Test 1 Solutions
Brendan Meeder

---

1 Signed Cycle Graphs

There are at least three different ways to approach this problem, and I'll go over two of the simplest arguments.

Proof 1: For a given n, enumerate the edges of the cycle $e_1, \ldots, e_n$ and we let $s_1, \ldots, s_n$ be the sign assigned to $e_1, \ldots, e_n$. Let f be the function $f(s_1, \ldots, s_n) = (-s_1, s_2, \ldots, s_n)$ (the function which takes an assignment of signs to the edges and flips the sign of e_1). Notice that this function is its own inverse and takes sign assignments of balanced cycles to sign assignments of unbalanced cycles (and visa versa). We should think of f as a mapping between the sets B and U of assignments of signs that result in balanced and unbalanced cycles, respectively. Every assignment $(s_1, \ldots, s_n)$ in U gets mapped to by f from the assignment $(-s_1, \ldots, s_n)$ in B, so it is surjective. Furthermore, the function f is clearly injective, so it's a bijection between sign assignments of balanced and unbalanced cycles. Therefore, exactly half of the 2^n signed cycles are balanced.

Proof 2: Again, enumerate the edges and arbitrarily pick an assignment of signs $s_1, \ldots, s_{n-1}$ for $e_1, \ldots, e_{n-1}$, respectively. There is only one choice of s_n which makes the cycle balanced: $s_n = \prod_{i=1}^{n-1} s_i$. Each balanced cycle can be uniquely associated with (put into a bijection with) exactly one of the 2^{n-1} choices of signs for $e_1, \ldots, e_{n-1}$, as once those signs are fixed, the choice for s_n is forced. Thus, there are 2^{n-1} many balanced cycles.

Leaving an answer in terms of a sum of binomial coefficients, without simplifying to the answer of 2^{n-1}, demonstrates that you remembered the definition of a balanced cycle. However, that proof doesn't provide the insight that half of all signed cycle graphs are balanced.

2 Status Prestige

In status prestige the prestige of a node i is the sum of the prestige of all of the nodes that 'think highly' of i. Usually you'd have to solve some linear system, but in this case the graph is acyclic so we simply write out the equations and forward solve the system:

$$\begin{aligned}
P_1 &= 1 \\
P_2 &= P_1 + P_5 \\
P_3 &= P_2 + P_4 + P_5 \\
P_4 &= P_2 + P_5 \\
P_5 &= P_1
\end{aligned}$$

Solving the system, we get $P_1 = 1, P_5 = 1, P_2 = 2, P_4 = 3, P_3 = 6$.

3 Designing a Viral Marketing Strategy

Some key points that we were looking for here:

- Measuring centrality and influence: We can use existing examples of one user exerting influence over another (maybe their posts are retweeted or liked, links that they share are frequently clicked on, etc.). We can also use the social graph structure to figure out traditional network positions such as bridges or structural holes. Since an actual structural hole is unlikely to exist in the network, you could come up with a new measure that is similar to what a structural hole captures. For example, rather than requiring it to be a cut vertex, you might say it's a node such that its neighborhood can be separated into chunks which are relatively large (say, at least 5% of its neighborhood), and which are not highly connected to each other.

- Running experiments. You can test how accurate your model is by getting some user to perform a particular behavior for you. Once you understand the way that a celebrity or some other user influences their peers, you can then better talk to someone who wants to publicize something on the network.

- Running simulations: If we build a model, such as an independent cascade model, we can perform simulations to estimate how messages spread in the network. If we had a threshold model, we could randomize over some range of thresholds we think users might have and run many simulations to see which groups of users are blocking clusters, which groups are extremely likely to adopt the product, etc.

- Algorithms: You will probably need to run your algorithms in some distributed manner as the social networks at hand are extremely large.

4 Centrality Measures

For the degree centrality, we see that the black node is connected to all other nodes, thus $deg(B)/(n-1) = 6/6 = 1$.

For closeness centrality, we notice that $d(B, v) = 1$ for all vertices. Thus,

$$C_C(B) = (n-1)\left(\sum_v d(B,v)\right)^{-1} = (n-1)/(n-1) = 1.$$

For betweenness centrality, label the non-black vertices 1-6, labeling the upper left vertex 1 and continue in the clockwise direction. If we consider the definition for betweenness centrality

$$C_B(black) = \sum_{i<j} \frac{short_{i,j}(black)}{short_{i,j}}$$

we see that we need to look at 15 pairs of vertices. For seven of these pairs,

$$\{(1,2),(2,3),(3,4),(4,5),(5,6),(1,6),(2,6)\},$$

we see that the corresponding term in the betweenness centrality calculation is zero as they are vertices adjacent to each other in G. For seven additional pairs,

$$\{(1,3),(1,5),(2,4),(2,5),(3,5),(3,6),(4,6)\},$$

there are two length two paths between i and j, and one of them goes through the black vertex; each of these pairs will contribute $1/2$ to the betweenness centrality calculation. Finally, the remaining $(1,4)$ pair contributes 1 to the sum as the only shortest path from 1 to 4 goes through the black node. Thus, the betweenness centrality of the black node is 4.5.

5 Transitive Triangles

For part a, we let $X_{(a,b,c)}$ be in indicator random variable that the *ordered* triple (a, b, c) forms a transitive triangle. This happens if and only if the directed edges $(a, b), (b, c)$, and (a, c) are present. Each edge is present, independently, with probability p. Thus, $E[X_{(a,b,c)}] = p^3$. The random variable

$$X = \sum_{(a,b,c)} X_{(a,b,c)}$$

counts the number of transitive triangles in the graph. The summation is taken over the $n(n-1)(n-2) = 3!\binom{n}{3}$ many ordered triples of vertices. By linearity of expectation, we expect there to be $p^3 n(n-1)(n-2)$ transitive triples in the graph.

For part b, we didn't have you give a solution that works for arbitrary n, so most people gave a directed 4-cycle as an example. For even n, and easy general solution is as follows: let $n = 2k$ split the vertices into two sets S_1 and S_2 of k vertices each. For every $a \in S_1$ and $b \in S_2$, make the directed edge (a, b). There are $k^2 = n^2/4$ many edges in this graph, and there are clearly no transitive triples as there aren't any undirected 3-cycles. In fact, the construction generalizes: pick any direction for each of the k^2 many edges in the complete k, k bipartite graph.

6 Diameter 2 G(n,p) Graphs

As per the hint, we're going to look at BAD event 'there exists some pair of vertices (i, j) such that $d(i, j) > 2$ (there is no length 1 or length 2 path between i and j).' The simple bad events here are $\{B_{i,j}\}_{i<j}$ which occur when, for the pair of vertices i and j, there doesn't exist a length 1 or length 2 path between i and j. We see that BAD happens if and only if at least one of the $B_{i,j}$ occur, so $Pr[BAD] = Pr[\cup_{i<j} B_{i,j}]$.

Let's examine the probability of $B_{i,j}$ occurring for an arbitrary (i, j) pair. There isn't an edge between i and j with probability 0.99. For the length two paths, we need to consider the $n-2$ intermediate nodes $k \neq i, j$ through which a length 2 path could go. The probability that there isn't a length two path through a particular k is $(1 - 0.01^2)$. In order for $B_{i,j}$ to happen, we need for there to be no (i, j) edge and no length 2 path through *each one of the* $n-2$ intermediate nodes k. By the rule of product for probabilities,

$$Pr[B_{i,j}] = 0.99(1 - 0.01^2)^{n-2} \leq 0.9999^{n-2}.$$

A quick aside: Some people made the mistake of saying that the probability of there being no i, j path through k is 0.99^2, but that's actually too strong since that's the probability that *neither* of the (i, k) and (k, j) edges are present. It might seem like this is inconsequential, but for bounding arguments, you need to make sure that you're estimate is loose on the 'correct' side. What I mean by this is that if you're examining upper bounds for A of the form $A \leq B$, you can make a sloppy estimate C of B, *as long as you know that* $C \geq B$. In this case, the estimate 0.99^2 is smaller than the actual probability, so it's not ok to use this probability in your bounding argument.

Back to the proof: At this point we use the union bound,

$$Pr[BAD] = Pr\left[\cup_{i<j} B_{i,j}\right] \leq \sum_{i<j} Pr[B_{i,j}] \leq \binom{n}{2} 0.9999^{n-2} \leq n^2 0.9999^{n-2}$$

and we see that $0 \leq Pr[BAD] \leq n^2 0.9999^{n-2} \Rightarrow Pr[BAD] \to 0$ as $n \to \infty$. Thus, $G(n, 0.01)$ graphs asymptotically almost surely have diameter no more than 2.

A common mistake was to use the 'rule of sum' instead of 'rule of product' when calculating $Pr[B_{i,j}]$ (getting something which is $.99(1 - 0.01^2(n-2))$)). It's always good to sanity check your assertions as you go along- in this case you could identify that something is amiss since for $n > 10000$ you get that $B_{i,j}$ occurs with a negative probability. Some students argued that the *expected* number of length 2 paths is $\Theta(n)$, and that this is good enough to prove $G(n, 0.01)$ graphs have diameter at most 2 a.a.s. The first part of this claim is in fact correct, but arriving at the conclusion via this argument actually has a subtle error in it. It might intuitively seem like if a sequence of random variables $X_1, X_2, \ldots$ has unbounded expected value ($E[X_n] \to \infty$ as $n \to \infty$) that would be good enough to argue that $Pr[X_n = 0] \to 0$ as $n \to \infty$. However, *in general this is not true*: for example, let X_n be a random variable such that $X_n = 0$ with probability $1 - 1/n$ and $X_n = n^2$ with probability $1/n$. $E[X_n] = n$, which becomes unbounded as $n \to \infty$, but X_n actually takes on the value 0 almost surely! The problem is that you need to show the variance of the X_n isn't 'too big' relative to the expected value of X_n to avoid examples like the one I just gave. In the problem you were given it is the case that the distribution of the number of length 2 paths from i to j in $G(n,p)$ is sufficiently concentrated about its mean to make the argument go through, but you need to do a bit more work than simply showing the expected number of paths grows unbounded.

7 Independent Cascades

Label the vertices in the top row (left to right) as 1, 2, and 3, and the vertices in the bottom row 4, 5, and 6. We let X_i be the indicator random variable that node i gets infected in the cascade and $X = \sum_{1 \le i \le 6} X_i$.

As vertices 1,2, and 3 are infected with probability 1, we have $E[X_1] = E[X_2] = E[X_3] = 1$. $X_4 = 1$ with probability 0.5, and $X_5 = 1$ only if both edges in the path from the black node to node 5 are present; this happens with probability 0.25. This gives us that $E[X_4] = 0.5, E[X_5] = 0.25$. $X_6 = 0$ only if both attempts to infect v_6 fail, which happens with probability 0.25; therefore $E[X_6] = 0.75$. Adding everything together, we get $E[X] = 4.5$. A common mistake was to forget the black node, or to say that v_6 gets infected with probability 1.