. 2 - g
8. 5 - h
9. 8 - i
10. 9 - j

Problem 2 (10 points). Consider the network in which there is a vertex for each word in the English language, and there is an edge between two words if they have appeared together in the same sentence in a New York Times article published in the last year. For example, there would be an edge between the words "dissident" and "journalist" due to a sentence in Saturday's NYT, but it might be quite unlikely there is an edge between (say) "alpine" and "dissident". For each of the universal network properties discussed in class and readings, indicate whether you think this network will exhibit that property, and briefly but clearly justify your answer.

2 points each. -1 if justification is not strong.

1. Giant component - **yes.**
    a. Linking words such as "a", "the" are likely to connect most words, which automatically implies a giant component.
2. Heavy tail distribution - **yes.**
    a. Linking words have a much higher degree distribution than less common words, so the degree follows a heavy-tailed distribution.
    b. Independently, Zipf's law on the frequency of terms in language applies here, so the distribution is probably heavy-tailed.
3. Small diameter - **yes.**
    a. Linking words join most words with chains of length 2-3, which automatically implies a small diameter.
4. High clustering - **yes.**
    a. Topically-related words are more likely to be used together - the example of "alpine", "dissident" and "journalist" illustrates this. Words in the topic cluster of

“conflict” such as dissident and journalist are more likely to be connected to each other than to words in other topic clusters like alpine.
    b. This is independent of linking words, which on their own might not produce clustering (which lead to many mistakes on the exam)
5. Sparsity - **yes.**
    a. Take any individual word, such as “network”. The number of contexts in which “network” might be used in a sentence is much smaller than the universe of possible sentences, so “network” has a very sparse neighborhood.
    b. This is true for every word we can think of (with the exception of linking words), so the network is sparse.

Problem 3 (15 points). Consider the network shown below.
(a) Ignoring the dotted line, calculate the clustering coefficient of this network.

(6+4 + 2(1/6) + 2(3/10)) / 14

4 points
(b) Calculate the overall edge density of this network.
22/(14 choose 2) = 22/(7*13)=22/91
2 points

(c) Calculate the edge density on just the left side of the dotted line, and also on just the right side of the dotted line. Exclude the edges crossing the dotted line, as they are neither on the left side nor the right side.

10/21 for both

3 points
(d) Calculate the edge density between the vertices on the left and right sides of the dotted line.
2/49
3 points

(e) Based on your calculations above and class discussions, would you say this network is highly clustered or not? Briefly but precisely justify your answer.
Yes, because clustering coefficient >> edge density
3 points - 2 points for saying yes, 1 point for explanation

Problem 4 (15 points). The table below is reproduced from Wikipedia, and was discussed in class. In it, d refers to (worst-case) diameter, and k refers to degree.

a) As clearly and precisely as you can, explain what the numbers in the table are measuring. Do the numbers in the table represent empirical measurements, or are they from mathematical Theories?

5 points

The maximum number of nodes you can have with a graph with worst case diameter that is less than or equal to k, and and average degree less than or equal to d. +3 points

Saying that the numbers are mathematical theories +2 points (Ok if they say empirical but explanation shows they know its mathematical exercise)

b) What can you observe about the growth rate of the numbers? Explain why it makes broad sense.

5 points

With respect to k, growth is exponential / very fast. With respect to d, growth is polynomial. If the student only says that the numbers increase, this is ok but needs to specify in both. Should also include that growth with respect to d is not as fast as growth with respect to k (-1 if not). Any explanation that relates the concepts of growth rate of nodes versus increasing diameter and degree (+1). 0/1 if no attempt at an explanation is made.

c) Consider the entry for 7 041 746 081 , which is approximately the size of the world population. Are the values for which this entry is achieved consistent with the readings from class? Justify your answer.

5 points

Yes, need to talk about the context of the number (# of ppl in the world) and what degree / diameter then means. They need to know the bounds of diameter / degree.

Full credit - only if totally correct. Partial credit for having some idea but not fully explained.

Common mistakes:

- Saying no because of Travers and Milgrams experiment (6 degrees). This is average case diameter, while here we are looking at worst-case diameter.
- Mentioning sparsity - to measure sparsity we need to know more information about the edges in the graph, so this is not a valid explanation.

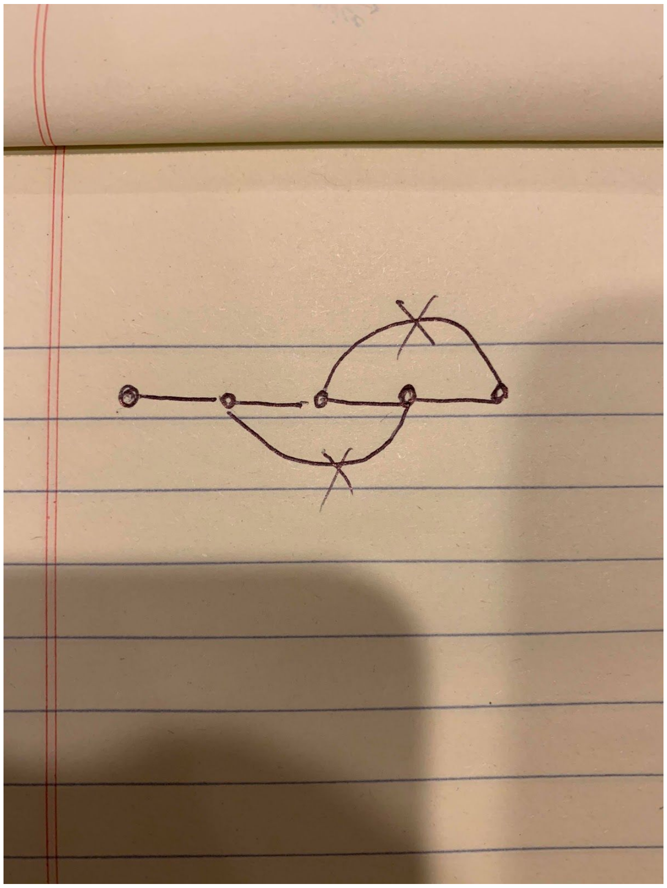
Problem 5 (15 points). Consider the network shown below.

(a) What is its worst-case diameter?

4

+2 if correct

(b) Clearly add two edges so that the worst-case diameter decreases by exactly 1.

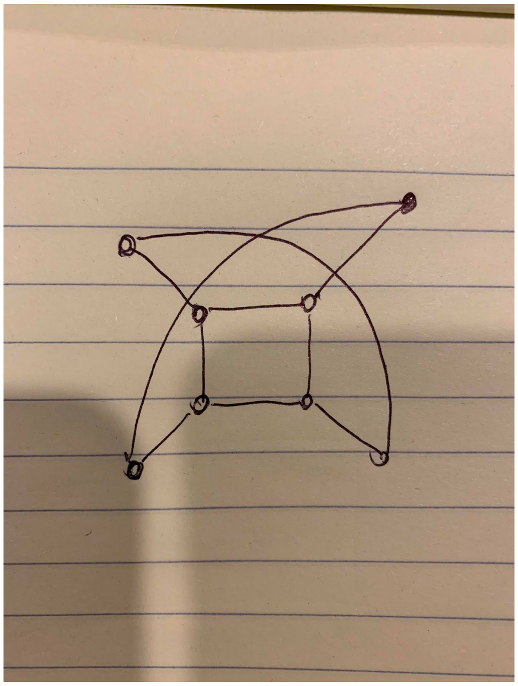


+1 for drawing two edges
+2 if correct (3 / 3)

(c) Draw a network of 5 vertices and 6 edges such that the worst-case diameter is 3, and when you remove 2 edges of your choice (clearly marked on your drawing), the worst-case diameter becomes 4.

+5 for worst case diameter
- +3 for correct worse case diameter
- +2 for having correct number of vertices and edges

+5 for removing two edges that lowers worst-case diameter to 4



Problem 6 (10 points). Fill in the following table by writing either “yes” or “no” in each entry.

__Erdos-Renyi__

Giant Component - yes
Small Diameter - yes
Sparsity - yes
Heavy Tails - no
High Clustering - no

__Preferential Attachment__

Giant Component - yes
Small Diameter - yes
Sparsity - yes
Heavy Tails - yes
High Clustering - no

+1 for each

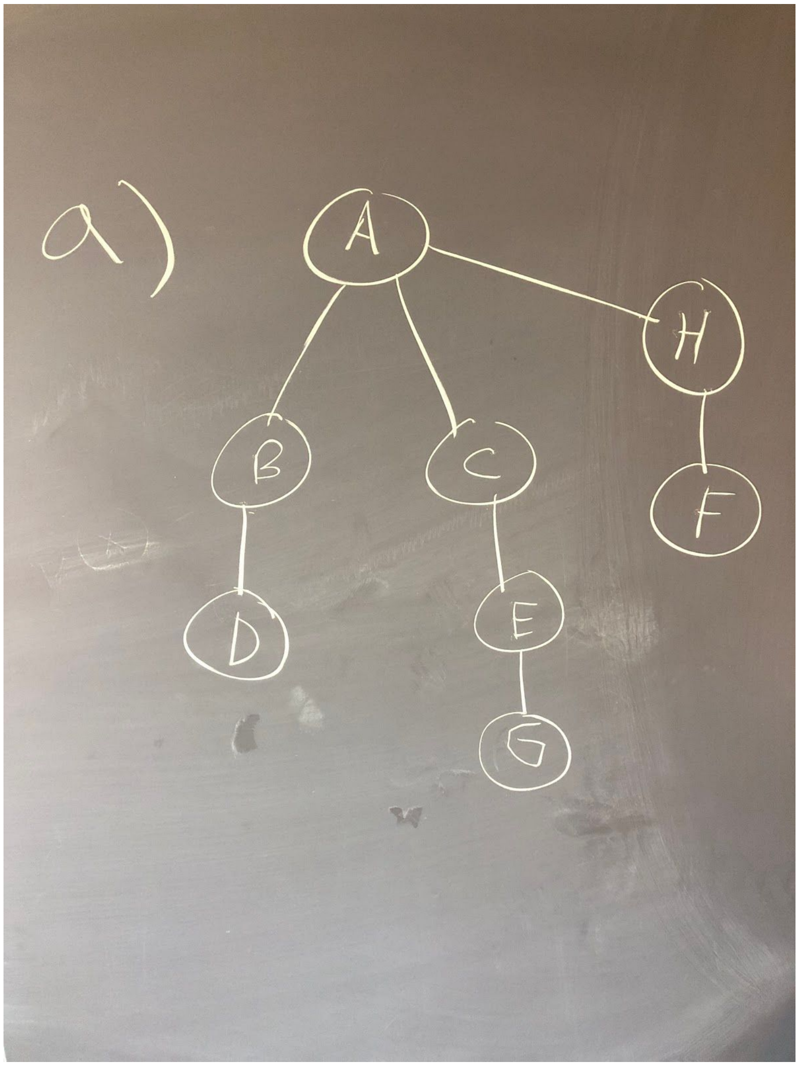
Problem 7 (10 points).
(photos of work at the end of document)
(a) Draw the cascade tree for network above and the given timestamps of when each

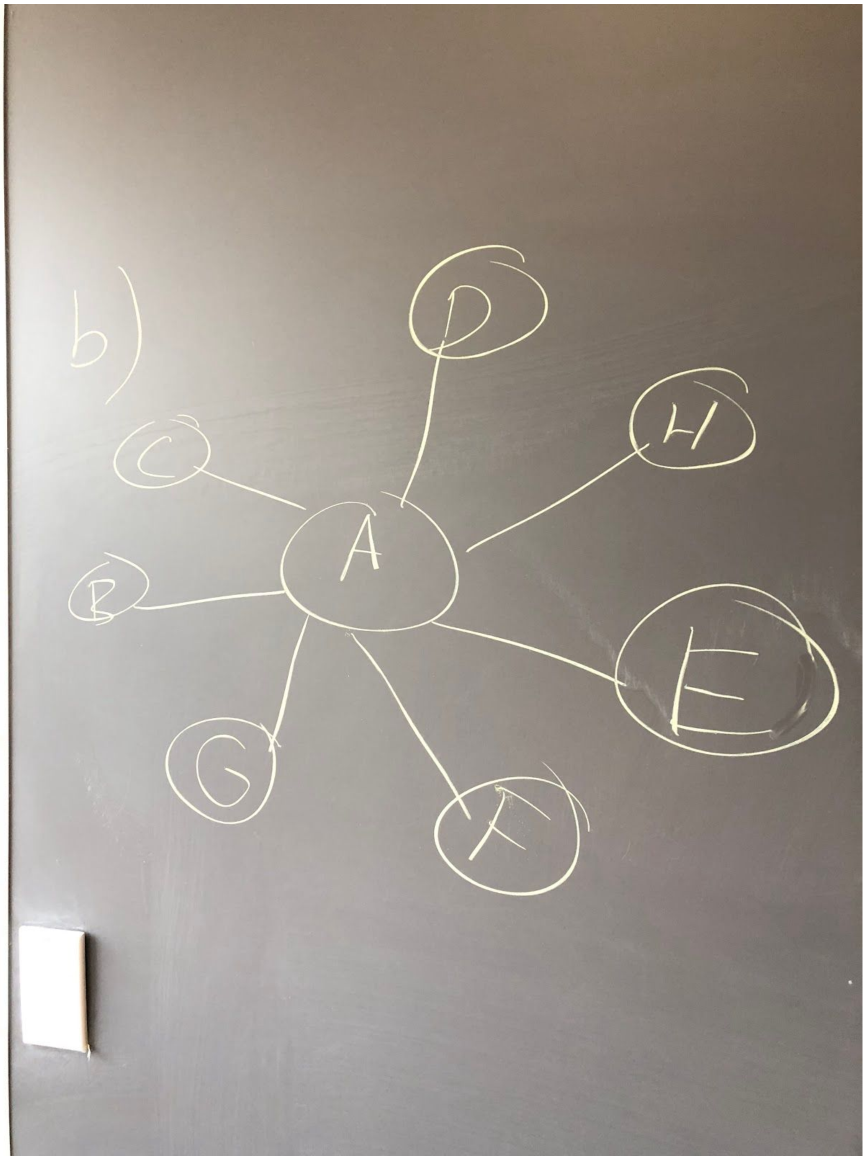
vertex shared. Break ties by attributing the cause of resharing to the neighbor who shared most recently.
+2
(b) Do the same, except break ties by attributing the cause of resharing to the neighbor who shared it first.
+3
(c) Draw the cascade tree for network above and the given timestamps of when each vertex shared. Break ties by attributing the cause of resharing to the neighbor who shared most recently.
+2
(d) Compute the virality of the cascade tree for part (c).
+3
Answer: 29/15
Numerator: sum of all the pairwise distances in the cascade tree
Denominator: How many pairwise distances there are

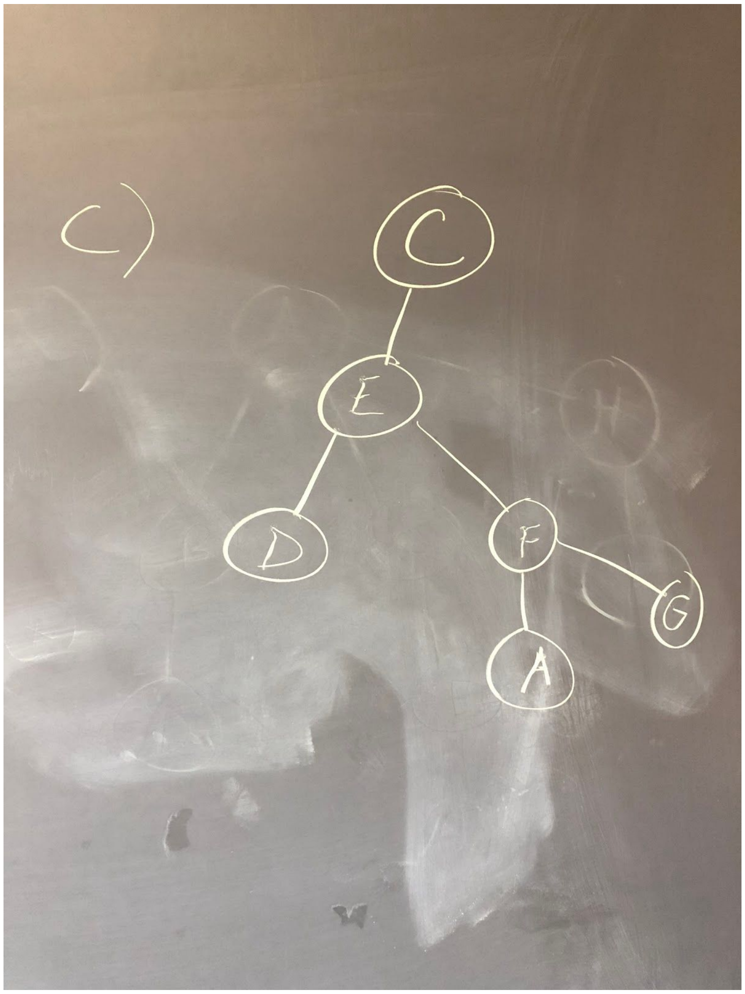
Problem 8 (15 points). One of the assigned readings begins with the following sentence: A crucial task in the analysis of on-line social-networking systems is to identify important people — those linked by strong social ties — within an individual's network neighborhood. Here we investigate this question for a particular category of strong ties, those involving spouses or romantic partners.
(a) Briefly but clearly describe what the source of data for the study was. Be sure to indicate exactly what information was used, and what information was not used.
+2 for saying facebook data
+1 for saying they looked at mutual friends
+2 for saying they did not look at anything except network structure. No demographics or anything

(b) Briefly but clearly describe the question or hypothesis the authors were investigating.
+3 for saying predicting relationships
+2 for saying based only on network structure

(c) Briefly but clearly describe the findings or results of the study.
+3 for talking about successful prediction of relationships based purely on network structure
+2 for differentiating between the effects of dispersion and embeddedness


a)
A
H
B
C
F
D
E
G


b)
D
C
H
A
B
E
G
F


c)
C
E
D
F
G
A

d) $$\begin{aligned} &d_{AC} + d_{AD} \\ &+ d_{AE} + d_{AF} + d_{AG} \\ &+ d_{CE} + d_{CD} + d_{CF} \\ &+ d_{CG} + d_{DE} + d_{DF} + d_{DG} \\ &+ d_{EF} + d_{EG} + d_{FG} \end{aligned}$$

d)

$$
\begin{array}{rrrrrrrr}
 & 3 & + & 3 & + & 2 & + & 1 \\
+ & 2 & + & 1 & + & 2 & + & 2 \\
+ & 3 & + & 1 & + & 2 & + & 3 \\
+ & 1 & + & 2 & + & 1 & & \\
\hline
 & 9 & & 7 & & 7 & & 6
\end{array}
$$

$$= 29$$