Global clustering coefficient in scale-free networks

Liudmila Ostroumova Prokhorenkova

Yandex

NET 2015

Clustering coefficient

Global clustering coefficient of a graph G:

$$C_1(G) = \frac{3\#(\text{triangles in } G)}{\#(\text{pairs of adjacent edges in } G)}$$

Clustering coefficient

Global clustering coefficient of a graph G:

$$C_1(G) = \frac{3\#(\text{triangles in } G)}{\#(\text{pairs of adjacent edges in } G)}$$

Average local clustering coefficient

- T^i is the number of connected neighbors of a vertex i
- P_2^i is the number of pairs of neighbors
- $C(i) = \frac{T^i}{P_2^i}$ is the local clustering coefficient for a vertex i
- $C_2(G) = \frac{1}{n}\sum_{i=1}^{n} C(i)$ – average local clustering coefficient

Motivation

Generalized preferential attachment models[1]:

- Power-law degree distribution with parameter $1+\gamma$, $\gamma > 1$
- Constant average local clustering for any $\gamma > 1$
- Constant global clustering coefficient for $\gamma > 2$

[1] L. Ostroumova, A. Ryabchenko, and E. Samosvat, *Generalized preferential attachment: tunable power-law degree distribution and clustering coefficient*, Proc. WAW'13, 2013.

Motivation

Generalized preferential attachment models[1]:

- Power-law degree distribution with parameter $1 + \gamma$, $\gamma > 1$
- Constant average local clustering for any $\gamma > 1$
- Constant global clustering coefficient for $\gamma > 2$

Challenge: suggest a model with power-law degree distributions with $1 < \gamma < 2$ and constant global clustering coefficient

[1] L. Ostroumova, A. Ryabchenko, and E. Samosvat, *Generalized preferential attachment: tunable power-law degree distribution and clustering coefficient*, Proc. WAW'13, 2013.

Motivation

Generalized preferential attachment models[1]:

- Power-law degree distribution with parameter $1+\gamma$, $\gamma > 1$
- Constant average local clustering for any $\gamma > 1$
- Constant global clustering coefficient for $\gamma > 2$

Challenge: suggest a model with power-law degree distributions with $1 < \gamma < 2$ and constant global clustering coefficient

Answer: it is impossible

[1] L. Ostroumova, A. Ryabchenko, and E. Samosvat, *Generalized preferential attachment: tunable power-law degree distribution and clustering coefficient*, Proc. WAW'13, 2013.

Scale-free graphs

- We consider a sequence of graphs $\{G_n\}$
- Each graph G_n has n vertices
- The degrees ξ_i are i.i.d. random variables following a *regularly varying* distribution:

$$1 - F(x) = L(x)x^{-\gamma}, \quad x > 0,$$

where $L(\cdot)$ is a slowly varying function: $\lim_{x\to\infty} \frac{L(tx)}{L(x)} = 1$ for any $t > 0$

Scale-free graphs

- We consider a sequence of graphs $\{G_n\}$
- Each graph G_n has n vertices
- The degrees ξ_i are i.i.d. random variables following a *regularly varying* distribution:

$$1 - F(x) = L(x)x^{-\gamma}, \quad x > 0,$$

  where $L(\cdot)$ is a slowly varying function:
  $\lim_{x\to\infty} \frac{L(tx)}{L(x)} = 1$ for any $t > 0$
- Consider only the case $1 < \gamma < 2$

Scale-free graphs

- We consider a sequence of graphs $\{G_n\}$
- Each graph G_n has n vertices
- The degrees ξ_i are i.i.d. random variables following a *regularly varying* distribution:

$$1 - F(x) = L(x)x^{-\gamma}, \quad x > 0,$$

  where $L(\cdot)$ is a slowly varying function:
  $\lim_{x\to\infty} \frac{L(tx)}{L(x)} = 1$ for any $t > 0$
- Consider only the case $1 < \gamma < 2$
- If $\sum_{i=1}^{n} \xi_i$ is odd, then we replace ξ_n by $\xi_n + 1$

Simple graphs: results

With high probability:

1. A simple graph G_n with such degree distribution exists
2. Global clustering coefficient tends to zero for **any** sequence of such graphs

Existence

Theorem

For any δ such that $1 < \delta < \gamma$ with probability $1 - O\left(n^{1-\delta}\right)$ there exists a simple graph on n vertices with the degree distribution defined above.

Existence

Theorem

For any δ such that $1 < \delta < \gamma$ with probability $1 - O\left(n^{1-\delta}\right)$ there exists a simple graph on n vertices with the degree distribution defined above.

Theorem [Erdős–Gallai]

A sequence of non-negative integers $d_1 \geq \ldots \geq d_n$ can be represented as the degree sequence of a finite simple graph on n vertices if and only if

1. $d_1 + \ldots + d_n$ is even;
2. $\sum_{i=1}^{k} d_i \leq k(k-1) + \sum_{i=k+1}^{n} \min(d_i, k)$ holds for $1 \leq k \leq n$.

Global clustering coefficient: upper bound

Theorem

For any $\varepsilon > 0$ and any α such that $0 < \alpha < \frac{1}{\gamma+1}$ with probability $1 - O(n^{-\alpha})$

$$C_1(G_n) \leq n^{-\frac{(2-\gamma)}{\gamma(\gamma+1)}+\varepsilon}$$

Upper bound: proof

$$C_1(G_n) = \frac{3 \cdot T(n)}{P_2(n)}$$

Upper bound: proof

$$C_1(G_n) = \frac{3 \cdot T(n)}{P_2(n)}$$

For any x:

$$T(n) \leq \left|\{i : \xi_i > x\}\right|^3 + \sum_{i:\xi_i \leq x} \xi_i^2$$

Upper bound: proof

$$C_1(G_n) = \frac{3 \cdot T(n)}{P_2(n)}$$

For any x:

$$T(n) \le \left|\{i : \xi_i > x\}\right|^3 + \sum_{i : \xi_i \le x} \xi_i^2$$

With probability $1 - O\left(\frac{x^\gamma}{n\, L(x)}\right)$:

$$\left|\{i : \xi_i > x\}\right| \le (1+\varepsilon) n\, x^{-\gamma} L\left(x\right)$$

$$\sum_{i : \xi_i \le x} \xi_i^2 \le (1+\varepsilon) \frac{4-\gamma}{2-\gamma} n\, x^{2-\gamma} L\left(x\right)$$

Upper bound: proof

Fix $x = n^{\frac{1}{\gamma+1}}$. For any $\delta > 0$ with probability $1 - O(n^{-\alpha})$:

$$T(n) \leq n^{\frac{3}{\gamma+1}+\delta}$$

Upper bound: proof

Fix $x = n^{\frac{1}{\gamma+1}}$. For any $\delta > 0$ with probability $1 - O(n^{-\alpha})$:

$$T(n) \leq n^{\frac{3}{\gamma+1}+\delta}$$

$$P_2(n) > n^{\frac{2}{\gamma}-\delta}$$

Upper bound: proof

Fix $x = n^{\frac{1}{\gamma+1}}$. For any $\delta > 0$ with probability $1 - O(n^{-\alpha})$:

$$T(n) \leq n^{\frac{3}{\gamma+1}+\delta}$$

$$P_2(n) > n^{\frac{2}{\gamma}-\delta}$$

Taking small enough δ, we obtain

$$C_1(G_n) \leq n^{\varepsilon-\frac{2-\gamma}{\gamma(\gamma+1)}}$$

Karamata's theorem

Theorem [Karamata]

Let L be slowly varying and locally bounded in $[x_0, \infty]$ for some $x_0 \geq 0$. Then

1. for $\alpha > -1$

$$\int_{x_0}^{x} t^{\alpha} L(t) dt = (1 + o(1))(\alpha + 1)^{-1} x^{\alpha+1} L(x), \quad x \to \infty\,.$$

2. for $\alpha < -1$

$$\int_{x}^{\infty} t^{\alpha} L(t) dt = -(1+o(1))(\alpha+1)^{-1} x^{\alpha+1} L(x), \quad x \to \infty\,.$$

Global clustering coefficient: lower bound

Theorem

For any $\varepsilon > 0$ and any α, $0 < \alpha < \min\{\frac{\gamma\varepsilon}{\gamma+2}, \frac{1}{\gamma+1}, \gamma - 1\}$, with probability $1 - O(n^{-\alpha})$ there exists a graph with the required degree distribution and

$$C_1(G_n) \geq n^{-\frac{(2-\gamma)}{\gamma(\gamma+1)} - \varepsilon}$$

Idea of the proof

- $\mathsf{P}\left(P_2(n) \leq n^{\frac{2}{\gamma}+\delta}\right) = 1 - O\left(n^{-\alpha}\right)$

Idea of the proof

- $\mathsf{P}\left(P_2(n) \leq n^{\frac{2}{\gamma}+\delta}\right) = 1 - O\left(n^{-\alpha}\right)$
- Let us order the degrees: $d_1 \geq \ldots \geq d_n$
- $d_k \approx \left(\frac{n}{k}\right)^{\frac{1}{\gamma}}$, so $d_k \approx k$ for $k \approx n^{\frac{1}{\gamma+1}}$

Idea of the proof

- $\mathsf{P}\left(P_2(n) \leq n^{\frac{2}{\gamma}+\delta}\right) = 1 - O\left(n^{-\alpha}\right)$
- Let us order the degrees: $d_1 \geq \ldots \geq d_n$
- $d_k \approx \left(\frac{n}{k}\right)^{\frac{1}{\gamma}}$, so $d_k \approx k$ for $k \approx n^{\frac{1}{\gamma+1}}$
- With probability $1 - O\left(n^{-\alpha}\right)$ we can construct a clique of size $n^{\frac{1}{\gamma+1}-\delta}$

Idea of the proof

- $\mathsf{P}\left(P_2(n) \le n^{\frac{2}{\gamma}+\delta}\right) = 1 - O\left(n^{-\alpha}\right)$
- Let us order the degrees: $d_1 \ge \ldots \ge d_n$
- $d_k \approx \left(\frac{n}{k}\right)^{\frac{1}{\gamma}}$, so $d_k \approx k$ for $k \approx n^{\frac{1}{\gamma+1}}$
- With probability $1 - O\left(n^{-\alpha}\right)$ we can construct a clique of size $n^{\frac{1}{\gamma+1}-\delta}$
- $3 \cdot T(n) \ge n^{\frac{3}{\gamma+1}-3\delta}$

Idea of the proof

- $\mathsf{P}\left(P_2(n) \leq n^{\frac{2}{\gamma}+\delta}\right) = 1 - O\left(n^{-\alpha}\right)$
- Let us order the degrees: $d_1 \geq \ldots \geq d_n$
- $d_k \approx \left(\frac{n}{k}\right)^{\frac{1}{\gamma}}$, so $d_k \approx k$ for $k \approx n^{\frac{1}{\gamma+1}}$
- With probability $1 - O\left(n^{-\alpha}\right)$ we can construct a clique of size $n^{\frac{1}{\gamma+1}-\delta}$
- $3 \cdot T(n) \geq n^{\frac{3}{\gamma+1}-3\delta}$
- Finally, $C_1(G_n) = \frac{3 \cdot T(n)}{P_2(n)} \geq \frac{n^{\frac{3}{\gamma+1}-3\delta}}{n^{\frac{2}{\gamma}+\delta}} = n^{-\frac{(2-\gamma)}{\gamma(\gamma+1)}-\varepsilon}$

Idea of the proof

- $\mathsf{P}\left(P_2(n) \leq n^{\frac{2}{\gamma}+\delta}\right) = 1 - O\left(n^{-\alpha}\right)$
- Let us order the degrees: $d_1 \geq \ldots \geq d_n$
- $d_k \approx \left(\frac{n}{k}\right)^{\frac{1}{\gamma}}$, so $d_k \approx k$ for $k \approx n^{\frac{1}{\gamma+1}}$
- With probability $1 - O\left(n^{-\alpha}\right)$ we can construct a clique of size $n^{\frac{1}{\gamma+1}-\delta}$
- $3 \cdot T(n) \geq n^{\frac{3}{\gamma+1}-3\delta}$
- Finally, $C_1(G_n) = \frac{3 \cdot T(n)}{P_2(n)} \geq \frac{n^{\frac{3}{\gamma+1}-3\delta}}{n^{\frac{2}{\gamma}+\delta}} = n^{-\frac{(2-\gamma)}{\gamma(\gamma+1)}-\varepsilon}$

Note: we have to prove that after we constructed a clique, **whp** we still can construct a graph without loops and multiple edges

Global clustering in weighted graphs

Global clustering coefficient for weighted graphs[2]:

$$C_1(G) = \frac{\text{total value of closed triplets}}{\text{total value of triplets}}$$

[2] T. Opsahl, P. Panzarasa, Clustering in weighted networks, Social Networks, 31(2), pp. 155–163 (2009)

Global clustering in weighted graphs

Global clustering coefficient for weighted graphs[2]:

$$C_1(G) = \frac{\text{total value of closed triplets}}{\text{total value of triplets}}$$

Value of a triplet:

- *arithmetic mean* of the weights of the ties
- *geometric mean*
- *maximum or minimum value*

[2] T. Opsahl, P. Panzarasa, Clustering in weighted networks, Social Networks, 31(2), pp. 155–163 (2009)

Global clustering in weighted graphs

Global clustering coefficient for weighted graphs[2]:

$$C_1(G) = \frac{\text{total value of closed triplets}}{\text{total value of triplets}}$$

Value of a triplet:

- *arithmetic mean* of the weights of the ties
- *geometric mean*
- *maximum or minimum value*
- *product*

[2] T. Opsahl, P. Panzarasa, Clustering in weighted networks, Social Networks, 31(2), pp. 155–163 (2009)

Global clustering in weighted graphs

Theorem

Fix any $\delta > 0$. For any α, $0 < \alpha < \frac{\gamma-1}{\gamma+1}$, with probability $1 - O\left(n^{-\alpha}\right)$ there exists a multigraph with the required degree distribution and

$$C_1(G_n) \geq \frac{2-\gamma}{10-3\gamma} - \delta$$

This theorem holds for any definition of global clustering coefficient

Idea of the proof

- **Whp** we can construct a clique on the set A of $n^{\frac{1}{\gamma+1}}$ vertices with largest degrees
- In addition, we connect all vertices from the set A only to each other

Idea of the proof

- **Whp** we can construct a clique on the set A of $n^{\frac{1}{\gamma+1}}$ vertices with largest degrees
- In addition, we connect all vertices from the set A only to each other
- The total value of closed triplets is at least $n^{\frac{3}{\gamma+1}}$

Idea of the proof

- **Whp** we can construct a clique on the set A of $n^{\frac{1}{\gamma+1}}$ vertices with largest degrees
- In addition, we connect all vertices from the set A only to each other
- The total value of closed triplets is at least $n^{\frac{3}{\gamma+1}}$
- The total value of all remaining triplets is at most $n^{\frac{3}{\gamma+1}}$

Concluding remarks

- In unweighted scale-free graphs with $1 < \gamma < 2$ the global clustering coefficient tends to zero
- We proposed a constructing procedure which allows to reach the obtained upper bound
- It is possible to construct a sequence of weighted scale-free graphs with an asymptotically constant global clustering coefficient

Thank you!

Questions?