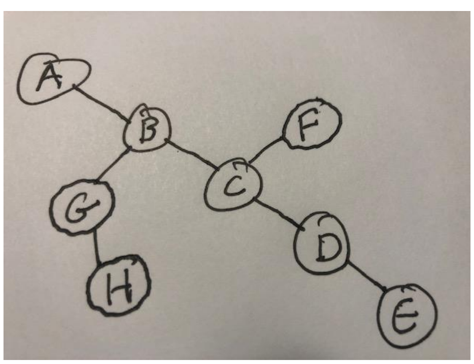
Homework 1
Networked Life (NETS 112)
Fall 2019
Prof Michael Kearns

Marks breakup:
Q1 - 20 points
Q2 - 10 points
Q3 - 15 points
Q4 - 15 points
Q5 - 10 points
Q6 - 10 points
Q7 - 20 points

Problem 1. For each of the following networks, compute the worst-case diameter, the average-case diameter, the edge density (number of edges divided by the maximum possible), and plot the degree distribution. Show your work where appropriate.

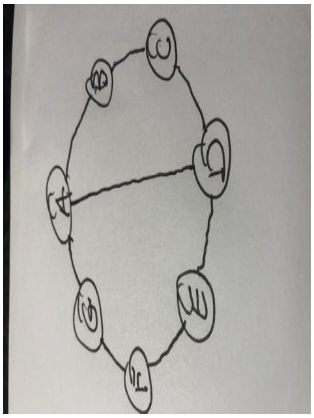
5 points for each example. Full if all correct, 4 if three correct, 3 if two correct, 2 if one correct.

(a)



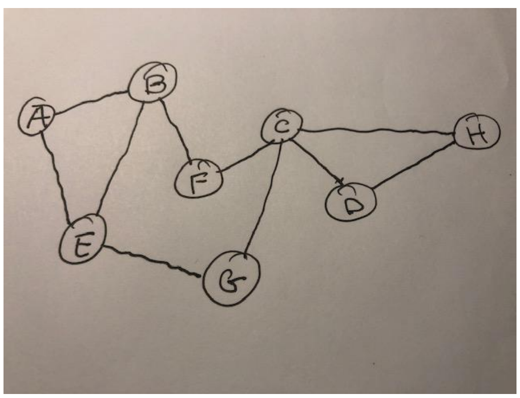
Worst-case - 5 (H → E)
Density - 7 actual edges. 8 choose 2 possible edges = 28 possible edges. Thus, **density = 0.25.**
Average-case - 68/28

(b)



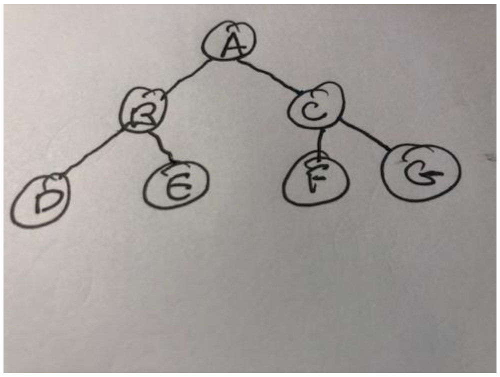
Worst-case - 3 (B -> E and others)
Edge density - 8 edges, 7 choose 2 = 21 possible edges → **density = 8/21.**
Average-case - 38/21

(c)



Worst-case - 4 (A → H and others)
Edge density - 10 edges, 8 choose 2 = 28 possible edges →
density = 5/14.
Average-case - 55/28

(d)



Worst-case - 4 (D → G and others)
Edge density - 6 edges, 7 choose 2 = 21 possible edges →
density = 2/7.
Average-case - 48/21

Problem 2. A *cycle* in a network is a sequence of vertices v1, v2, v3…,vn, v1 such that each vertex is distinct (other than the first and last), and each adjacent pair in the sequence are neighbors (i.e. v1 and v2 are neighbors, v2 and v3 are neighbors, etc. and vn and v1 are neighbors). Since we start at v1 and end at v1, and all vertices in between are distinct, we've taken a tour of the vertices in the sequence. Note that other edges may be present as well (e.g. v1 and v7 might be neighbors). As an example, in the network for problem 1(c), the sequence A, B, F, C, G, E, A is a cycle of length 6 edges.

We can also consider cycles in the directed network of the Web. For instance, the following sequence of URLs:

https://www.cis.upenn.edu/~mkearns/
https://www.cis.upenn.edu/~aaroth/
https://www.cse.huji.ac.il/~katrina/
https://www.cse.huji.ac.il/~katrina/research.html
http://www.cs.cmu.edu/~avrim/
http://hunch.net/~jl/
http://hunch.net/~jl/projects/projects.html
http://hunch.net/~jl/projects/RL/RL.html
https://www.cis.upenn.edu/~mkearns/

is a cycle of length 8: each of the pages above has a link to the next one in the sequence, and we end up back where we started. Note that the hard part of finding a long cycle is not simply finding a long list of links; it's then being able to find your way "home".

Using any methods you like, find the longest cycle you can in the Web, and list the corresponding sequence of URLs. Carefully describe your methodology. We will spot-check for correctness of your cycles (i.e. that each page links to the next). Rewards of some kind will be given for the longest cycle found, and for the most creative/interesting. Try to avoid making the problem easy or trivial by having repeated links from the same top-level domain (e.g. by having all of your links begin with www.cnn.com).

Full points for creative solution with explanation.

-3 for poor explanation
-2 for too short (under 3 jumps)
-5 for no explanation

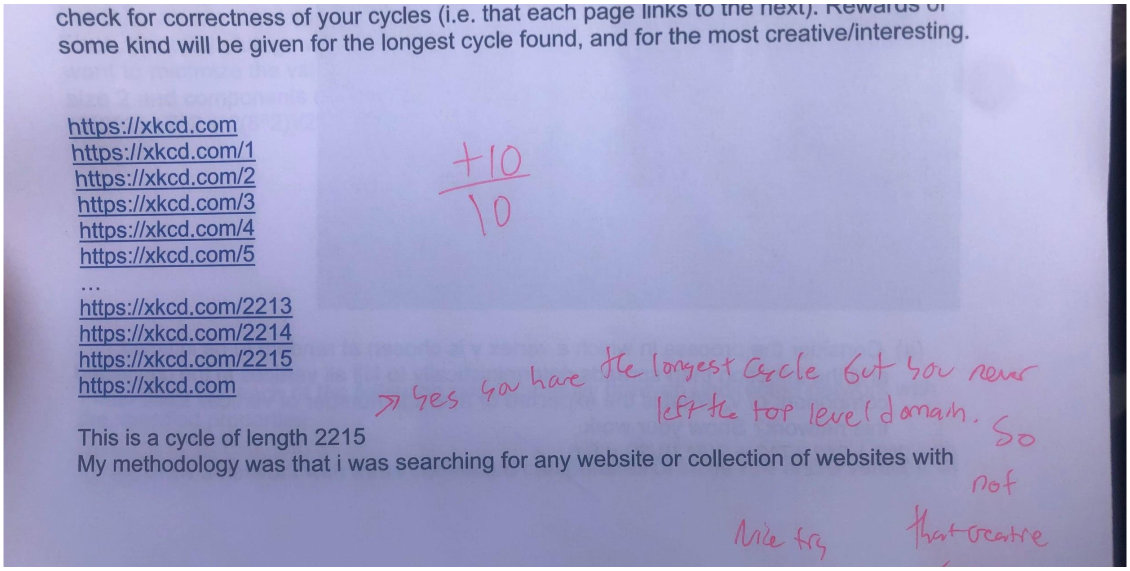
Biggest creative cycle so far, Damian Krupa, 56, bloomberg.com

Cycle of length 56:

https://www.bloomberg.com/ -> https://www.facebook.com/Bloomberglp -> https://www.facebook.com/BloombergCareers/ -> https://www.bloomberg.com/careers -> https://about.bgov.com/ -> https://about.bgov.com/government-affairs/ -> https://about.bgov.com/government-affairs/ -> https://www.linkedin.com/company/bloomberg-government/ -> https://www.linkedin.com/company/bloomberg-bna/ -> https://www.linkedin.com/company/bloomberg/ -> https://www.linkedin.com/company/the-wall-street-journal/ -> https://www.wsj.com/ -> https://classifieds.wsj.com/advertise/careers/ -> https://www.dowjones.com/ -> https://www.dowjones.com/legal-notices/ -> https://www.dowjones.com/contact/ -> https://www.facebook.com/dowjones -> https://www.facebook.com/WallStreetDaily -> https://www.wallstreetdaily.com/ -> https://www.facebook.com/SVBizjournal -> https://siliconvalleybusinessjournal.com/ -> https://www.bizjournals.com/sanjose/news/ -> https://www.bizjournals.com/seattle/ -> https://www.facebook.com/pugetsoundbusinessjournal -> https://www.facebook.com/SeattleBusinessMag -> http://seattlebusinessmag.com -> https://twitter.com/seattlebusiness -> https://twitter.com/PSBJ -> https://www.bizjournals.com/seattle/ -> https://www.bizjournals.com/newyork/ -> https://twitter.com/NYBizJournal -> https://twitter.com/bizjournals -> https://twitter.com/FastCompany -> https://www.fastcompany.com/ -> https://www.fastcompany.com/technology -> https://www.linkedin.com/company/fast-company/ -> https://www.linkedin.com/company/forbes-magazine/ -> https://www.forbes.com/ -> https://www.forbes.com/worlds-billionaires -> https://www.forbes.com/careers-at-forbes -> https://www.forbes.com/careers-at-forbes/software-engineering -> https://twitter.com/forbescareers -> https://twitter.com/metronewyork -> https://twitter.com/nypost -> https://twitter.com/CNN -> https://www.cnn.com/ -> https://www.cnn.com/us -> https://www.cnn.com/specials/us/crime-and-justice -> https://www.instagram.com/cnn/ -> https://www.instagram.com/CNNBusiness/ -> https://www.cnn.com/business -> https://money.cnn.com/data/markets/ -> https://www.cnn.com/business/media -> https://twitter.com/CNNbusiness -> https://twitter.com/business -> https://www.bloomberg.com/

Methodology: I started at the Bloomberg main site, and since almost all major finance-related companies have links to their social media on their websites, and links to their websites on their social media accounts, I was able to smoothly go from companies' websites to their Facebook, Twitter, Instagram or LinkedIn profiles, and vice versa. Thanks to the "Affiliated pages" feature on social media portals, I was able to easily navigate between different finance-related social media profiles, and then from there go to those companies' websites, explore 2-3 pages there, and go back to a different social media profile to find the profile of another company. Since I stack to finance-related companies, I was able to reach the CNN website, from where I could go to the CNN Business page, and then go back to the Bloomberg main page through the "Affiliated pages" feature on the social media site.

And Lucas Weiner for actual longest but not creative.



check for correctness of your cycles (i.e. that each page links to the next). Rewards of some kind will be given for the longest cycle found, and for the most creative/interesting.

https://xkcd.com
https://xkcd.com/1
https://xkcd.com/2
https://xkcd.com/3
https://xkcd.com/4
https://xkcd.com/5
…
https://xkcd.com/2213
https://xkcd.com/2214
https://xkcd.com/2215
https://xkcd.com

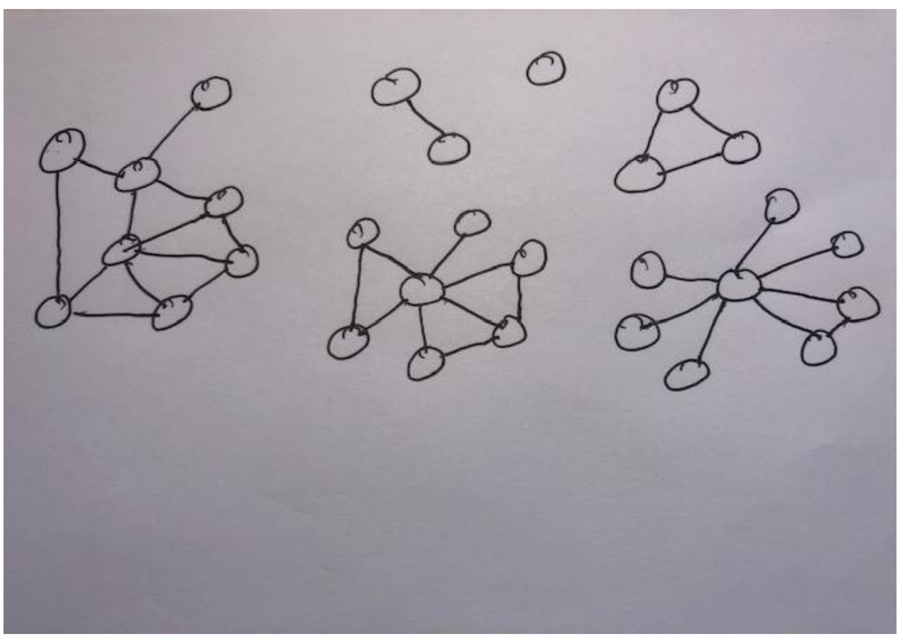
This is a cycle of length 2215
My methodology was that i was searching for any website or collection of websites with

+10 / 10

> yes you have the longest cycle but you never left the top level domain. So not that creative

Nice try

Problem 3. Consider the following network, which consists of multiple connected components:



(a) Consider the process in which a vertex v is chosen at random to be infected, and the infection then spreads deterministically to kill all vertices in the connected component of v. What is the expected or average number of vertices killed for this network? Show your work.

Summing over components, E(killed) = 8*8/29 + 8*8/29 + 7*7/29 + 3*3/29 + 2*2/29 + 1*1/29 = (1/29)*($8^2 + 8^2 + 7^2 + 3^2 + 2^2 + 1^2$) = **191/29.**

(b) Suppose you are allowed to "immunize" exactly one vertex that can no longer be infected, and cannot infect others. Which vertex would you choose to make the average number of vertices killed as small as possible? What would the new value for this average be? Show your work.

The **vertex in the middle of the rightmost connected component (with 8 vertices)** is the correct one to immunize. Then it splits into tive connected components of 1 vertex and one connected component of 2 vertices. Then subtract 8*8/29 from previous E(killed) and add 5*1/29 + 2*2/29. This results in (191-64 + 9)/29 i.e **136/29.**

Alternative approaches that yield 135/29, 137/29, 136/28, etc based on assumptions about the immunized vertex were given full credit.

(c) Suppose you are forced to add an edge between vertices in different connected components. Which edge would you choose to make the average number of vertices killed as small as possible? What would the new value for this average be? Show your work.

An edge connecting the 2-component and the 1-component is the correct edge to add. Then modify your answer from part a) or part b) by subtracting 1*1/29 + 2*2/29 and adding 3*3/29. Answers: **195/29** and others (based on consistency with part b).

Problem 4. For each of the following parts, carefully draw an undirected network with the specified properties.

(a) Draw a network with three connected components of sizes 7, 4 and 3 vertices, in which the total number of edges is 12 and the worst-case diameter (longest shortest-path distance) of the largest component is 5.

(b) Draw a network with three connected components of sizes 5, 4 and 4 vertices, in which the total number of edges is 19 and the worst-case diameter of the largest component is 1.

(c) Draw a network with three connected components of sizes 12, 5 and 3 vertices, in which the total number of edges is 17 and the worst-case diameter of the largest component is 7.

Individual correct solutions, each part worth 5. Wrong diameter size -3. Wrong number edges -2

Problem 5. In this problem you are asked to carefully define, create and visualize a network of your own choosing, ideally reflecting a personal interest. You should begin by carefully specifying what the vertices are in your network, and stating exactly what constitutes an edge between two vertices. As long as you're precise, your network can be about any topic you like; the vertices don't have to correspond to people at all. As one example, the vertices in your network could represent your favorite musical artists (people or bands), and the edges could represent pairs of artists who have collaborated on a song, for some precise definition of collaboration.

Carefully define your vertices and edge definition, making sure your network has at least 15 vertices and 20 edges. Specify any research or data sources you use to identify the edges in your network. Draw a clear visualization of your network, showing all vertices and edges. Compute the worst-case diameter, edge density and degree distribution for your network.

-1 for underspecified edges (e.g. "knowing each other")
-1 for each of diameter/density/distribution not provided

Problem 6. For this problem, you will need a standard 6-sided die in order to make random choices. For each of the following networks, you are to repeat the following experiment **10 times**:

- Use your die to choose randomly choose a "start" vertex in the network.
- From that start vertex, roll your die to pick a random neighbor of your start vertex, and "move" to the chosen vertex.
- Repeat this process for a total of **15 steps**, always choosing a random neighbor of your current vertex and moving to that neighbor.

Note that if your current vertex has, say, 3 neighbors, you can just use the 1, 2 and 3 faces of your die to pick your next move. Also, if your current vertex has only one neighbor, you don't need to roll your die, your next move will simply be to that neighbor.

As an example, a random walk in the network for part (a) below might look like the following sequence of 15 steps:

D, E, D, C, A, B, C, B, C, D, E, A, C, D, E

We will say that this sequence **terminates** at vertex E.

For each of the networks below, perform 10 separate 15-step random walks, and explicitly write out the resulting sequence of vertices visited. Then annotate each vertex in the network diagram with the total number of times that a sequence terminated on that vertex.

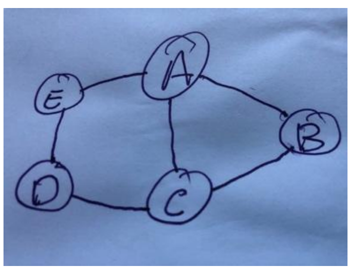
10pts total

5pts for a-c

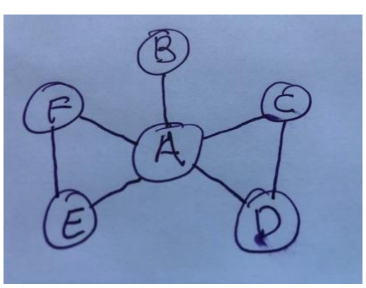
5pts for d

- 2 pts discussing results
- 3 pts discussing limit
- -1 if using “center” w/o clear definition
- -2 if not mentioning degree (number of neighbors, or any similar concept) when discussing infinity

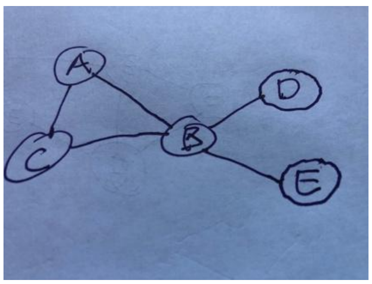
(a)



(b)



(c)



(d) Based on the counts of terminal vertices, do you see any pattern or relationship between the terminal counts and the network structure? Can you guess how the counts would compare to each other across vertices if you were to perform an infinite number of random walks?

The vertices with the highest degree are the most common terminal vertices. If this were extended to infinite random walks, the counts would become proportional to the degree.

Problem 7. Write a brief (say about a page) essay summarizing and synthesizing the readings and lectures on any one of the four course “modules” so far other than the introductory material (i.e. one of “Structural Properties of Networks”, “Contagion in

Networks", "Machine Learning and Social Networks", and "Navigation in Networks" on the course schedule).

-5 if missing 1 important information from the module (readings, topics)

(For example, when discussing the structural properties of networks, one should include discussion of the squash network reading and the Erdos number project.)