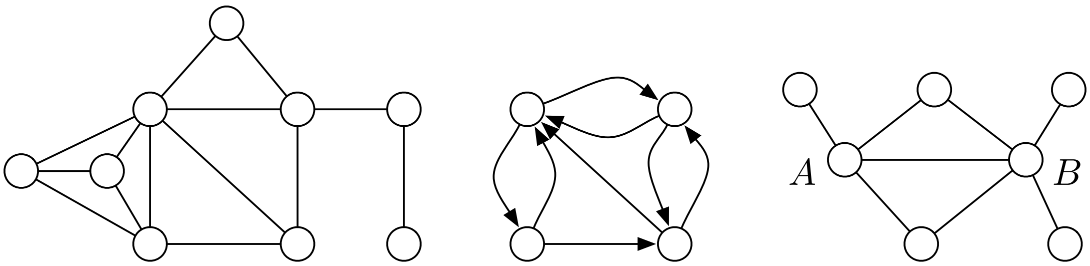
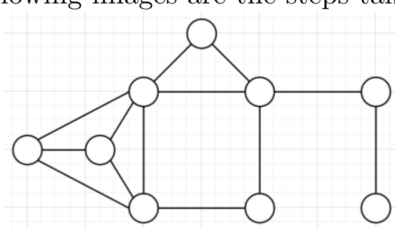
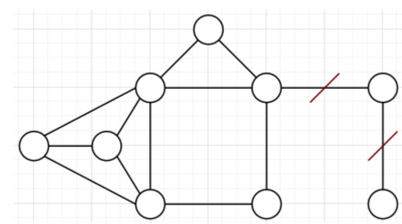
CSCI 5352 Network Analysis and Modeling Fall 2020
Prof. Daniel Larremore
Problem Set 2
Student: Alex Book; Collaborators: None

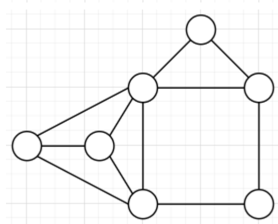
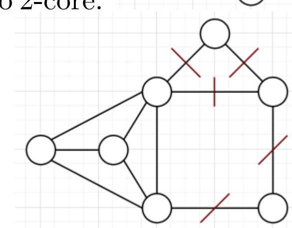
1. (13 pts) Consider the following three networks:

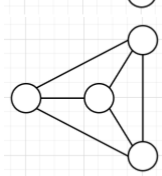


(A) (B) (C)

- (4 pts) Find a 3-core in network (A).
  The following images are the steps taken to find a 3-core:

  1-core:  transition from 1-core to 2-core: 

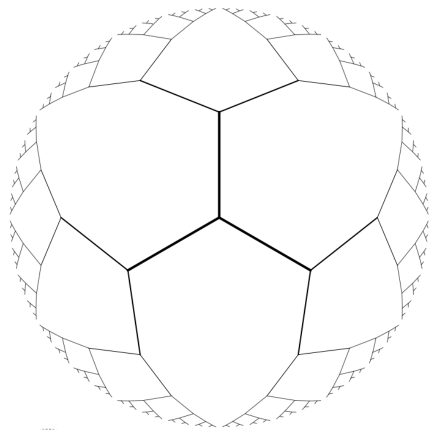
  2-core:  transition from 2-core to 3-core: 

  3-core: 

- (5 pts) What is the reciprocity of network (B)?
  Total edges = 8
  Reciprocated edges = 6
  Reciprocity = Reciprocated edges/Total edges = 6/8 = .75

- (4 pts) What is the cosine similarity of vertices A and B in network (C)?
  $\sigma_{AB} = \frac{n_{ij}}{\sqrt{k_A k_B}} = \frac{2}{\sqrt{4*5}} = \frac{1}{\sqrt{5}}$

2. (15 pts) A Cayley tree is a symmetric regular tree in which each vertex is connected to the same number k of others, until we get out to the leaves, like the figure below, with $k = 3$. Show that the number of vertices reachable in d steps from the central vertex is $k(k-1)^{d-1}$ for $d \geq 1$. Then give an expression for the diameter of the network in terms of k and the number of vertices n. State whether this network displays the "small-world effect," defined as having a diameter that increases as $O(\log n)$ or slower.



Proof by induction:

Base Case:
Let $d = 1$. In 1 step, there would be k nodes reachable from the central vertex (the k nodes it is connected to).

$$k(k-1)^{d-1} = k(k-1)^{1-1} = k \tag{1}$$

Let $d = 2$. In 2 steps, there would be $k(k-1)$ nodes reachable from the central vertex (the k nodes it is directly connected to, plus the $k-1$ nodes that each of *those* are connected to).

$$k(k-1)^{d-1} = k(k-1)^{2-1} = k(k-1) \tag{2}$$

Inductive Hypothesis:
Assume that for $d = i$, where $i \geq 1$, the number of nodes reachable from the central vertex is represented by $s_i = k(k-1)^{d-1}$.

Inductive Step:
Let $d = i + 1$. We can say that the number of nodes reachable s from the central vertex in any given step is equal to the number reachable from the last step multiplied by the quantity $(k-1)$ (as $k-1$ new nodes are reached from each existing 'leaf' node).

$$\begin{aligned}
s_{i+1} &= (k-1) * s_i \\
s_{i+1} &= (k-1) * k(k-1)^{i-1} && \text{by inductive hypothesis} \\
s_{i+1} &= k(k-1)^i && \text{by multiplication}
\end{aligned}$$

The diameter of a Cayley tree can be expressed as twice the number of steps taken d, as one edge is added on either 'side' of the tree with each step.
The total number of nodes n can be expressed as:

$$n = 1 + \sum_{i=1}^{d} k(k-1)^{i-1} \tag{3}$$
$$n = 1 + k\frac{1-(k-1)^d}{1-(k-1)} \tag{4}$$
$$n = 1 - (1-(k-1)^d) \tag{5}$$
$$n = (k-1)^d \tag{6}$$
$$\log(n) = d\log(k-1) \tag{7}$$
$$d = \frac{\log(n)}{\log(k-1)} \tag{8}$$
$$2d = \frac{2\log(n)}{\log(k-1)} \tag{9}$$
$$\text{diameter} = \frac{2\log(n)}{\log(k-1)} \sim O(\log(n)) \tag{10}$$

Therefore, this network does indeed display the "small-world effect." Note that the above holds true for $k \geq 3$. $k = 1$ allows for only two total nodes and thus isn't considered. $k = 2$ yields the following:

$$n = 1 + \sum_{i=1}^{d} 2(2-1)^{i-1} \tag{11}$$
$$n = 1 + \sum_{i=1}^{d} 2 \tag{12}$$
$$n = 1 + 2d \tag{13}$$
$$2d = n - 1 \tag{14}$$
$$\text{diameter} = n - 1 \tag{15}$$
$$\text{diameter} \sim O(n) \tag{16}$$

Thus we can say that for $k \geq 3$, Cayley Trees display the "small-world effect."

3. (35 pts total) In this question, we will investigate several properties of online social networks by analyzing the Facebook100 ("FB100") data set, which you may download via the link posted on Piazza. Each of the 100 plaintext ASCII files in the FB100 folder contains an edge list for a 2005 snapshot of a Facebook social network among university students and faculty within some university. Interpret this edge list as a simple graph.[1]

   (a) (5 pts) In most social networks, we observe a surprising phenomenon called the *friendship paradox*. Let k_u denote the degree of some individual u, and let some edge $(u, v) \in E$. The paradox is that the average degree of the neighbor $\langle k_v \rangle$ is *greater* than the average degree $\langle k_u \rangle$ of the vertex. That is, on average, each friend of yours has more friends than you.

   The mean neighbor degree (MND) of a network is defined as

$$\langle k_v \rangle = \frac{1}{2m} \sum_{u=1}^{n} \sum_{v=1}^{n} k_v A_{uv} \quad . \tag{17}$$

   Derive an expression for $\langle k_v \rangle$ in terms of the average squared-degree $\langle k^2 \rangle$ and the average degree $\langle k \rangle$. Show your work.

   Note: In an undirected graph, the adjacency matrix is symmetrical across the diagonal, so $A_{ij} = A_{ji}$.

$$\langle k \rangle = \frac{1}{n} \sum_{i=1}^{n} k_i = \frac{2m}{n} \tag{18}$$

$$\langle k^2 \rangle = \frac{1}{n} \sum_{i=1}^{n} k_i^2 \tag{19}$$

$$k_i = \sum_{j=1}^{n} A_{ij} \tag{20}$$

$$\langle k^2 \rangle = \frac{1}{n} \sum_{i=1}^{n} k_i * k_i = \frac{1}{n} \sum_{i=1}^{n} k_i \sum_{j=1}^{n} A_{ij} \tag{21}$$

$$\langle k^2 \rangle = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} k_i A_{ij} \tag{22}$$

$$\text{substituting u,v in for i,j we get: } \langle k^2 \rangle = \frac{1}{n} \sum_{u=1}^{n} \sum_{v=1}^{n} k_u A_{uv} \tag{23}$$

$$\frac{\langle k^2 \rangle}{\langle k \rangle} = \langle k_v \rangle = \frac{1}{2m} \sum_{u=1}^{n} \sum_{v=1}^{n} k_u A_{uv} \tag{24}$$

[1]The data were kindly provided by A.L. Traud, P.J. Mucha and M.A. Porter, as part of their paper "Social Structure of Facebook Networks," *Physica A* **391**, 4165–4180 (2012), which is freely available at `http://arxiv.org/abs/1102.2166` or `http://bit.ly/1ztbVoS`.

(b) (15 pts) Now, using all 100 of the FB100 networks, make a figure showing a scatterplot of the ratio $\langle k_v \rangle / \langle k_u \rangle$ as a function of the mean degree $\langle k_u \rangle$. Include a horizontal line representing the line of "no paradox," and label the nodes corresponding to Reed, Bucknell, Mississippi, Virginia, and UC Berkeley. (Remember: figures without axes labels will receive no credit.)

Comment on the degree to which we do or do not observe a friendship paradox across these networks as a group. Comment on whether there is any dependency between the magnitude of the paradox (the size of the MND, relative to the size of the mean degree) and the network's mean degree. A few points of extra credit will be awarded to an explanation of why we should, in fact, expect to see a friendship paradox in these networks, and that identifies the conditions under which we should expect to see *no* paradox.

There does indeed seem to be a friendship paradox across all schools, as can be seen in the figure generated by my code. There doesn't seem to be any direct relationship between the magnitude of the paradox and the network's mean degree.

(c) (15 pts) A related phenomenon in social networks is the *majority illusion*. Let $x \in \{0, 1\}$ be a binary-valued vertex-level property, and let $q = \frac{1}{n}\sum_u x_u$ be the fraction of vertices that exhibit this property. If we set $q < 0.5$, then this property appears only in a minority of nodes. The majority illusion occurs when $q < 0.5$, but the majority of a node's neighbors, on average, exhibit that property, that is, $\langle x_v \rangle > 0.5$.

Explain in words and mathematics how this can be possible.

This is likely due to many of the nodes that exhibit the property being well-connected (having a high degree). Say, if a few nodes (with the property) have 100 or more connections each, while many times more nodes (without the property) only have a few connections each (with many of those connections being with the nodes that have the property). On average, any given node wouldn't have the property, but would likely be connected to nodes with the property. This is exemplified in everyday life by the social media influencers that have millions of followers, but follow far fewer people. When influencers advertise a product, their followers (who likely follow numerous similar influencers) are likely to feel that it is very popular, despite few people actually advertising it.

(d) (20 pts *extra credit*) Another common property of social networks is that they have very small diameters relative to their total size. This property is sometimes called the "small-world phenomenon" and is the origin of the popular phrase "six degrees of separation".[2]

- For each FB100 network, compute (i) the diameter $\ell_{\max}$ of the largest component of the network and (ii) the mean geodesic distance $\langle \ell \rangle$ between pairs of vertices in the largest component of the network. Make two figures, one showing $\ell_{\max}$ versus network size n and one showing $\langle \ell \rangle$ versus the size of the largest component n.

[2]This term originated in a play written by John Guare in 1990, which was turned into a 1993 movie starring Will Smith. The concept, however, was originated by Stanley Milgram, working in 1967, who was the first to measure the lengths of paths in large social networks. This is the same Stanley Milgram who did important but problematic work on obedience. Dive into this story at the Wikipedia page for "Milgram experiment" when you have time...

Comment on the degree to which these figures support the six-degrees of separation idea.

- Briefly discuss whether and why you think the diameter of Facebook has increased, stayed the same, or decreased relative to these values, since 2005. (Recall that Facebook now claims to have roughly 10^9 accounts.)