Exercises: Influence and opinion dynamics

Master ACN: Random graphs and epidemics

January 15, 2016

1. Granovetter'78 [M. Granovetter. Thresholds Models of Collective Behavior. The American Journal of Sociology, 83(6), May 1978, 1420-1443]

   Assume a complete graph; we are considering collective action here. There are N individuals and everyone observes everyone else's action. Each node i has a threshold θ_i, $0 \leq \theta_i \leq 1$. Node i will adopt the new behaviour if at least a fraction θ_i of people have already adopted it. We consider discrete time, and we observe the population of adopters. Let $f(x)$ be the fraction of people with threshold $\theta_i \leq x/N$. $f(x)$ is non-decreasing. Let $n(t)$ be the number of adopters at step t.

   (a) Start with $n(0) = 0$. Then $n(1) = f(n(0)); n(2) = f(n(1)) \ldots$. Simulate this collective behaviour, for θ_i taken from a uniform distribution. Plot the population of adopters vs the threshold values $\tilde{\theta}_i = \theta_i * N$. What size is the resulting cascade?

   (b) Do the same for $\tilde{\theta}_i$ taken from a Normal distribution, with the same mean $(N/2)$ and for many values of variance (start with a small variance, like $< N/10$ and go as large as you can). How does the size of the cascade change as the variance increases?

   (c) Try these experiments for various values of N.

2. Who is influential?
   Let's consider a graph (data set given in class) 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.

   Bonus question: what is the complexity of the greedy algorithm?