Lecture Notes: Social Networks: Models, Algorithms, and Applications
Lecture 1: Jan 26, 2012
Scribes: Geoffrey Fairchild and **Jason Fries**

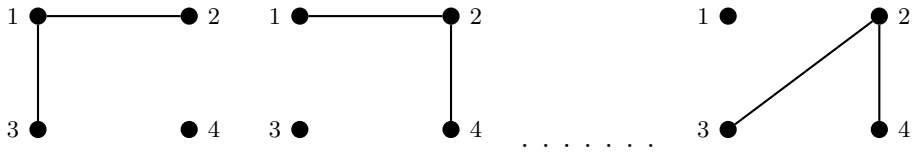
1 Random Graph Models for Networks

1.1 Graph Modeling

A *random graph* is a graph that is obtained by randomly sampling from a collection of graphs. This collection may be characterized by certain graph parameters having fixed values.

Definition 1 *$G(n, m)$ is the graph obtained by sampling uniformly from all graphs with n vertices and m edges.*

For example, given $n = 4$ and $m = 2$ with the vertex set $V = \{1, 2, 3, 4\}$ we could obtain any one of these graphs:



(a) G_1 (b) G_2 (d) G_{15}

Figure 1: Possible random graph instances for $n = 4, m = 2$ resulting in a *state space* Ω of size 15

The probability of selecting a graph G_N requires determining the size of the set of all possible graph outcomes, computed as choosing from all possible pairs of nodes n, all possible m edge combinations.

Definition 2 *The total number of possible random graphs given n vertices and m edges is*

$$|\Omega| = \frac{1}{\binom{\binom{n}{2}}{m}}$$

1.2 Erdős-Renyi Model

The above approach constitutes the *sampling view* of generating a random graph. Alternatively we can take a *constructive view* where we start with vertex set $V = \{1, 2, 3....n\}$, and selecting uniformly at random one edge from those edges not yet chosen, repeating this m times.

Definition 3 *$G(n, p)$ is the random graph obtained by starting with vertex set $V = \{1, 2, 3...n\}$, letting $0 \leq p \leq 1$, and connecting each pair vertices $\{i, j\}$ by a edge with probability p*

This model is typically referred to as the Erdos-Renyi (ER) Random Graph Model, outlined by Erdős and Renyi in two papers from 1959 and 1960 [2, 3]. While the model bears their names, their work initially examined the properties of the $G(n, m)$ model, only later expanding to analyze

the $G(n,p)$ model. Both variants were independently proposed by Solomonoff and Rapaport in 1951[5] and Gilbert in 1959[4].

In analyses, the $G(n,m)$ model is not as easy to deal with mathematically as the similar (though not exact) graph $G(n,p)$, so in practice $G(n,p)$ is more commonly used today. The equivalence of $G(n,m)$ and $G(n,p)$ can be noted by setting $\binom{n}{2}\ p = M$, and observing that as $n \to \infty$ $G(n,p)$ should behave similarly to $G(n,m)$, as by virtue of the law of large numbers, $G(n,p)$ will contain approximating the same number of edges as $G(n,m)$.

1.2.1 Probabilistic Characteristics of $G(n,p)$

Definition 4 *The expected number of edges in* $G(n,p) = \binom{n}{2}\ p$

For example, if we wanted the generate linear number of edges, sparse graphs need p should be on the order of $\frac{1}{n}$.

Definition 5 *For the distribution of number of edges in* $G(n,p)$*, let* x *be the random variable dependent on the number of edges in*

$$Prob[X = x] = \binom{\binom{n}{2}}{x}\ p^x\ (1-p)^{\binom{n}{2}-x}$$

This takes the form of a binomial distribution, and the implication of this definition is that edges are concentrated around the mean with high probability.

Definition 6 *The expected degree of* $G(n,p) = (n-1)\ p$

Definition 7 *For the degree distribution of* $G(n,p)$*, fix a vertex* v *and let* y *be the number of edges incident on* v

$$Prob[Y = y] = \binom{n-1}{y} p^y (1-p)^{n-n-y}$$

This is why the Erdős-Renyi graphs are said to have a binomial degree distribution.

Definition 8 *We can use a poisson approximation to compute an expected degree distribution of* $G(n,p)$ *as follows:*

If we fix $(n-1)p$ *to a constant* c *– the expected degree – then*

$$\binom{n-1}{y}\ p^y\ (1-p)^{n-1-y} \longrightarrow \underbrace{\frac{c^y e^{-c}}{y!}}_{\text{poisson distribution with paramter } c}$$

Definition 9 *The expectation of the local clustering clustering coefficient in* $G(n,p)$ *is* p

Recall the definition of the *local clustering coefficient* as:

$$cc(v) = \frac{\text{pairs of neighbors of } v \text{ connected by edges}}{\text{total pairs of } v}$$

The expected value of $cc(v)$ is calculated as

$$\mathrm{E}[cc(v)] = \sum_{d=0}^{n-1} \mathrm{E}[\ cc(v) \mid deg(v) = d\ ]\ Prob[\ deg(v) = d\ ]$$

Observe that the conditional expectation of cc(v) given deg(v)=d reduces to p

$$\mathrm{E}[\ cc(v) \mid deg(v) = d\ ] = \frac{p\binom{d}{2}}{\binom{d}{2}} = p$$

Leaving the equation

$$p \sum_{d=0}^{n-1} Prob[\ deg(v) = d\ ]$$

The sum probability of all possible outcomes is, of course, equal to 1, leaving our final equation as

$$p \cdot 1 = p$$

Given a formal definition of the clustering coefficient cc for a random graph $G(n, p)$ we can revisit the 1998 paper in Nature by Watts and Strogatz[6] and now compute the cc of a corresponding random graph.

	N	average degree	cc_1	cc_1 of corresponding random graph
actors network	225226	61	0.79	0.00027
power grid	4941	2.67	0.080	0.005
C. elegans	282	14	0.28	0.05

Table 1: Comparing observed networks against "corresponding" random graphs.

For example,the corresponding random graph for the actors' network would be

$$\begin{aligned} n &= 225226 \\ c = (n-1)p &= 61 \\ p = 61/225225 &= 0.00027 \end{aligned}$$

1.2.2 "Small World" Property of $G(n, p)$

We will show that the diameter of $G(n, p)$ is

$$ln_c\ n \frac{ln\ n}{ln\ c}, \text{where } c = p(n-1)$$

For example, consider an *acquaintance network* of every human being on earth, currently estimated at 7 billion people. If every individual has, on average, 1000 acquaintances, our graph diameter is calculated as

$$\frac{ln\ 7x10^9}{ln\ 1000} = 3.33...$$

Definition 10 *The diameter of graph G(V,E), where distance = the shortest path between u, v is*

$$\max_{u,v \in V}\ distance(u, v)$$

Remember however that we are acting on random graphs, meaning that diameter is itself a random variable. The diameter referred to here more correctly thought of as the *expected diameter* of graph G, formally stated as

$$Prob[\ distance(u, v) > \frac{ln\ n}{ln\ c}\ ] \longrightarrow 0 \text{ as } n \to \infty$$

Note that, counter perhaps to our intuition, this expected diameter value does not lie in the middle of roughly an equal number of graphs with low diameter and and graphs with high diameters. In reality, as $n \to \infty$, there are a diminishing number of graphs with a diameter larger than $\frac{ln\ n}{ln\ c}$. A formal proof[1] of the expected diameter of a random graph is outside the scope of this text, but we can construct a heuristic argument that gives some intuition into the problem.

Fix a vertex v, and $c = (n-1)p$. Divide our graph into two sets of nodes, *reached* and *remaining*. At each level we continuing adding edges to unreached nodes such that the number of vertices reachable in s hops us c^s, where $c^s = n$ and $s = \frac{ln\ n}{ln\ c}$.

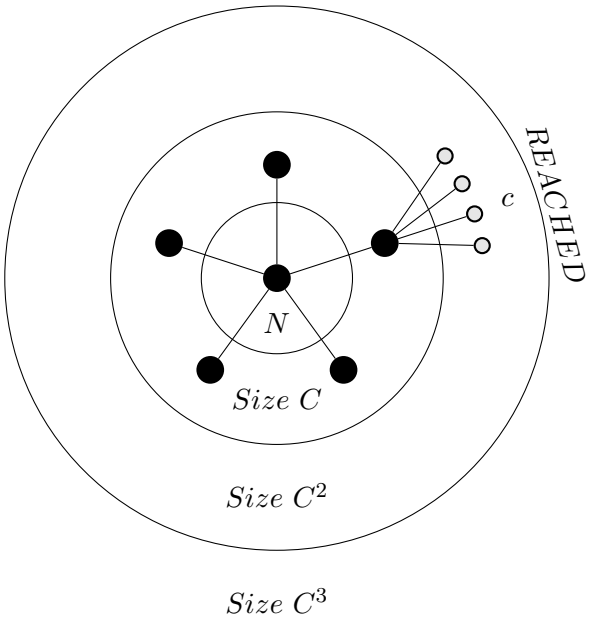


Figure 2: A "heuristic" argument proof for expected graph diameter. As our graph grows we add unreached nodes to add to our graph.

As a heuristic proof, there are of course problems with this . Eventually our *reached* set will be larger than our *remaining*, for example. Next class will discuss some of these points and ways to address them.

References

[1] B. Bollobás. *Random graphs*, volume 73. Cambridge Univ Pr, 2001.

[2] P. Erdös and A. Rényi. On random graphs, i. *Publicationes Mathematicae (Debrecen)*, 6:290–297, 1959.

[3] P. Erdős and A. Rényi. *On the evolution of random graphs*. Akad. Kiadó, 1960.

[4] E.N. Gilbert. Random graphs. *The Annals of Mathematical Statistics*, 30(4):1141–1144, 1959.

[5] R. Solomonoff and A. Rapoport. Connectivity of random nets. *Bulletin of Mathematical Biology*, 13(2):107–117, 1951.

[6] Duncan J. Watts and Steven H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393:440–442, 1998.