Giant Component and Component Size

Bowen Gong | 10/20/2019
Presentation Structures

- Introduction & motivation
- Significant formulas
- Characteristics of giant component
  - Uniqueness of giant component
  - Component size distribution
  - Average component size
- NetLogo simulation
- ER networks simulations
Introduction

- **Random network**
  - G(N, L): L links will be randomly placed, connecting N nodes.
  - G(N, p): probability connecting 2 nodes is p.
  - Probability of N-node network having L link:
    - \[ p_L = \left( \frac{L}{N(N-1)} \right) p^L (1 - p)^{\frac{N(N-1)}{2} - L} \]
  - Expectation of #links in a random network:
    - \[ \langle L \rangle = p \frac{N(N-1)}{2} \]
    - \[ \langle k \rangle = p(N - 1) \]

- The mean degree in a graph with m edges is \( 2m/n \)
- \[ \langle k \rangle = \sum_0^{\binom{n}{2}} \frac{2m}{n} P(m) = (n - 1)p \]
Introduction

- **Random network**
  - Binomial model describing the possibility of a node having degree $k$
    - $p_k = \binom{k}{N-1}p^k(1-p)^{N-1-k}$
  - Poisson model if $k \ll N$:
    - $p_k = e^{-\langle k \rangle} \frac{\langle k \rangle^k}{k!}$

Motivation

Graph structure in the web network

- A giant web network with 200 million pages (nodes) and 1.5 billion links (edges) was studied.
- Over 90% nodes are in a single component.
- Giant component is made of 4 parts. (giant strongly connected component, IN, OUT, Tendrils)

Web's macroscopic structure*

Motivation

Giant Component in Real Network*

- Astrophysics citation network: 99% nodes in weak connected component (WCC), 37% in strongly connected component (SCC)*

- Wikipedia talk page: 99% nodes in WCC, 30% in SCC*

- Enron email network: 92% nodes in the giant component.

- Yeast protein interaction: 93% nodes in the giant component

- How does the giant component form? What is its size distribution?

*From the notes of CSE 158, Web Mining and Recommender Systems, lecture 12, Julian McAuley, UCSD
Clustering happens when more and more nodes were connected.

Size of largest cluster > threshold, a **large cluster** emerges rapidly.

Take a closer look:

- $p=0$, no edge & no connection & cluster & largest cluster size 1
- $p=1$, 1 cluster & largest cluster size $N$
- in general, if $u$ as prob. a node not in giant component
- $u_1 = 1 - p + pu$

Size of largest is independent of $N$:

\[ u = u_1^{(N-1)} = (1-p+pu)^{N-1} \]

Size of largest is dependent of $N$:
Significant Formula

- Sometimes to have a large component in the network, occupying most of the network. E.g. network, transportation. Meaningless if nodes are not connected.

- Is there a transition point between $p = 0$ and $p = 1$?

- Absolutely!

- Define giant component as the network component with size grows in proportion to graph size $N$. 
Occurrence of Giant Network

- \( u = u_1^{(N-1)} = (1-p+pu)^{(N-1)} \)
- Replace \( p = \langle k \rangle / N \), \( u = 1 - S \). \( S \) is \( N_G/N \):
  - \( S = 1 - e^{-\langle k \rangle S} \) implicit func. on \( S \).
  - Critical point: \( \langle k \rangle = 1 \)
    - \( \langle k \rangle < 1 \) no giant component
    - \( \langle k \rangle \geq 1 \) has giant component
  - Simulation (right fig): 10000 nodes, 40 realizations
Occurrence of Giant Component

- Five stages

1. $p < \frac{\langle k \rangle}{N}$, $0 < \langle k \rangle < 1$
   - All components are either tree or single node
   - Largest component is a tree, the largest component size will not exceed $O(\log(N))$
   - Numerous small component
   - No giant component in graph
   - $N_G/N = 0$
Occurrence of Giant Component

• Five stages

2. $p < 1/N$, $\langle k \rangle = 1$
   - $N_G/N = 0$
   - Giant component in the corner
   - Largest component size
     - $((\langle k \rangle - 1 - \log\langle k \rangle)^{-1}(\log\langle k \rangle - 2.5 \log\log N)$, approximately $N^{2/3}$, no other component will contain more than $O(\log(N))$
     - A jump with several magnitudes can be observed
Occurrence of Giant Component

- Five stages:
  3. $p > 1/N, \langle k \rangle > 1$
    - One giant component distinguishable
    - All other components small
    - Giant component keeps absorb smaller components.
    - The larger the $\langle k \rangle$, the larger the giant component.
Occurrence of Giant Component

- Five stages:

4. \( p > \frac{\ln N}{N}, \langle k \rangle > \ln N \)

  - Almost one giant component
  - Minor isolated nodes
Occurrence of Giant Component

- Five stages:

5. $p$ keeps increasing
   - Nodes not only connected, but also the average degree of the nodes is going to be equal.
Characteristics of Giant Component

- **Existence of a giant component when** $\langle k \rangle > 1$

  - Heuristic argument: Suppose there is a small set of connected vertices in the graph.
  - The set is composed with core and periphery $P$.
  - Image adding vertices from periphery to the core, if $s$ denotes (core + periphery), each time $P_{k+1}$ becomes:
    
    $$P_{k+1} (N - s) = p (n - s) * P_k = c * P_k$$

  - If $c > 1$, it grows exponentially. Will be a point that the size of the component will become a significant fraction of the network. (giant network)

* Paris Siminelakis lecture note, University of Athens
Characteristics of Giant Component

- **Uniqueness of the giant component**
  - There is only one giant component in the graph.
  - Assume there are two giant components with size: $S_1N$, $S_2N$
  - If they are separate, no edge shall be allowed between them.
  - Then, the probability is:
    \[
    P(S_1 \neq S_2) = \left(1 - \frac{\langle k \rangle}{n-1}\right)^{S_1S_2N^2} \sim 0
    \]
  - There is only one unique giant component exists.

*Paris Siminelakis lecture note, University of Athens*

Probability of the random network having two giant components is trivial *
Characteristics of Giant Component

- **Component size distribution**
  - Apart from the giant component, the other nodes are in the form of small components.
  - Assume $\pi_s$ to be the fraction of small component that has size $s$, size distribution follow:
    - $\pi_s = \frac{e^{-s\langle k \rangle} (s \langle k \rangle)^{s-1}}{s!}$
    - When $\langle k \rangle$ is small, no giant component in the graph. Small component with size 1 occupy large number of nodes.
    - When $\langle k \rangle$ is equal or larger than 1, there is a unique giant component. Probability vs. size exponentially decay.
    - $\pi_s \approx s^{-1.5} e^{-\langle k \rangle - 1) s + (s - 1) \ln \langle k \rangle}$
    - When $\langle k \rangle = 1$, $\pi_s = s^{-1.5}$
Characteristics of Giant Component

- **Average component size**
  - Average component size is not variant with size of random network \( N \).
  - Average component size (biased):
    - \( \langle s \rangle = \frac{1}{1 - \langle k \rangle + \frac{\langle k \rangle N_G}{N}} \)
    - Bias comes from the fact that node is more likely to join component with larger size.
  - After the correction, average component size (not biased) can be calculated with:
    - \( \langle s \rangle = \frac{2}{2 - \langle k \rangle + \frac{\langle k \rangle N_G}{N}} \)
N = 150

Pay attention to how fast the size of giant component change when $\langle k \rangle > 1$
ER Network Simulation

- 10000 nodes, 100 realizations, take average:

- Focus on critical point:
  - (1) $p = 1/N$, $\langle k \rangle = 1$;
    - Size increases quickly when $\langle k \rangle \Rightarrow 1$
    - Exponentially increase
  
  - (2) $p = \ln(N)/N$, $\langle k \rangle = \ln(N)$;
    - Size not in giant component is small ($\sim 1$)
    - Size keeps decreasing when $\langle k \rangle$ increases
Conclusion

• Basic intro to random network and significant formulas
• Existence of giant component in a graph
• Uniqueness of giant component (only 1)
• Component size of giant component
• Average component size
• NetLogo simulation
• ER network showing fast increase.

• The preparation of the slides is greatly impacted by lecture notes entitled “Networks and Random Graphs” given by Paris Siminelakis in School of Electrical and Computer Engineering at the National Technical University of Athens.
• Also see the book *Network An Introduction* by Mark Newman from University of Michigan.
Thank you