Preferential attachment models and their generalizations

Liudmila Ostroumova, Andrei Raigorodskii

Yandex
Lomonosov Moscow State University
Moscow Institute of Physics and Technology

June, 2013

Experimental observations

Examples of large real-world networks:

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

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

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

- Constant clustering coefficient (many triangles)

Preferential attachment: definition

Preferential attachment (Barabási and Albert, Bollobás and Riordan):

- Start with a small graph
- At every step we add a 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
- Usually m edges are drawn independently or one by one
- After n steps we obtain a graph G_m^n

Preferential attachment: properties

Theorem [Bollobás–Riordan–Spencer–Tusnády]

Let $m \geq 1$ and $\epsilon > 0$ be fixed, and put $\alpha_{m,d} = \frac{2m(m+1)}{d(d+1)(d+2)}$. Then **whp** we have

$$(1-\epsilon)\alpha_{m,d} \leq \frac{\#_m^n(d)}{n} \leq (1+\epsilon)\alpha_{m,d}$$

for every d in the range $m \leq d \leq n^{\frac{1}{15}}$

Preferential attachment: properties

Theorem [Bollobás–Riordan–Spencer–Tusnády]

Let $m \geq 1$ and $\epsilon > 0$ be fixed, and put $\alpha_{m,d} = \frac{2m(m+1)}{d(d+1)(d+2)}$. Then **whp** we have

$$(1-\epsilon)\alpha_{m,d} \leq \frac{\#_m^n(d)}{n} \leq (1+\epsilon)\alpha_{m,d}$$

for every d in the range $m \leq d \leq n^{\frac{1}{15}}$

In particular, **whp** for all d in this range we have $\frac{\#_{a,m}^n(d)}{n} = \Theta\left(d^{-3}\right)$.

Preferential attachment: properties

Theorem [Bollobás–Riordan–Spencer–Tusnády]

Let $m \geq 1$ and $\epsilon > 0$ be fixed, and put $\alpha_{m,d} = \frac{2m(m+1)}{d(d+1)(d+2)}$. Then **whp** we have

$$(1-\epsilon)\alpha_{m,d} \leq \frac{\#_m^n(d)}{n} \leq (1+\epsilon)\alpha_{m,d}$$

for every d in the range $m \leq d \leq n^{\frac{1}{15}}$

In particular, **whp** for all d in this range we have $\frac{\#_{a,m}^n(d)}{n} = \Theta\left(d^{-3}\right)$.

Theorem (Bollobás–Riordan)

Fix an integer $m \geq 2$ and a positive real number ϵ. Then **whp** G_m^n is connected and has diameter $\mathrm{diam}\left(G_m^n\right)$ satisfying

$$(1-\epsilon)\log n/\log\log n \leq \mathrm{diam}\left(G_m^n\right) \leq (1+\epsilon)\log n/\log\log n$$

Preferential attachment: generating

Algorithm 1: Preferential attachment

input : Number of vertices n, vertex out-degree $m \geq 1$
output: Graph $G = (\{1, \ldots, n\}, E)$

```
M: array of length 2mn
for v ⟵ 0 to n − 1 do
    for i ⟵ 1 to m do
        M[2(mv + i)] ⟵ v
        draw r ∈ {1, ..., 2(mv + i)} uniformly at random
        M[2(mv + i) − 1] ⟵ M[r]
E ⟵ ∅
for i ⟵ 1 to mn do
    E ⟵ E ∪ {M[2i], M[2i − 1]}
```

Buckley–Osthus and Móri models

- Fix some positive constant a – "initial attractiveness".
- Start with a graph with one vertex and m loops.
- At nth each step add one vertex and add m edges one by one.
- The probability to add an edge ni is proportional to $indeg(i) + ma$.

Buckley–Osthus and Móri models

- Fix some positive constant a – "initial attractiveness".
- Start with a graph with one vertex and m loops.
- At nth each step add one vertex and add m edges one by one.
- The probability to add an edge ni is proportional to $indeg(i) + ma$.

Theorem (Buckley–Osthus)

Let $m, a \geq 1$ be fixed integers then for all $0 \leq d \leq n^{1/100(a+1)}$ **whp**

$$\frac{\#^n_{a,m}(d)}{n} \sim C(a,m)d^{-2-a}.$$

Buckley–Osthus and Móri models

- Fix some positive constant a – "initial attractiveness".
- Start with a graph with one vertex and m loops.
- At nth each step add one vertex and add m edges one by one.
- The probability to add an edge ni is proportional to $indeg(i) + ma$.

Theorem (Buckley–Osthus)

Let $m, a \geq 1$ be fixed integers then for all $0 \leq d \leq n^{1/100(a+1)}$ **whp**

$$\frac{\#^n_{a,m}(d)}{n} \sim C(a,m)d^{-2-a}.$$

Theorem (Grechnikov)

Let $a > 0$ be fixed real and $\psi(n) \to \infty$ when $n \to \infty$, then **whp** we have

$$\left| \#^n_{a,m}(d) - \frac{\mathrm{B}(d + ma, a + 2)}{\mathrm{B}(ma, a + 1)} n \right| \leq \left(\sqrt{d^{-a-2} n} + d^{-1} \right) \psi(n).$$

Clustering coefficient

Global clustering coefficient of a graph G:

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

Clustering coefficient

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_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
- $C_2(n) = \frac{1}{n}\sum_{i=1}^{n} C(i)$ – average local clustering coefficient

Theorem (Bollobás)

Let $m \geq 1$ be fixed. The expected number of triangles in G_m^n is given by

$$(1 + o(1)) \frac{m(m-1)(m+1)}{48} (\log n)^3$$

as $n \to \infty$.

Triangles in PA models

Theorem (Bollobás)

Let $m \geq 1$ be fixed. The expected number of triangles in G_m^n is given by

$$(1 + o(1)) \frac{m(m-1)(m+1)}{48} (\log n)^3$$

as $n \to \infty$.

Theorem (Bollobás)

Let $m \geq 1$ be fixed. The expected value of the global clustering coefficient is

$$\mathrm{E} C_1(G_m^n) \sim \frac{m-1}{8} \frac{(\log n)^2}{n}$$

as $n \to \infty$.

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.

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.

Problems:

- Parameter of the power-law degree distribution is $\gamma = 3$
- Global clustering coefficient tends to zero.

Polynomial model: definition

Polynomial model: definition

- Put $m = 2p$

Polynomial model: definition

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

Polynomial model: definition

- 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: definition

- 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
- α – probability of an indegree preferential step
  β – probability of an edge preferential step
  δ – probability of a random step

Polynomial model: definition

- 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
- α – probability of an indegree preferential step
  β – probability of an edge preferential step
  δ – probability of a random step

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

Polynomial model: properties

Degree distribution

Let $N_n(d)$ be the number of vertices with degree d. Then for $d < n^{\frac{2\alpha+\beta}{2(2\alpha+\beta+1)}}$ **whp**

$$N_n(d) \sim C(m, \alpha, \beta) d^{-1-\frac{2}{2\alpha+\beta}} n.$$

Polynomial model: properties

Degree distribution

Let $N_n(d)$ be the number of vertices with degree d. Then for $d < n^{\frac{2\alpha+\beta}{2(2\alpha+\beta+1)}}$ **whp**

$$N_n(d) \sim C(m,\alpha,\beta) d^{-1-\frac{2}{2\alpha+\beta}} n.$$

Average local clustering

If $\beta > 0$, then **whp**

$$C_2(n) \geq C(m,\beta) > 0\,.$$

Polynomial model: properties

Degree distribution

Let $N_n(d)$ be the number of vertices with degree d. Then for $d < n^{\frac{2\alpha+\beta}{2(2\alpha+\beta+1)}}$ **whp**

$$N_n(d) \sim C(m, \alpha, \beta) d^{-1-\frac{2}{2\alpha+\beta}} n.$$

Average local clustering

If $\beta > 0$, then **whp**

$$C_2(n) \geq C(m, \beta) > 0\,.$$

Global clustering

(1) If $2\alpha + \beta < 1$ then **whp** $C_1(n) \sim const$.

(2) If $2\alpha + \beta = 1$ then **whp** $C_1(n) \propto (\log n)^{-1}$.

(2) If $2\alpha + \beta > 1$ then **whp** $C_1(n) \propto n^{1-2\alpha-\beta}$.

Polynomial model: generating

Algorithm 2: Polynomial model

input : Number of vertices n, out-degree $m = 2p$, $\alpha, \beta, \delta \geq 0$: $\alpha + \beta + \delta = 1$
output: Graph $G = (\{1, \ldots, n\}, E)$

for $v \longleftarrow 0$ **to** $n-1$ **do**
  for $i \longleftarrow 1$ **to** p **do**
    $M[2(mv + 2i - 1)] \longleftarrow v$; $M[2(mv + 2i)] \longleftarrow v$
    switch *the value of* $r \xleftarrow{\text{sample}} U[0,1]$ **do**
      case $r < \alpha$
        draw $r_1, r_2 \in \{1, \ldots, mv + 2i\}$ uniformly at random
        $M[2(mv + 2i) - 1] \longleftarrow M[2r_1 - 1]$;
        $M[2(mv + 2i) + 1] \longleftarrow M[2r_2 - 1]$
      case $r < \alpha + \beta$
        draw $r_1 \in \{1, \ldots, mv + 2i\}$ uniformly at random
        $M[2(mv + 2i) - 1] \longleftarrow M[2r_1]$;
        $M[2(mv + 2i) + 1] \longleftarrow M[2r_1 + 1]$
      otherwise
        draw $r_1, r_2 \in \{1, \ldots, v\}$ uniformly at random
        $M[2(mv + 2i) - 1] \longleftarrow r_1$; $M[2(mv + 2i) + 1] \longleftarrow r_2$

for $i \longleftarrow 1$ **to** mn **do**
  $E \longleftarrow E \cup \{M[2i], M[2i + 1]\}$

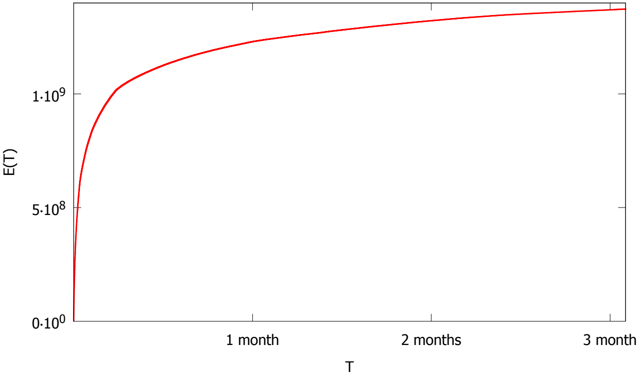
Recency property

Let E be the number of edges in a graph. Some edges connect nodes with age difference less than T. Denote the number of such edges by $E(T)$.

Recency property

Let E be the number of edges in a graph. Some edges connect nodes with age difference less than T. Denote the number of such edges by $E(T)$.

We noticed that news-related part of the Web has so-called recency property.



Recency sensitive models

At each step a new vertex with m edges is added. Neighbors of new vertex are chosen with probabilities proportional to their attractiveness.

Recency sensitive models

At each step a new vertex with m edges is added. Neighbors of new vertex are chosen with probabilities proportional to their attractiveness. Attractiveness function:

$$f_\tau(d,q,a) = (1 \text{ or } q) \cdot (1 \text{ or } d) \cdot \left(1 \text{ or } e^{-\frac{a}{\tau}}\right) \ ,$$

where q is quality of a vertex, d is degree of a vertex, and a is age of a vertex.

Recency sensitive models

At each step a new vertex with m edges is added. Neighbors of new vertex are chosen with probabilities proportional to their attractiveness. Attractiveness function:

$$f_\tau(d, q, a) = (1 \text{ or } q) \cdot (1 \text{ or } d) \cdot \left(1 \text{ or } e^{-\frac{a}{\tau}}\right) ,$$

where q is quality of a vertex, d is degree of a vertex, and a is age of a vertex.

- $f_\tau(d, q, a) = d$ leads to preferential attachment
- $f_\tau(d, q, a) = q \cdot d$ leads to fitness model

Recency sensitive models: properties

Degree distribution:

- If $f_\tau = d$ (preferential attachment), then $\#(d) \propto d^{-3}$

Recency sensitive models: properties

Degree distribution:

- If $f_\tau = d$ (preferential attachment), then $\#(d) \propto d^{-3}$
- If $f_\tau = q\, e^{-\frac{a}{\tau}}$ and q is distributed according to the power law with parameter γ, then $\#(d) \sim d^{-\gamma}$

Degree distribution:

- If $f_\tau = d$ (preferential attachment), then $\#(d) \propto d^{-3}$
- If $f_\tau = q\, e^{-\frac{a}{\tau}}$ and q is distributed according to the power law with parameter γ, then $\#(d) \sim d^{-\gamma}$
- If $f_\tau = d\, q\, e^{-\frac{a}{\tau}}$ and q is exponentially distributed with parameter μ, then $\#(d) \sim d^{-C(\tau)}$

Recency sensitive models: properties

Degree distribution:

- If $f_\tau = d$ (preferential attachment), then $\#(d) \propto d^{-3}$
- If $f_\tau = q\, e^{-\frac{a}{\tau}}$ and q is distributed according to the power law with parameter γ, then $\#(d) \sim d^{-\gamma}$
- If $f_\tau = d\, q\, e^{-\frac{a}{\tau}}$ and q is exponentially distributed with parameter μ, then $\#(d) \sim d^{-C(\tau)}$

Recency property

$E - E(T) \propto e^{\frac{-T}{\tau}}$

Function Sample(V, W)

input : array V[*1*. . . *n*], array W[*1*. . . *n+1*]

`// assume that` W[*1*] $= 0$ `and` W[*i+1*] `is the sum of weights of` V[*1*]$, \ldots,$V[*i*]

output: V[*i*] with probability proportional to W[*i+1*] - W[*i*]

$\xi \xleftarrow{\text{sample}} U[0,1]$;

$x \longleftarrow \xi \cdot$ W[*n+1*];

`// find` V[*r-1*] `where` $r = \arg\min\{k : \mathsf{W}[k] > x\}$

$l \longleftarrow 1, r \longleftarrow n+1$;

while *l* < *r* **do**

  mid $= \lfloor \frac{l+r}{2} \rfloor$;

  if W[*mid*] > *x* **then**

    $r =$ mid ;

  else

    $l =$ mid $+ 1$;

return V[*r*]

Recency sensitive models: generating

Algorithm 3: Recency sensitive model

input : Number of vertices n, out-degree m, quality distribution $\mathfrak{Q}$,
attractiveness function $f(d, q, a)$

output: Graph $G = (\{1, \ldots, n\}, E)$

W[*1*] $\longleftarrow 0, i \longleftarrow 1$;
for *new* $\longleftarrow 1$ **to** n **do**
    d[*new*] $\longleftarrow m$; q[*new*] $\xleftarrow{\text{sample}} \mathfrak{Q}$;
    for $k \longleftarrow 1$ **to** m **do**
        old $\longleftarrow$ `Sample(`*V*, *W*`)` ;
        W[*i+1*] $\longleftarrow$ W[*i*] $+ f($d[*old*]+1, q[*old*], -old$) - f($d[*old*], q[*old*], -old$)$;
        V[*i*] $\longleftarrow$ old ;
        $E \longleftarrow E \cup \{new, old\}$;
        d[*old*] $\longleftarrow$ d[*old*] $+ 1$;
        $i \longleftarrow i + 1$;
    W[*i+1*] $\longleftarrow$ W[*i*] $+ f($d[*new*], q[*new*], -new$)$;
    V[*i*] $\longleftarrow$ new ;
    $i \longleftarrow i + 1$;