Random Graphs III
Dmitri Loguinov
Texas A&M University

March 28, 2017
Agenda

• Random-graph models
  - $G(n, k_{out})$
  - $G(n, r)$

• Percolation of monotone properties

• Relationship between connectivity and degree
Random Graph Models 9

• The next model is \(G(n, k_{\text{out}})\)
  - Each node randomly selects \(k\) nodes from the remaining \(n - 1\) vertices to be its neighbors
  - This is a **directed** graph
  - The out-degree is \(k\), but what is the in-degree?

• The probability that node \(v\) is selected by \(u\) is:

\[
p_{uv} = 1 - \prod_{i=1}^{k} \left(1 - \frac{1}{n - i}\right) = \frac{k}{n - 1}
\]

• Since there are \(n - 1\) nodes \(u\), the in-degree of \(v\) is:

\[
d_v^{\text{in}} \sim B(n - 1, p_{uv})
\]
Random Graph Models 10

- Easy to notice that $E[d_v^{in}] = k$

- Graphs $G(n,k_{out})$ are often found in user-driven networks with freedom to choose neighbors
  - A prime example would be a P2P system, where each joining user randomly selects $k$ other peers to be his/her neighbors

- Another version of this model is the social network
  - Each person randomly selects $k$ other people in the world to be their friends

- **Q:** What is the probability that each person in an $n$-node network is connected to every other one through a chain of acquaintances?
The final model is $G(n,r)$

- A fairly recent development that models wireless ad-hoc networks (i.e., users form a graph using each other’s computers as routers)
- Two nodes $u$ and $v$ are linked if and only if $||u - v|| \leq r$
- Parameter $r$ is the communication radius and $||.||$ is some norm (usually Euclidean) in the 2D space
Random Graph Models 7

- Formalization
  - Coordinates $(x,y)$ of each node in $G(n,r)$ are iid uniform random variables in $[0,A]$.
  - Nodes scattered in a 2D square whose area is $A^2$.
  - Edge $(u,v)$ exists iff distance between $u$ and $v$ is less than $r$.

- Define node density $\rho$: 
  $$\rho = \frac{n}{A^2}$$

- Number of nodes in an area of size $S$?
  - Binomial with parameters $(n, S/A^2)$.
  - The mean is simply $\rho S$. 
• **Q:** What is the expected number of neighbors (degree) of each node?

\[ E[d_i] = \rho \pi r^2 - 1 \]

• **Q:** How many wireless sensors with a communication radius of 200 feet does one need to scatter in a zone 50x50 miles to obtain a connected network with high probability?

• This relatively simple question requires rather complex mathematical tools, which we omit here

Graph Properties

- Definition: a property $Q$ is **monotone** if whenever $G$ has $Q$, every graph with the same number of nodes containing $G$ also has $Q$
  - Monotonicity applies to addition of new edges

- Which of these are monotone?
  - Graph contains a triangle (3-cycle)
  - Graph contains a complete sub-graph of order 5
  - The number of edges is odd
  - Graph is connected

- We will understand connectivity of random graphs using an **edge-growth** process similar to $G(n,M)$
Graph Properties 2

- **Theorem**: if $Q$ is monotone and $M_1 < M_2$, then:
  \[ P(G(n, M_1) \text{ has } Q) \leq P(G(n, M_2) \text{ has } Q) \]

- **Theorem**: if $Q$ is monotone and $p_1 < p_2$, then
  \[ P(G(n, p_1) \text{ has } Q) \leq P(G(n, p_2) \text{ has } Q) \]

- **Definition**: we say that almost every (a.e.) graph has $Q$, if $P(\text{graph has } Q) \to 1$ as $n \to \infty$
  - We may occasionally produce graphs in which $Q$ does not hold, but the fraction of such graphs tends to 0 as $n \to \infty$

- Alternatively we may say that $Q$ holds almost surely
  - Meaning with probability $1-o(1)$ as $n \to \infty$
Graph Properties 3

• Erdos was first to discover that certain monotone properties in $G(n, M(n))$ appeared “suddenly”
  – There is a threshold function $M^*(n)$ such that:

\[
\frac{M(n)}{M^*(n)} \rightarrow 0 \Rightarrow P(G(n, M(n)) \text{ has } Q) \rightarrow 0
\]

  – and:

\[
\frac{M(n)}{M^*(n)} \rightarrow \infty \Rightarrow P(G(n, M(n)) \text{ has } Q) \rightarrow 1
\]

• Example: $M^*(n) = n^{1.5}$; find some examples of $M(n)$ for each of the two cases above
Graph Properties 4

• Threshold functions are not unique and differ by a fixed factor:
  – For any two threshold functions
    \[ M_1^*(n) = O(M_2^*(n)), \quad M_2^*(n) = O(M_1^*(n)) \]

• To better understand behavior of random graphs, examine their evolution as defined below:
  – Graph \( G_0 \) starts with \( n \) nodes and no edges
  – \( G_t \) is obtained from \( G_{t-1} \) by adding a random edge
  – \( G_t \) has \( t \) edges and is called a graph process

• Clearly, \( G_0 \subset G_1 \subset \ldots \subset G_t \)
Graph Properties 5

• One particularly interesting monotone property \( Q \) of a graph process \( G_t \) is its connectivity
  - Once the graph is connected at some time \( t_Q \), it stays connected throughout the remainder of the process

• More formally, \( t_Q \) is called the hitting time of \( Q \):
  \[
  t_Q = \min \{ t \geq 0 : G_t \text{ has } Q \}
  \]

• Theorem: the hitting time is almost always “close” to the threshold function \( M^*(n) \) of \( Q \)
  - The closeness is in the asymptotic sense
Our first main result is the threshold function for graphs $G(n,p)$ and $G(n,M)$

- Both thresholds are equivalent to accumulation of average degree $\log(n)$
- However, we can do even better

**Theorem**: assume that

$$M = \frac{n(\log n + c(n))}{2}, \quad p = \frac{\log n + c(n)}{n}$$

- then:

$$P(G \text{ is connected}) \to e^{-e^{-c(n)}}$$
Graph Properties 7

• As soon as $G_t$ accumulates $n(\log n + c(n))/2$ edges, it becomes connected:
  - With probability 1 if $c(n) \to \infty$
  - With probability 0 if $c(n) \to -\infty$

• For example, $M = n(\log n + \log\log n)/2$ is sufficient for connectivity of a.e. graph
  - Similarly, $M = n(\log n - \log\log n)/2$ guarantees that almost no graph is connected

• Q: What is the asymptotic probability of connectivity of random graphs under these conditions?
  - Suppose $p = 2\log(n)/n$, or $p = \log(n)/n$ or $p = \sqrt{n}/n$
  - What if $p = (\log^2(n) - 3\log(n))/n$?
• It is often convenient to express thresholds in terms of degree
  - Recall that expected degree $E[d] = 2M/n \approx np$

• Restating the same thresholds, we have:

$$E[d] = \log n + c(n) \Rightarrow P(G \text{ is connected}) = e^{-e^{-c}}$$

  - If $c$ tends to $+\infty$ or $-\infty$, we have two options: 1) a.e. graph is connected or 2) a.e. graph is disconnected

• This effect is known as **percolation**
  - Originated in physics under percolation theory
  - Used to model forest fires, social networks, and various physical/chemical problems
Graph Properties 9

• In general, how does $G_t$ look for different $t$?
  - This is a stochastic process with a vast number of properties

• Theorem: as time $t$ evolves, $G_t$ goes through the following three stages:
  - 1) If $E[d(t)] < 1$, then a.e. component is a tree (i.e., no cycles in the graph) and the size of each component is $O(\log n)$
  - 2) As soon as $E[d(t)]$ becomes larger than 1 (i.e., $E[d] = c > 1$) cycles start to emerge and we have the formation of one giant component of size $\Theta(n^{2/3})$; all smaller components are of size $O(\log n)$
  - 3) The giant component grows 4 times faster than $t$ and “swallows” all other components by the time $E[d(t)]$ reaches and slightly exceeds $\log n$