Large Cliques in a Power-law Random Graph

Svante Janson    Tomasz Luczak    and Ilkka Norros

September 24, 2009

Presented by: Dung Nguyen
Problems

- What is the size of largest cliques on power-law random graph?
- Does there exist an efficient algorithm which find nearly optimal maximum cliques with high probability (whp)?
1. Poissonian Model

2. Greedy Algorithms

3. Main Results
   - Theorems
   - Proof

4. Conclusion
Each node $i$ is assigned weight $W_i$ randomly with power-law tail distribution

$$P(W > x) = ax^{-\alpha}, \ x \geq x_0$$

Largest weight $W_{\text{max}} = \max_i W_i$

$$P(W_{\text{max}} > tn^{1/\alpha}) \leq nP(W > tn^{1/\alpha}) = O(t^{-\alpha})$$

Number of edges between each pair $\{i,j\}$ of vertices is Poisson distributed random number with expectation:

$$E(E_{ij}) = \lambda_{ij} = b \frac{W_i W_j}{n}$$

Delete duplicate edges to get simple graph, vertices $i$ and $j$ are joined by an edge with probability

$$p_{ij} = 1 - e^{-\lambda_{ij}}$$
Greedy algorithms

- Greedy 1: Check the vertices in order of decreasing weights and select every vertex that is adjacent to every selected vertex. Denote the output is greedy clique $K_{gr}$.

- Greedy 2: Check the vertices in order of decreasing weights and select every vertex that is adjacent to every vertex with higher weight. Denote the output is quasi top clique $K_{qt}$.

- Greedy 3: Stop with first failure. Denote the output is full top clique $K_{ft}$.

\[ |K_{ft}| \leq |K_{qt}| \leq |K_{gr}| \leq |K_{max}| \]

where $K_{max}$ is the largest clique.
Denote the size of largest cliques is $\omega(G(n, \alpha))$.

**Theorem (1)**

(i) If $0 < \alpha < 2$, then

$$\omega(G(n, \alpha)) = (c + o_p(1))n^{1-\alpha/2}(\log n)^{-\alpha/2},$$

where $c = ab^{\alpha/2}(1 - \alpha/2)^{-2}$

(ii) If $\alpha = 2$, then $\omega(G(n, \alpha)) = O_p(1)$; that is, for every $\epsilon > 0$

there exists a constant $C_\epsilon$ such that $P(\omega(G(n, \alpha)) > C_\epsilon) < \epsilon$ for every $n$. However, there is no fixed finite bound $C$ such that

$\omega(G(n, \alpha)) \leq C$ whp

(iii) If $\alpha > 2$, then $\omega(G(n, \alpha)) \in \{2, 3\}$ whp. Moreover, the

probabilities of each of the events $\omega(G(n, \alpha)) = 2$ and

$\omega(G(n, \alpha)) = 3$ tend to positive limits.
Ratio of greedy algorithms

Theorem (2)

If $0 < \alpha < 2$, then $K_{gr}$ and $K_{qt}$ both have size $(1 + o_p(1))\omega(G(n, \alpha))$; in other words

$$|K_{gr}|/|K_{max}| \xrightarrow{p} 1 \text{ and } |K_{qt}|/|K_{max}| \xrightarrow{p} 1.$$  

On the other hand,

$$|K_{ft}|/K_{max} \xrightarrow{p} 2^{\alpha/2}.$$  

Corollary (3)

For every $\alpha > 0$ there exists an algorithm which whp finds in $G(n, \alpha)$ a clique of size $(1 + o(1))\omega(G(n, \alpha))$ in polynomial time.
Case $\alpha < 2$

Partition the vertex set to "dense set" and "sparse set"

$$V_s^- = \{ i : W_i \leq s\sqrt{n \log n} \} \text{ and } V_s^+ = \{ i : W_i > s\sqrt{n \log n} \}$$
Case \( \alpha < 2 \)

- Partition the vertex set to "dense set" and "sparse set"
  
  \[ V_s^- = \{ i : W_i \leq s\sqrt{n\log n} \} \text{ and } V_s^+ = \{ i : W_i > s\sqrt{n\log n} \} \]

- \( |V_s^-| \) is large but the size of its maximum clique is "small".
- \( |V_s^+| \) is small but the size of its maximum clique is "large".

\[ p_{ij} \leq 1 - n^{-bs^2}, \text{ if } i, j \in V_s^- \]
\[ p_{ij} > 1 - n^{-bs^2}, \text{ if } i, j \in V_s^+ \]
Case $\alpha < 2$

- Partition the vertex set to "dense set" and "sparse set"
  \[ V_s^- = \{ i : W_i \leq s\sqrt{n\log n} \} \quad \text{and} \quad V_s^+ = \{ i : W_i > s\sqrt{n\log n} \} \]
- $|V_s^-|$ is large but the size of its maximum clique is "small".
- $|V_s^+|$ is small but the size of its maximum clique is "large".
  \[
  p_{ij} \leq 1 - n^{-bs^2}, \quad \text{if } i, j \in V_s^-
  \]
  \[
  p_{ij} > 1 - n^{-bs^2}, \quad \text{if } i, j \in V_s^+
  \]
- Which $s$ is suitable?
Case $\alpha < 2$

- $|V_s^+| = (1 + o(1))as^{-\alpha} n^{1-\alpha/2} \log n^{-\alpha/2}$ whp
- $\omega(G[V_s^-]) \leq 2n^{bs^2} \log n$ whp
- $\omega(G(n, \alpha)) \leq \omega(G[V_s^-]) + |V_s^+|$ whp
Case $\alpha < 2$

- $|V_s^+| = (1 + o(1))as^{-\alpha}n^{1-\alpha/2} \log n^{-\alpha/2}$ whp
- $\omega(G[V_s^-]) \leq 2n^{bs^2} \log n$ whp
- $\omega(G(n, \alpha)) \leq \omega(G[V_s^-]) + |V_s^+|$ whp
- Choose $s$ to match the exponent of $n$ in two components of the right side
- $s = (1 - \epsilon)b^{-1/2}(1 - \alpha/2)^{1/2}$ whp
- $\omega(G(n, \alpha)) \leq (1 + o(1))(1 - \epsilon)^{-\alpha}cn^{1-\alpha/2} \log n^{-\alpha/2}$ whp
Case $\alpha < 2$

- Use $K_{qt}$ as a lower bound
- $V_s^+$ is dense, almost vertices in $V_s^+$ belong to $K_{qt}$. Choose $s = (1 + \epsilon)b^{-1/2}(1 - \alpha/2)^{1/2}$:
  \[ |V_s^+ \setminus K_{qt}| \leq Cn^{-\epsilon(1-\alpha/2)}|V_s^+ \setminus K_{qt}| \]
Case $\alpha < 2$

- Use $K_{qt}$ as a lower bound
- $V^+_s$ is dense, almost vertices in $V^+_s$ belong to $K_{qt}$. Choose $s = (1 + \epsilon) b^{-1/2}(1 - \alpha/2)^{1/2}$:
  \[|V^+_s \setminus K_{qt}| \leq Cn^{-\epsilon(1-\alpha/2)} |V^+_s \setminus K_{qt}|\]
- $\omega(G(n, \alpha)) \geq |K_{gr}| \geq |K_{qt}| \geq |V^+_s| - |V^+_s \setminus K_{qt}|$
  \[= (1 + o(1))(1 + \epsilon)^{-\alpha} cn^{1-\alpha/2} \log n^{-\alpha/2}\]
Case $\alpha = 2$

- If $\omega(G(n, \alpha)) > m$ then number of cliques of size 4 is greater than $\binom{m}{4}$
- Estimate the number of cliques of size 4
- Calculate the number of 4-vertex cliques on two condition: $W_{\text{max}} < An^{1/2}$ and other.
  
  \[
  E(X_4 | \{W_i\}_i^n) \leq b^6 (n^{-3/2} \sum_i W_i^3)^4 \\
  E(n^{-3/2} \sum_i W_i^3; W_{\text{max}} \leq An^{1/2}) = O(nAn^{1/2}) \\
  P(n^{-3/2} \sum_i W_i^3 > t) \leq t^{-1} E(n^{-3/2} \sum_i W_i^3; W_{\text{max}} \leq An^{1/2}) + P(W_{\text{max}} > An^{1/2}) \leq CA t^{-1} + CA^{-2}
  \]
Case $\alpha = 2$

- If $\omega(G(n, \alpha)) > m$ then number of cliques of size 4 is greater than $\binom{m}{4}$
- Estimate the number of cliques of size 4
- Calculate the number of 4-vertex cliques on two condition: $W_{\text{max}} < An^{1/2}$ and other.

\[
E(X_4|\{W_i\}_i^n) \leq b^6(n^{-3/2} \sum_i W_i^3)^4
\]
\[
E(n^{-3/2} \sum_i W_i^3; W_{\text{max}} \leq An^{1/2}) = O(nAn^{1/2})
\]
\[
P(n^{-3/2} \sum_i W_i^3 > t) \leq t^{-1}E(n^{-3/2} \sum_i W_i^3; W_{\text{max}} \leq An^{1/2}) + P(W_{\text{max}} > An^{1/2}) \leq CA^{-1} + CA^{-2}
\]
- Choose $A = t^{1/3}: n^{-3/2} \sum_i W_i^3 = O_p(1)$
- $\omega(G(n, \alpha)) = O_p(1)$
Approximation algorithm

- Test all group of 4 vertices
- Number of 4-vertex cliques is less than loglogn with high probability
- Test all sets of of 4-vertex cliques
Case $\alpha > 2$

- Same method as the case $\alpha = 2$, $P(\omega(G(n, \alpha) \geq 4) \rightarrow 0$
- $\omega(G(n, \alpha)) \leq 3$ whp
Case $\alpha > 2$

- Same method as the case $\alpha = 2$, $P(\omega(G(n, \alpha) \geq 4) \rightarrow 0$
- $\omega(G(n, \alpha)) \leq 3$ whp
- $P(\omega(G(n, \alpha)) = 2) \rightarrow e^{-\frac{1}{6}(bE(W^2))^3}$
- $P(\omega(G(n, \alpha)) = 3) \rightarrow 1 - e^{-\frac{1}{6}(bE(W^2))^3}$
Conclusion

- Weight of each vertex is assigned randomly, making the model more flexible.
- Partitioning the vertex set can exploit the connectivity property.

There are transition points in the size of cliques. The size of connected components is greater than some threshold. Are these exponent factors of real networks the transition points of some property?
Conclusion

- Weight of each vertex is assigned randomly, making the model more flexible.
- Partitioning the vertex set can exploit the connectivity property.
- Apply partitioning methods to problems related to connectivity properties.
- There are transition points in the size of cliques, size of connected components. Are exponent factors of real networks transition points of some property?