Aaron Orenstein

1) By definition of odd, $n = 2k + 1, k \in Z$.

$$C_i = \frac{n}{\sum_j d_{ij}} = \frac{2k+1}{(1+2+\cdots+k)+(1+2+\cdots+k)} = \frac{2k+1}{k(k+1)} = \frac{n}{\left(\frac{n-1}{2}\right)\left(\frac{n+1}{2}\right)} = \frac{4n}{(n-1)(n+1)}$$

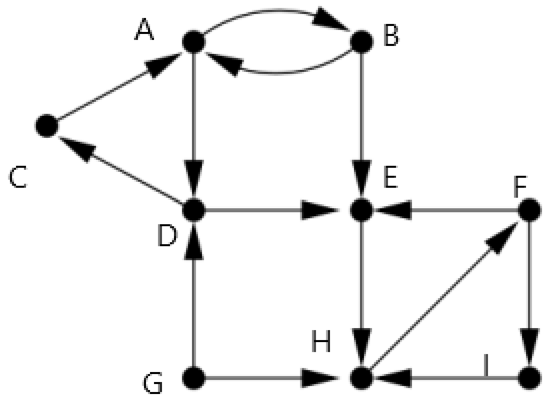
2)

a.

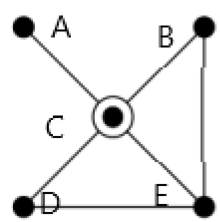
i. The whole graph is a 3-core as every node has edges to 3 nodes

ii. There is no 3-core. Starting with the whole graph, we must remove the bottom node as it connects to only 2 nodes. Then the bottom corners connect to only 2 nodes and must be removed. This repeats until there are no nodes left. Thus there is no set of nodes that is a 3-core.

b. SCCs:
Reflexive Pairs: {AB}
Triangles: {ACD}, {EFH}, {FHI}
We can combine SCCs if there are edges to and from nodes in each SCC: {EFHI}, {ABCD}



c. $C_A = \frac{0}{0} = 0, C_B = \frac{1}{1} = 1, C_C = \frac{2}{6} = \frac{1}{3}, C_D = \frac{1}{1} = 1, C_E = \frac{2}{3} = \frac{2}{3}$



d. 1 = left, 2 = right

$e_1 = \frac{6}{20} = 0.3, r_1 = \frac{8}{20} = 0.4$

$e_2 = \frac{10}{20} = 0.5, r_2 = \frac{12}{20} = 0.6$

$Q = (0.3 - 0.4^2) + (0.5 - 0.6^2) = 0.28$

e. Every pair of nodes has one shortest path that either start/ends in the center or goes through the center. Thus $x_i = n(n-1)$

3)

a. The diameter is 1 because any pair of nodes must have an edge between them (by def. of clique) which is also the shortest path.

b. The diameter is $dL = d(n-1)$ (so $2L$ for the square case). The length of the geodesic path between any two nodes is the Manhattan distance between them. The furthest nodes are the opposite corners which must travel L edges in each dimension.

c. In the first step, there are k reachable nodes. For each subsequent step, every reachable node can move to $k-1$ new nodes. Then at the end, there are $k(k-1)^{d-1}$ nodes. To travel between two nodes, we must move down to the common ancestor and then back up to the second node. At the worst case, the common ancestor Is the center node. Then the diameter is $2d$, where d is the number of branches outwards.

$n = 1 + \sum_{i=1}^{d} k(k-1)^{i-1} = 1 + k\sum_{i=0}^{d-1}(k-1)^i = 1 + \frac{k}{2-k}\left(1-(k-1)^d\right)$

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

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

Diameter = $2d = 2\log\left(1 - \frac{(n-1)(2-k)}{k}\right) / \log(k-1)$

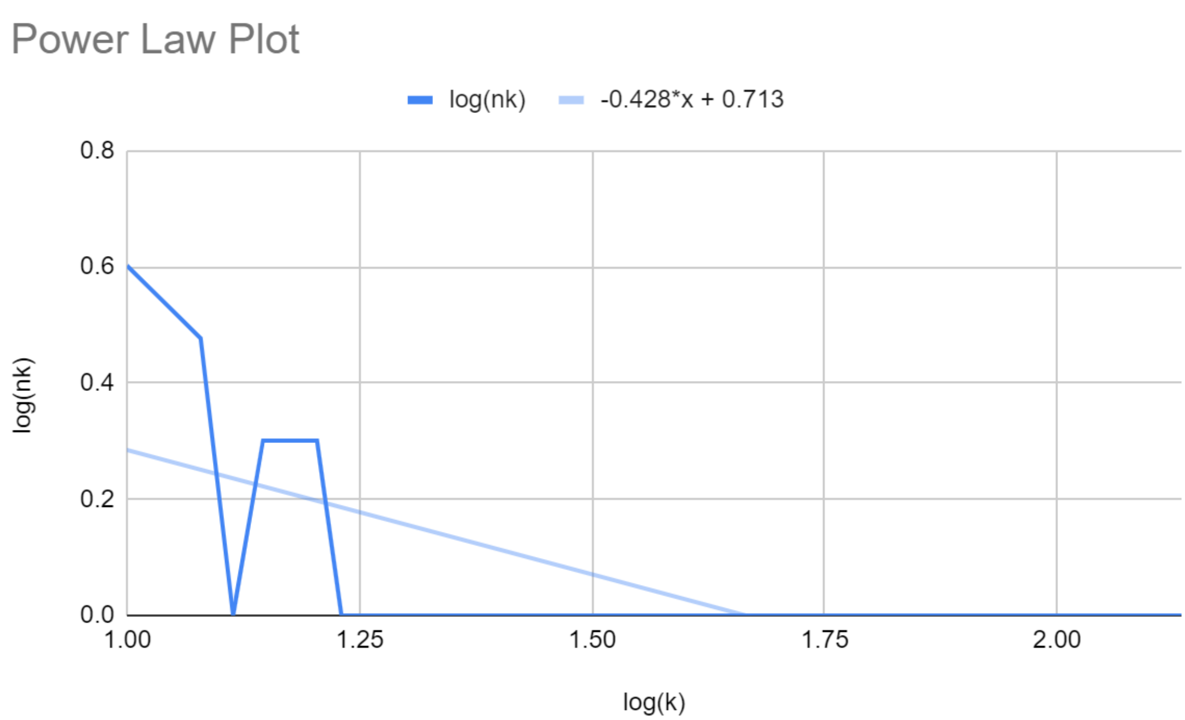
d.

i. CWE. A clique has a constant diameter.

ii. Not CWE. The diameter of the lattice increases linearly with n.

iii. CWE. The math in (c) reduces to $\log(n)$ scaling since the k terms are constant. The number of nodes increases exponentially as the diameter increases linearly. Thus if the nodes increases linearly, the diameter is increasing logarithmically.

4) Here's a log-log plot of the degree counts vs degree with a linear regression (calculated by google sheets). The slope of the best-fit line is $-\alpha$. Then $\alpha = 0.428$.



5) Let group 1 have r nodes and group 2 have $n-r$ nodes.

a. $e_1 = \frac{2r-2}{2n-2} = \frac{r-1}{n-1}, r_1 = \frac{1+2r-2}{2n-2} = \frac{2r-1}{2n-2}$

$$e_2 = \frac{2n-2r-2}{2n-2} = \frac{n-r-1}{n-1}, r_2 = \frac{1+2n-2r-2}{2n-2}$$

$$Q = \frac{r-1}{n-1} + \frac{n-r-1}{n-1} - \frac{4r^2-4r+1}{4(n-1)^2} - \frac{(2n-2r-1)^2}{4(n-1)^2}$$

$$= \frac{n-2}{n-1} - \frac{4n^2+8r^2+2-8nr-4n}{4(n-1)^2}$$

$$= \frac{4n^2-12n+8-4n^2-8r^2-2+8nr+4n}{4(n-1)^2}$$

$$= \frac{-8r^2+6+8nr-8n}{4(n-1)^2} = \frac{3-4n+4nr-4r^2}{2(n-1)^2}$$

b. By definition of even, $n = 2k, k \in Z$. So $Q = \frac{3-8k+8kr-4r^2}{2(2k-1)^2}$

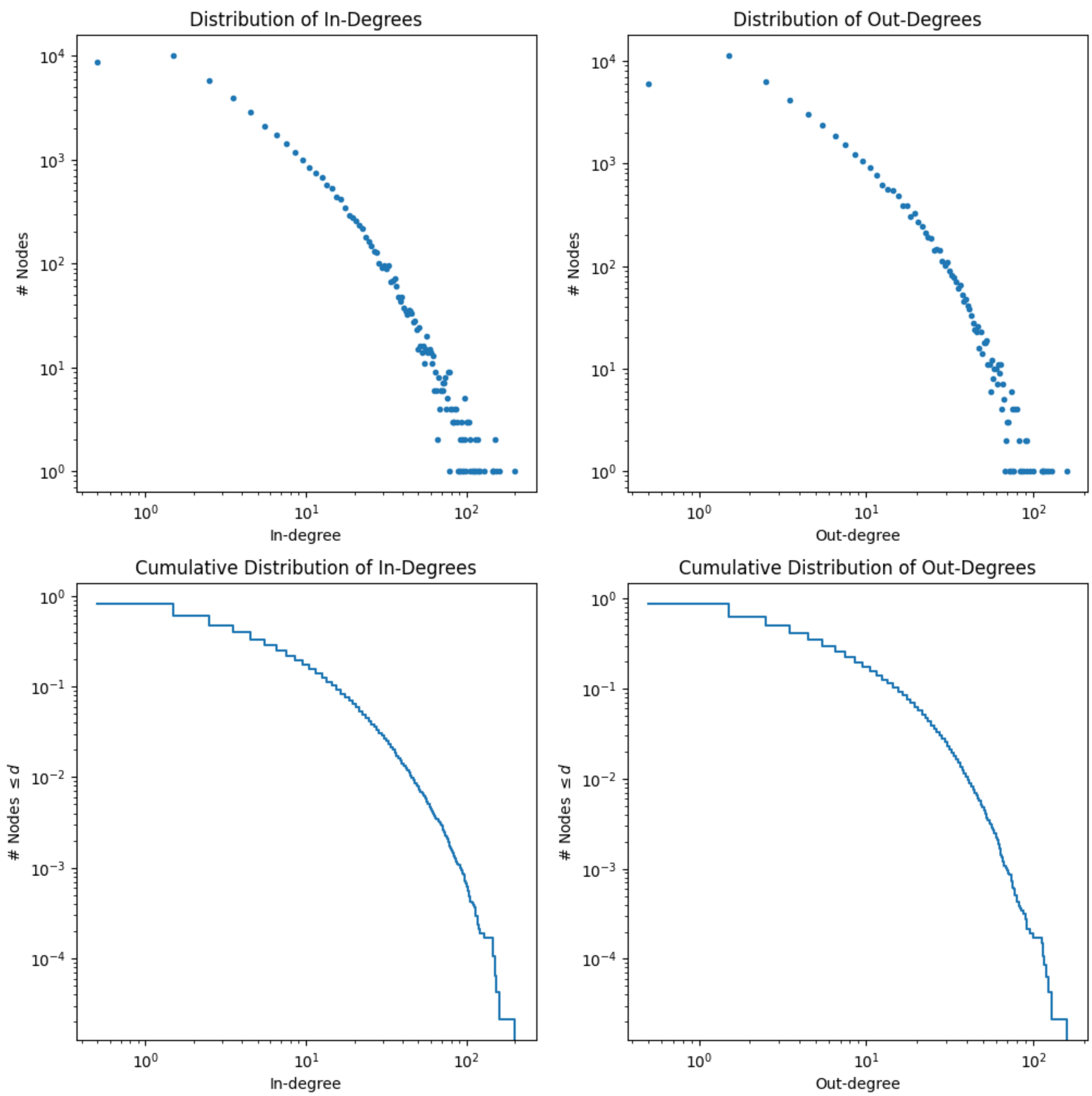
$$\frac{dQ}{dr} = \frac{8k-8r}{2(2k-1)^2} = 0 \Rightarrow k - r = 0 \Rightarrow r = k$$

$\frac{d^2Q}{dr^2} = \frac{-8}{2(2k-1)^2} < 0$ so this point is a local (and global) maximum for the modularity

6)

a.

b.



c. The distributions and CCDFs are nearly identical for wall and friend edges. This indicates that users tend to post on their friends' walls. For every friend edge, there is a corresponding edge where each person posted on the other's wall. However, we ignored the time component of each post. If we restricted the recency of posts, the distributions would likely look different as there would be friends who had not posted on each other's walls recently. The wall distributions would have smaller degree than the friend edges.

7)

Random (each edge gets a random score between 0-1)
Averaged over 100 trials
Top 8 Score: 287.29951185708813
Accuracy: 0.11 / 8
AUROC: 0.503936335403727

a. preferential_attachment
   Top 8 Score: 1273
   Accuracy: 1 / 8
   AUROC: 0.6956521739130435

b. common_neighbors
   Top 8 Score: 275
   Accuracy: 1 / 8
   AUROC: 0.6696428571428571

c. jacquard
   Top 8 Score: 303.8
   Accuracy: 0 / 8
   AUROC: 0.5786749482401656

d. adamic_adar
   Top 8 Score: 216.85352682995904
   Accuracy: 1 / 8
   AUROC: 0.7304606625258798

The random links method achieves an average auroc value of 0.50 as expected. All the other methods achieve and auroc value above 0.5 (worst is jacquard, best is adamic_adar). This shows that they outperform random link creation. All but jacquard are able to predict one of the removed links. Jacquard still has a better-than-random auroc because it assigned higher scores to the removed links (relative to other missing edges).