Mini-projects

Master ACN: Random graphs and epidemics

January 21, 2016

These exercises will be graded, towards determining your final grade for the class. Only two out of three will be graded (you can choose which three). The third problem is related to the Jan 29th class lecture, and will be given then.

1. Who is influential?
   Let's consider a graph of individuals. We would like to find the set of most influential individuals. The size of a cascade is defined as the the total number of individuals in the graph that are infected when the spreading stops. For the Independent Cascade Model starting with an initial set of size k, simulate the process, to calculate the size of the cascade. Plot the cascade size against k choosing the set k initially infected individuals according to: the degree centrality (those with highest degree), distance centrality (those with lowest average distance to other nodes), the greedy method we discussed in class, and a random set of k nodes. Note that for greedy method, you will need to estimate the influence function $f(S)$. This can be done by sampling M infections over the graph. Try this sampling for M large enough that the error is small enough.

   Use the SNAP dataset (`https://snap.stanford.edu/data/index.html`), in particular the data from collaboration networks. Use a smaller dataset to start and try again with a larger dataset to see how your results differ.

   (a) What value of M did you use (for your chosen dataset)?
   (b) Choose a dataset from those on the cited website. The dataset contains only undirected weightless edges. Create two graphs - one where all edges have the same weight $p, 0 < p < 1$, and one where each edge (u, v) has a weight p_{uv} chosen uniformly between 0 and 1. Fix the graphs before the next step (i.e. do not draw p and p_{uv} for each run for the next plots). Plot cascade size against k, for k from 1 to 30 for the four methods mentioned above, for each graph.
   (c) Bonus question: what is the complexity of the greedy algorithm?

2. A coordination game
   First, consider the following two player coordination game. The two players, or two nodes are connected to each other and can adopt one of two behaviours (or opinions), A, or B. A node's payoff is higher if he adopts the same behaviour as his neighbour's. If they both adopt A, they both get a payoff of $a > 0$. If they both adopt B, they both get a payoff of $b > 0$. If they have differing behaviours, they get 0.

   Now consider a graph of many nodes. Each node now plays this coordination game with each of his neighbours. Let p represent the fraction of a node's neighbours that adopt A. Then, a node with degree d will choose A when $apd > b(1-p)d$, so when $p > q$, where $q = b/(a+b)$.

(a) Now consider some general graph, and a set S or early, hard-wired adopters of A. Does A spread monotonically throughout the graph? Prove your answer.

(b) Consider an infinite line graph where everyone starts with B. Starting with one node that adopts A, for what values of q will the cascade end with every node adopting A? For what values of q will everyone adopt A in an infinite grid graph?

(c) Now consider the possibility of maintaining both behaviours, AB, at a cost c. Let's say we have an infinite line graph, with all nodes initially in B. The payoff for a node is the sum of payoffs over the two edges minus any cost. The payoff matrix for a given edge (u, v) is given below (in addition to the below, a node pays a cost of c for maintaining AB - once, not for each edge):

$u \setminus v$	A	B	AB $(-c)$
A	(a, a)	(0,0)	(a, a)
B	$(0, 0)$	(b,b)	(b, b)
AB $(-c)$	(a, a)	(b, b)	$(\max(a, b), \max(a, b))$

Now let's say the first node in the line changes to behaviour A. Now with $b = 1$, $a > b$, for what values of (a, c) does A spread?

(d) Can you derive more general conditions (i.e. do not set b=1)?