Generalized preferential attachment

Liudmila Ostroumova
Yandex
Lomonosov Moscow State University

Joint work with A. Ryabchenko and E. Samosvat

November, 2012
Experimental observations

Examples of large real-world networks:

- World-wide web
- Social networks
- Biological and chemical systems
- Neural networks

Typical properties:

- Sparse graphs (\( n \) vertices, \( mn \) edges)
- Small diameter
- Power law degree distribution

\[
|\{v : \text{deg}(v) = d\}| \approx c d^{\gamma}, \quad 2 < \gamma < 3
\]

Constant clustering coefficient
Experimental observations

Examples of large real-world networks:

- World-wide web
- Social networks
- Biological and chemical systems
- Neural networks

Typical properties:

- Sparse graphs \((n \text{ vertices}, \; m n \text{ edges})\)
- Small diameter
- Power law degree distribution

\[
\frac{\left| \{v : \deg(v) = d\} \right|}{n} \approx \frac{c}{d^\gamma}, \quad 2 < \gamma < 3
\]

- Constant clustering coefficient
Global clustering coefficient of a graph $G$:

$$C_1(n) = \frac{3 \# \text{(triangles in } G \text{)}}{\# \text{(pairs of adjacent edges in } G \text{)}}.$$
Global clustering coefficient of a graph $G$:

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

Average local clustering coefficient

- $T^i$ is the number of edges between the neighbors of a vertex $i$
- $P^i_2$ is the number of pairs of neighbors
- $C(i) = \frac{T^i}{P^i_2}$ is the local clustering coefficient for a vertex $i$
- $C_2(n) = \frac{1}{n} \sum_{i=1}^{n} C(i)$ – average local clustering coefficient
Idea of preferential attachment [Barabási, Albert]:

- Start with a small graph
- At every step we add new vertex with \( m \) edges
- The probability that a new vertex will be connected to a vertex \( i \) is proportional to the degree of \( i \)
Idea of preferential attachment [Barabási, Albert]:

- Start with a small graph
- At every step we add new vertex with \( m \) edges
- The probability that a new vertex will be connected to a vertex \( i \) is proportional to the degree of \( i \)

Theorem [Bollobás, Riordan]

Let \( f(n), n \geq 2, \) be any integer-valued function with \( f(2) = 0 \) and \( f(n) \leq f(n + 1) \leq f(n) + 1 \) for every \( n \geq 2 \), such that \( f(m) \to \infty \) as \( n \to \infty \). Then there is a random graph process \( T(n) \) satisfying the conditions of Barabási and Albert such that, with probability 1, \( T(n) \) has exactly \( f(n) \) triangles for all sufficiently large \( n \).
Start from an arbitrary graph $G_{m_0}^{n_0}$ with $n_0$ vertices and $mn_0$ edges.
Start from an arbitrary graph $G_{m}^{n_0}$ with $n_0$ vertices and $mn_0$ edges. We make $G_{m}^{n+1}$ from $G_{m}^{n}$ by adding a new vertex $n + 1$ with $m$ edges.
Start from an arbitrary graph $G_{m_0}^{n_0}$ with $n_0$ vertices and $mn_0$ edges.
We make $G_{m_1}^{n+1}$ from $G_{m_0}^{n}$ by adding a new vertex $n + 1$ with $m$ edges.
The probability that the degree of a vertex $i$ increases by one equals

$$A \frac{\deg(i)}{n} + B \frac{1}{n} + O \left( \frac{(\deg(i))^2}{n^2} \right)$$
- Start from an arbitrary graph $G_{m}^{n_0}$ with $n_0$ vertices and $mn_0$ edges.
- We make $G_{m}^{n+1}$ from $G_{m}^{n}$ by adding a new vertex $n + 1$ with $m$ edges.
- The probability that the degree of a vertex $i$ increases by one equals
  \[ A \frac{\text{deg}(i)}{n} + B \frac{1}{n} + O \left( \frac{(\text{deg}(i))^2}{n^2} \right) \]
- The probability of adding a multiple edge is $O \left( \frac{(\text{deg}(i))^2}{n^2} \right)$.
Start from an arbitrary graph $G_{mn_0}^{n_0}$ with $n_0$ vertices and $mn_0$ edges. We make $G_{m+1}^{n+1}$ from $G_m^n$ by adding a new vertex $n + 1$ with $m$ edges. The probability that the degree of a vertex $i$ increases by one equals

\[ A \frac{\deg(i)}{n} + B \frac{1}{n} + O \left( \frac{(\deg(i))^2}{n^2} \right) \]

The probability of adding a multiple edge is $O \left( \frac{(\deg(i))^2}{n^2} \right)$. 

$2mA + B = m$, $0 \leq A \leq 1$
Triangles property:
The probability that the degree of two vertices $i$ and $j$ increases by one equals

$$e_{ij} \frac{D}{mn} + O\left(\frac{d_i^m d_j^m}{n^2}\right)$$

Here $e_{ij}$ is the number of edges between vertices $i$ and $j$ in $G^n_m$ and $D$ is a positive constant.
Fix some positive number $a$ – "initial attractiveness". (Bollobás–Riordan model: $a = 1$).

Start with a graph with one vertex and $m$ loops.

At $n$th each step add one vertex with $m$ edges.

We add $m$ edges one by one. The probability to add an edge $ni$ at each step is proportional to $\text{deg}(i) + a$. 

\[ \text{Outdegree: } m \]
\[ \text{Triangles: } D = 0 \]
\[ \text{Preferential attachment: } A = \frac{1}{1 + a} \]
\[ \text{Degree distribution: Power law with } \gamma = 2 + a \]
\[ \text{Global clustering: } \left(\log n\right)^{\frac{2}{n}} \text{ for } a = 1, \log n n \text{ for } a > 1 \]
Fix some positive number $a$ — "initial attractiveness". (Bollobás–Riordan model: $a = 1$).

Start with a graph with one vertex and $m$ loops.

At $n$th each step add one vertex with $m$ edges.

We add $m$ edges one by one. The probability to add an edge $ni$ at each step is proportional to $\text{deg}(i) + a$.

- **Outdegree**: $m$
- **Triangles**: $D = 0$
- **Preferential attachment**: $A = \frac{1}{1+a}$
- **Degree distribution**: Power law with $\gamma = 2 + a$
- **Global clustering**: $\frac{(\log n)^2}{n}$ ($a = 1$), $\frac{\log n}{n}$ ($a > 1$)
Holme–Kim model

**Idea:** To mix preferential attachment steps with the steps of triangle formation.
**Holme–Kim model**

**Idea:** To mix preferential attachment steps with the steps of triangle formation.

- Add a new vertex $v$ with $m$ edges
- Perform one PA step
- Then perform a triangle formation step with the probability $P_t$ or a PA step with the probability $1 - P_t$

**Triangle formation:** If an edge between $v$ and $u$ was added in the previous PA step, then add one more edge from $v$ to a randomly chosen neighbor of $u$. 
**Idea:** To mix preferential attachment steps with the steps of triangle formation.

- Add a new vertex $v$ with $m$ edges
- Perform one PA step
- Then perform a triangle formation step with the probability $P_t$ or a PA step with the probability $1 - P_t$

**Triangle formation:** If an edge between $v$ and $u$ was added in the previous PA step, then add one more edge from $v$ to a randomly chosen neighbor of $u$.

- **Outdegree:** $m$
- **Triangles:** $D = (m - 1)P_t$
- **Preferential attachment:** $A = \frac{1}{2}$
- **Degree distribution:** Power law with $\gamma = 3$
- **Average local clustering:** constant
- **Global clustering:** tends to zero
Random Apollonian networks

Outdegree: $m = 3$

Triangles: $D = 3$

Preferential attachment: $A = \frac{1}{2}$

Degree distribution: Power law with $\gamma = 3$

Average local clustering: constant

Global clustering: tends to zero

Liudmila Ostroumova

Generalized preferential attachment
Random Apollonian networks

Outdegree: $m = 3$

Triangles: $D = 3$

Preferential attachment: $A = 2$

Degree distribution: Power law with $\gamma = 3$

Average local clustering: constant

Global clustering: tends to zero
Random Apollonian networks

Outdegree: $m = 3$

Triangles: $D = 3$

Preferential attachment: $A = \frac{1}{2}$

Degree distribution: Power law with $\gamma = 3$

Average local clustering: constant

Global clustering: tends to zero

Liudmila Ostroumova

Generalized preferential attachment
Random Apollonian networks

Outdegree: \( m = 3 \)

Triangles: \( D = 3 \)

Preferential attachment: \( A = \frac{1}{2} \)

Degree distribution: Power law with \( \gamma = 3 \)

Average local clustering: constant

Global clustering: tends to zero

Liudmila Ostroumova

Generalized preferential attachment
Random Apollonian networks

Outdegree: \( m = 3 \)

Triangles: \( D = 3 \)

Preferential attachment: \( A = \frac{1}{2} \)

Degree distribution: Power law with \( \gamma = 3 \)

Average local clustering: constant

Global clustering: tends to zero
Random Apollonian networks

Outdegree: \( m = 3 \)

Triangles: \( D = 3 \)

Preferential attachment: \( A = 1 \)

Degree distribution: Power law with \( \gamma = 3 \)

Average local clustering: constant

Global clustering: tends to zero

Liudmila Ostroumova

Generalized preferential attachment
Random Apollonian networks

Outdegree: $m = 3$

Triangles: $D = 3$

Preferential attachment: $A = 1$

Degree distribution: Power law with $\gamma = 3$

Average local clustering: constant

Global clustering: tends to zero

Liudmila Ostroumova

Generalized preferential attachment
Random Apollonian networks

- **Outdegree:** $m = 3$
- **Triangles:** $D = 3$
- **Preferential attachment:** $A = \frac{1}{2}$
- **Degree distribution:** Power law with $\gamma = 3$
- **Average local clustering:** constant
- **Global clustering:** tends to zero
Polynomial model

- Put $m = 2p$
Polynomial model

- Put $m = 2p$
- Fix $\alpha, \beta, \delta \geq 0$ and $\alpha + \beta + \delta = 1$
Polynomial model

- Put \( m = 2p \)
- Fix \( \alpha, \beta, \delta \geq 0 \) and \( \alpha + \beta + \delta = 1 \)
- Add a new vertex \( i \) with \( m \) edges. We add \( m \) edges in \( p \) steps
Polynomial model

- Put \( m = 2p \)
- Fix \( \alpha, \beta, \delta \geq 0 \) and \( \alpha + \beta + \delta = 1 \)
- Add a new vertex \( i \) with \( m \) edges. We add \( m \) edges in \( p \) steps
- \( \alpha \) – probability of an indegree preferential step
- \( \beta \) – probability of an edge preferential step
- \( \delta \) – probability of a random step
Polynomial model

- Put $m = 2p$
- Fix $\alpha, \beta, \delta \geq 0$ and $\alpha + \beta + \delta = 1$
- Add a new vertex $i$ with $m$ edges. We add $m$ edges in $p$ steps
- $\alpha$ – probability of an indegree preferential step
- $\beta$ – probability of an edge preferential step
- $\delta$ – probability of a random step

**Edge preferential:** choose a random edge, add two edges between its endpoints and $i$
Put $m = 2p$

Fix $\alpha, \beta, \delta \geq 0$ and $\alpha + \beta + \delta = 1$

Add a new vertex $i$ with $m$ edges. We add $m$ edges in $p$ steps

- $\alpha$ – probability of an indegree preferential step
- $\beta$ – probability of an edge preferential step
- $\delta$ – probability of a random step

### Edge preferential: choose a random edge, add two edges between its endpoints and $i$

- **Outdegree:** $2p$
- **Triangles:** $D = \beta p$
- **Preferential attachment:** $A = \alpha + \frac{\beta}{2}$.
- **Degree distribution:** Power law with $\gamma = 1 + \frac{2}{2\alpha + \beta}$
- **Average local clustering:** constant
- **Global clustering:** constant for $A > 1/2$ ($\gamma > 3$), tends to zero for $A \leq 1/2$ ($2 < \gamma \leq 3$)
Global clustering

$\gamma = 3.5$

![Graph showing global clustering with parameters $\alpha = 0.4, \beta = 0$ and $\alpha = 0, \beta = 0.8$.](image)
Average local clustering

$\gamma = 3.5$

$\alpha = 0.4, \beta = 0$

$\alpha = 0, \beta = 0.8$

Liudmila Ostroumova

Generalized preferential attachment
Average local and global clustering depending on $A$

$\beta = 0.5, \gamma = 1 + 1/A$
Global and average local clustering depending on $n$

$\alpha = 0.5, \beta = 0.2 \Rightarrow \gamma = 8/3$
Degree distribution

Let $N_n(d)$ be the number of vertices with degree $d$ in $G^m_n$. Then for $k < n^{2+1/A}$ whp

$$N_n(d) \sim \frac{\Gamma \left( m + \frac{B+1}{A} \right)}{A \Gamma \left( m + \frac{B}{A} \right)} d^{-1 - \frac{1}{A}} n.$$
Degree distribution

Let $N_n(d)$ be the number of vertices with degree $d$ in $G_{m}^{n}$. Then for $k < n^{\frac{1}{2+1/A}}$ \textbf{whp}

$$N_n(d) \sim \frac{\Gamma \left( m + \frac{B+1}{A} \right)}{A \Gamma \left( m + \frac{B}{A} \right)} d^{-1-\frac{1}{A} n}.$$

Average local clustering

\textbf{Whp}

$$C_2(n) \geq \frac{1}{n} \sum_{i: \text{deg}(i) = m} C(i) \geq \frac{2cD}{m(m + 1)}.$$
Let $P_2(n)$ be the number of all path of length 2 in $G_m^n$.

<table>
<thead>
<tr>
<th>$P_2(n)$</th>
</tr>
</thead>
</table>
| (1) If $2A < 1$, then \textbf{whp} $P_2(n) \sim \left(2m(A + B) + \frac{m(m-1)}{2}\right) \frac{n}{1-2A}$.
| (2) If $2A = 1$, then \textbf{whp} $P_2(n) \propto n \log(n)$.
| (3) If $2A > 1$, then \textbf{whp} $P_2(n) \propto n^{2A}$.
Let $P_2(n)$ be the number of all paths of length 2 in $G^m_n$.

### $P_2(n)$

1. If $2A < 1$, then $\text{whp } P_2(n) \sim \left(2m(A + B) + \frac{m(m-1)}{2}\right) \frac{n}{1-2A}$.
2. If $2A = 1$, then $\text{whp } P_2(n) \propto n \log(n)$.
3. If $2A > 1$, then $\text{whp } P_2(n) \propto n^{2A}$.

### Triangles

$\text{Whp}$ the number of triangles $T(n) \sim Dn$. 

Liudmila Ostroumova | Generalized preferential attachment
Let $P_2(n)$ be the number of all path of length 2 in $G_m^n$.

### $P_2(n)$

1. If $2A < 1$, then \textbf{whp} $P_2(n) \sim \left(2m(A + B) + \frac{m(m-1)}{2}\right) \frac{n}{1-2A}$.
2. If $2A = 1$, then \textbf{whp} $P_2(n) \propto n \log(n)$.
3. If $2A > 1$, then \textbf{whp} $P_2(n) \propto n^{2A}$.

### Triangles

\textbf{Whp} the number of triangles $T(n) \sim Dn$.

### Global clustering

1. If $2A < 1$, then \textbf{whp} $C_1(n) \sim \frac{3(1-2A)D}{\left(2m(A+B)+\frac{m(m-1)}{2}\right)}$.
2. If $2A = 1$, then \textbf{whp} $C_1(n) \propto (\log n)^{-1}$.
3. If $2A > 1$, then \textbf{whp} $C_1(n) \propto n^{1-2A}$. 

Liudmila Ostroumova

Generalized preferential attachment
$P_2(n)$ and $T(n)$ in real networks

Retweet graph

- Number of P2
- $200 \cdot \text{(number of triangles)}$

Number of vertices

Slope: 2.3
Retweet graph

- Number of P2
- $200 \cdot \text{(number of triangles)}$

Slope: 2.3
**Generalized preferential attachment:**

- Power law degree distribution with any exponent $\gamma > 2$
- Constant average local clustering coefficient
- Constant global clustering coefficient for $\gamma > 3$
Generalized preferential attachment:

- Power law degree distribution with any exponent $\gamma > 2$
- Constant average local clustering coefficient
- Constant global clustering coefficient for $\gamma > 3$

Ways to overcome this obstacle:

- The number of added edges is a random variable (C. Cooper, 2006)
- A new vertex added at time $t$ generates $t^c$ edges (C. Cooper, P. Prałat, 2011)
- Adding edges between already existing nodes (e.g. the Cooper–Frieze model)