M375T/M396C: Topics in Complex Networks **Spring 2013**

Lecture 02 — 01/17/13

Lecturer: Ravi Srinivasan *Scribe: Tyson Loudon*

2.1 Mathematical description of a network

The most elementary notion of a *network* is simply that of a graph—a set of vertices and adjoining edges. Let $V_n = \{1, \ldots, n\}$ be a set of $n \in \mathbb{N}$ nodes, and $E_n = \{(i, j) :$ there is an edge connecting $i \in V_n$ to $j \in V_n\}$ be the set of edges. We define the graph $G_n = (V_n, E_n)$. The edges connecting the nodes can be *directed* or *undirected* depending on the network under consideration. To illustrate, a network with undirected edges is the Facebook social graph (friendship is considered to be mutual), while an example with directed edges is the graph of Twitter users (users can follow others without reciprocation).

In our discussion we will often find it useful to refer to the *distance* between two arbitrary nodes. Define d_{ij} to be the length of the shortest path that connects node i and node j (if there is no such path, define $d_{ij} = +\infty$). Then, the *diameter* d of the graph G_n is defined to be the length of the "longest" shortest path," i.e.

$$\operatorname{diam}(G_n) = \max_{i,j \in V_n} d_{ij}.$$

In addition, we will say that a subset $V' \subseteq V_n$ is *connected* if there exists a path between any two nodes in V'. Equivalently, the subgraph $G' = (V', E')$ of G_n (where E' consists of all edges between nodes in V') is connected if the diameter of G' is finite.

2.2 Random graph models

To begin studying networks, we seek simple mechanisms from which we can generate interesting and mathematically tractable models. Since the typical network we observe in reality is the byproduct of random processes with some underlying structure, it seems reasonable for us to use random graphs as a first approximation. We can then determine whether the model in question has properties that we observe in empirical studies (e.g., small-world effect, clustering, power law distributions, etc.). To summartize, we use the following method:

1. For a fixed $n \in \mathbb{N}$, put a probability distribution $\mathbb{P}_n$ on the set $\mathcal{G}_n$ of undirected graphs with n nodes. Since an edge may exist between any pair of nodes, the number of such graphs is $|\mathcal{G}_n| = 2^{\binom{n}{2}}$.

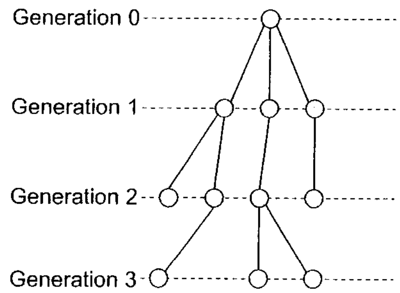

Figure 2.1. Tree structure of a Galton-Watson branching process.

2. Consider large networks, i.e., take $n \to \infty$.

3. Determine if desired properties are retained in the limiting graph. That is, for a given property A we would like to check that $\mathbb{P}_n\left(G_n \text{ satisfies property } A\right) \to 1$ as $n \to \infty$.

There is a vast array of random graph models, each of which has been proposed to reflect particular desired properties. In this course, we will only discuss a subset of such models which have played an especially significant role in improving our understanding of network structure.

2.3 Galton-Watson branching processes

The simplest example of a random graph is a *Galton-Watson branching process*, introduced by Sir Francis Galton in 1873 to represent the geneology of a population. This process is characterized by the probability distribution of the number of children of any given individual, known as the *offspring distribution.* Starting with one root node, we suppose that each node has a random number of children. Let ξ_i^n denote the number of children of the i^{th} member of the n^{th} generation. We assume that the set $\{\xi_i^n\}_{i,n\in\mathbb{N}}$ is comprised of independent and identically distributed (iid) random variables, with the same distribution as some random variable ξ. It can be seen that the resulting random graph is a tree, which we denote $\mathcal{T}$.
In our study of branching processes, we will observe what are known as *phase transitions* (sometimes referred to as *tipping points* in sociological literature). The most important of these in the context of population dynamics is a threshold for the probability p_{ext} that the population goes extinct. We will show that this transition is expressed in terms of the mean value $\mu = \mathbb{E}[\xi]$ of the offspring distribution, i.e., the average number of children of any given

individual. In particular, we observe the following three regimes

$$\begin{array}{rlcl} \text{Subcritical:} & \mu < 1 & \Longrightarrow & p_{\text{ext}} = 1 \\ \text{Critical:} & \mu = 1 \text{ and } \mathbb{P}(\xi = 1) < 1 & \Longrightarrow & p_{\text{ext}} = 1 \\ \text{Supercritical:} & \mu > 1 & \Longrightarrow & p_{\text{ext}} < 1 \end{array}$$

The intuition for this result is as follows. If each individual has, on average, less than one child then we can expect the population to become extinct since we expect each subsequent generation to consist of less individuals than the one before it. Analogously, if each individual has more than one child on average, we should expect that there is at least some nonzero chance that the population will survive. In the critical case, as long as we rule out the deterministic situation in which each individual has exactly one child, extinction should occur as well.
We will restate and prove this intuitive result next time but first we require a little more setup.

Definition. The *probability generating function* of a random variable Y is $\phi_Y(s) = \mathbb{E}\left[s^Y\right]$, which is well-defined for all $s \in \mathbb{C}$ such that $|s| < 1$.

If Y takes values in $\mathbb{N}$, then

$$\phi_Y(s) = \sum_{k=0}^{\infty} \mathbb{P}(Y = k)s^k = \mathbb{P}(Y = 0) + \sum_{k=1}^{\infty} \mathbb{P}(Y = k)s^k.$$

Notice that $\phi_Y(0) = \mathbb{P}(Y = 0)$ and $\phi'_Y(0) = \mathbb{P}(Y = 1)$. In general, we see that the distribution of Y can be recovered from the probability generating function since $\phi_Y^{(k)}(0)/k! = \mathbb{P}(Y = k)$. Note that $\phi'_Y(s)$ and $\phi''_Y(s)$ are both strictly positive for $0 \leq s \leq 1$ if $\mathbb{P}(Y = 0)$ and $\mathbb{P}(Y = 1)$ are nonzero. In this case, $\phi_Y(s)$ is an increasing function which is convex (i.e., concave up). In the next lecture, we will use these properties to study the behavior of the branching process in the regimes discussed above.

References

1. Acemoglu, D., & Ozdaglar A. (2009). 6.207/14.15, Networks, Fall 2009 [Lecture notes]. Retrieved from:
`http://economics.mit.edu/faculty/acemoglu/courses/`

2. Chaintreau, A. (2012). COMS 4995-1, Introduction to Social Networks, Fall 2012 [Lecture notes]. Retrieved from:
`http://www.cs.columbia.edu/~augustin/`

3. Draief, M., & Massoulié, L. (2010). Epidemics and rumours in complex networks. Cambridge University Press.

4. Durrett, R. (2006). Random graph dynamics (Vol. 20). Cambridge University Press.

Last edited: March 21, 2013