Mathematical Models of the WWW and related networks

Alan Frieze

The WWW is an example of a large real-world network.

The WWW is an example of a large real-world network.
It grows unpredictably (at least not by formal design)

The WWW is an example of a large real-world network.
It grows unpredictably (at least not by formal design)

Other examples:

The WWW is an example of a large real-world network.
It grows unpredictably (at least not by formal design)

Other examples:

- Internet

The WWW is an example of a large real-world network. It grows unpredictably (at least not by formal design)

Other examples:

- Internet
- Metabolic Networks

The WWW is an example of a large real-world network.
It grows unpredictably (at least not by formal design)

Other examples:

- Internet
- Metabolic Networks
- Social networks

The WWW is an example of a large real-world network.
It grows unpredictably (at least not by formal design)

Other examples:

- Internet
- Metabolic Networks
- Social networks
- Neural Networks

The WWW is an example of a large real-world network. It grows unpredictably (at least not by formal design)

Other examples:

- Internet
- Metabolic Networks
- Social networks
- Neural Networks
- Peer to Peer Networks

We assume that they arise via some random process.

We assume that they arise via some random process.

We model them as random graphs

We assume that they arise via some random process.

We model them as random graphs

Classical Erdős and Rényi Model: $G_{n,m}$

We assume that they arise via some random process.

We model them as random graphs

Classical Erdős and Rényi Model: $G_{n,m}$

Vertex set $\{1, 2, \ldots, n\}$, $N = \binom{n}{2}$
Each of the $\binom{N}{m}$ graphs with m edges is equally likely.

Problem

Problem

Suppose that $m = cn$, c constant.
For small k, the number n_k of vertices of degree k satisfies

$$n_k \sim \frac{c^k e^{-c}}{k!} n \qquad \mathbf{whp}$$

Problem

Suppose that $m = cn$, c constant.
For small k, the number n_k of vertices of degree k satisfies

$$n_k \sim \frac{c^k e^{-c}}{k!} n \qquad \textbf{whp}$$

In many real world cases

$$n_k \sim \frac{A}{k^\alpha} n \qquad \text{power law}$$

for some constants A, α e.g. Faloutsos,Faloutsos,Faloutsos

Fixed (expected) degree sequence models.

Fixed (expected) degree sequence models.

Choose a degree sequence $d_1, d_2, \ldots, d_n$ and then choose a graph G uniformly from graphs with this degree sequence.

Fixed (expected) degree sequence models.

Choose a degree sequence $d_1, d_2, \ldots, d_n$ and then choose a graph G uniformly from graphs with this degree sequence.

Bender and Canfield
McKay,Wormald
Bollobás
Molloy and Reed

Fixed (expected) degree sequence models.

Choose a degree sequence $d_1, d_2, \ldots, d_n$ and then choose a graph G uniformly from graphs with this degree sequence.

Bender and Canfield
McKay,Wormald
Bollobás
Molloy and Reed
Let $\Theta = \sum_i d_i(d_i - 2)$.

$\Theta < 0$ implies all components of G are small **whp**

$\Theta > 0$ implies G contains a giant component (size $\Omega(n)$) **whp**

Digraphs with a fixed degree sequence: Cooper and Frieze

Consider a random digraph D with n vertices and $\ell_{i,j}$ vertices of in-degree i and out-degree j.

Digraphs with a fixed degree sequence: Cooper and Frieze

Consider a random digraph D with n vertices and $\ell_{i,j}$ vertices of in-degree i and out-degree j.

$\theta n = \sum_{i,j} i\ell_{i,j}$ is the number of arcs in D.
$d = \sum_{i,j} ij\ell_{i,j}/(\theta n)$ is the expected out-degree of a randomly chosen arc.

Digraphs with a fixed degree sequence: Cooper and Frieze

Consider a random digraph D with n vertices and $\ell_{i,j}$ vertices of in-degree i and out-degree j.

$\theta n = \sum_{i,j} i\ell_{i,j}$ is the number of arcs in D.
$d = \sum_{i,j} ij\ell_{i,j}/(\theta n)$ is the expected out-degree of a randomly chosen arc. Then **whp**

$\theta < 0$ implies all strong components of G are small

$\theta > 0$ implies G has a giant strong component S (size $\Omega(n)$)

More on $d > 1$.

Let L^+ be the set of vertices with a giant “fan-out” and L^- be the set of vertices with a giant “fan-in”. Then **whp**

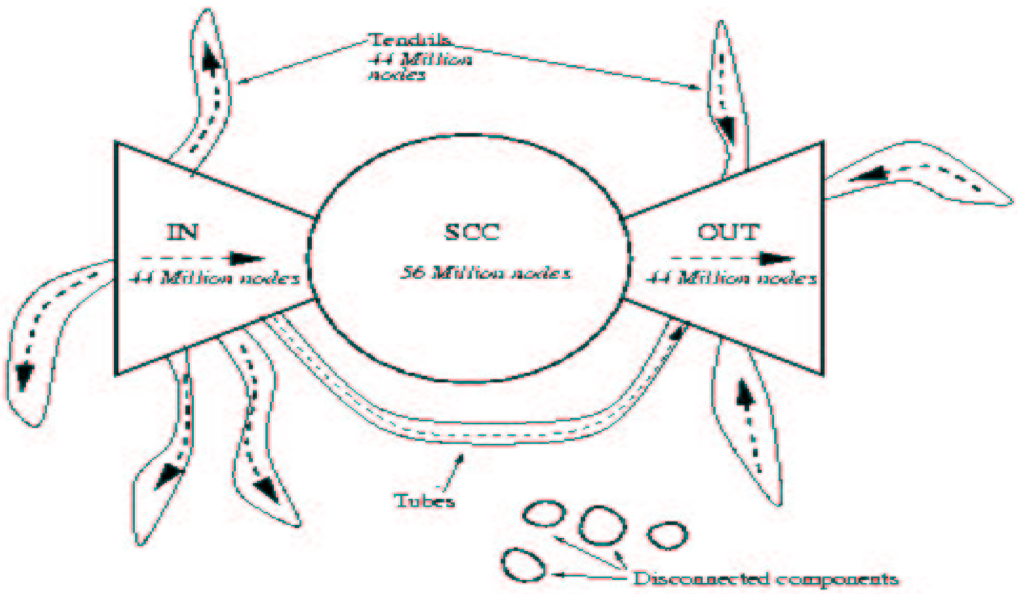
$$S = L^+ \cap L^-$$

More on $d > 1$.

Let L^+ be the set of vertices with a giant “fan-out” and L^- be the set of vertices with a giant “fan-in”. Then **whp**

$$S = L^+ \cap L^-$$

temp.gif (GIF Image, 522x394 pixels) file:///home/alan/texfiles/webgraph/talk/temp.gif



Eigenvalues – Mihail and Papadimitriou
Chung,Lu and Vu.

Eigenvalues – Mihail and Papadimitriou
Chung,Lu and Vu.

Faloutsos,Faloutsos,Faloutsos observed that the largest eigenvalues of the adjacency matrix followed a power law.

Eigenvalues – Mihail and Papadimitriou
Chung,Lu and Vu.

Faloutsos,Faloutsos,Faloutsos observed that the largest eigenvalues of the adjacency matrix followed a power law.

Expected degree model
Fix $a_1, a_2, \ldots, a_n$. Let $A = a_1 + \cdots + a_n$. Then put an edge between i and j with probability $a_i a_j / A$.

Eigenvalues – Mihail and Papadimitriou
Chung,Lu and Vu.

Faloutsos,Faloutsos,Faloutsos observed that the largest eigenvalues of the adjacency matrix followed a power law.

Expected degree model
Fix $a_1, a_2, \ldots, a_n$. Let $A = a_1 + \cdots + a_n$. Then put an edge between i and j with probability $a_i a_j / A$.

Let $\lambda_1 \geq \lambda_2 \geq \cdots$ be the largest eigenvalues of the adjacency matrix of the graph produced.

Suppose that $a_i = \frac{a_1}{i^\alpha}, 1/2 < \alpha < 1$ for $1 \leq i \leq n^\beta$, β sufficiently small.

Suppose that $a_i = \frac{a_1}{i^\alpha}, 1/2 < \alpha < 1$ for $1 \leq i \leq n^\beta$, β sufficiently small.

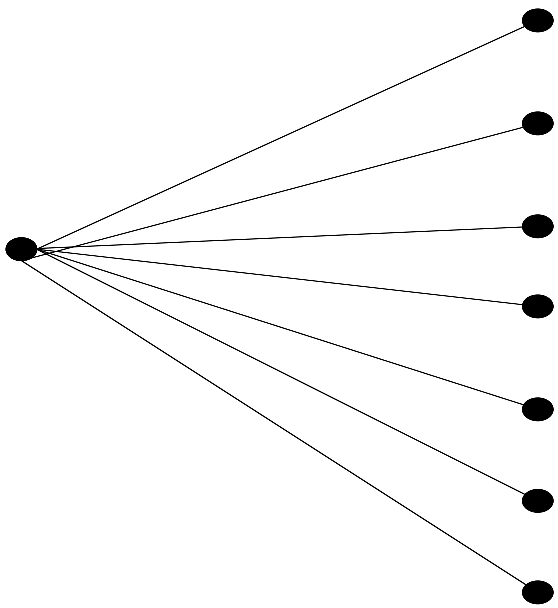
$$(1 - o(1))\sqrt{a_i} \leq \lambda_i \leq (1 + o(1))\sqrt{a_i} \qquad \textbf{whp}$$

for $i \leq n^\beta$.

Suppose that $a_i = \frac{a_1}{i^\alpha}, 1/2 < \alpha < 1$ for $1 \leq i \leq n^\beta$, β sufficiently small.

$$(1 - o(1))\sqrt{a_i} \leq \lambda_i \leq (1 + o(1))\sqrt{a_i} \qquad \mathbf{whp}$$

for $i \leq n^\beta$.



Star of degree d:
largest eigenvalue $d^{1/2}$

Dynamic models: Preferential Attachment Model (PAM).

Barabasi and Albert

Dynamic models: Preferential Attachment Model (PAM).
Barabasi and Albert
We build the graph dynamically:
At time t:

- Add a new vertex v_t.
- Connect v_t to m randomly chosen vertices $u_1, \ldots, u_m$ in $\{v_1, \ldots, v_{t-1}\}$.

Dynamic models: Preferential Attachment Model (PAM).
Barabasi and Albert
We build the graph dynamically:
At time t:

- Add a new vertex v_t.
- Connect v_t to m randomly chosen vertices $u_1, \ldots, u_m$ in $\{v_1, \ldots, v_{t-1}\}$.

$$\mathbf{Pr}(u_i = u) = \frac{deg_{t-1}(u)}{2m(t-1)}$$

Dynamic models: Preferential Attachment Model (PAM).
Barabasi and Albert
We build the graph dynamically:
At time t:

- Add a new vertex v_t.
- Connect v_t to m randomly chosen vertices $u_1, \ldots, u_m$ in $\{v_1, \ldots, v_{t-1}\}$.

$$\mathbf{Pr}(u_i = u) = \frac{deg_{t-1}(u)}{2m(t-1)}$$

The rich get richer in the WWW world.

Dynamic models: Preferential Attachment Model (PAM).
Barabasi and Albert
We build the graph dynamically:
At time t:

- Add a new vertex v_t.
- Connect v_t to m randomly chosen vertices $u_1, \ldots, u_m$ in $\{v_1, \ldots, v_{t-1}\}$.

$$\mathbf{Pr}(u_i = u) = \frac{deg_{t-1}(u)}{2m(t-1)}$$

The rich get richer in the WWW world.
Preferential attachment also used in models by Yule 1925 and Simon 1955.

Let $d_k(t)$ denote the expected number of vertices of degree k at time t.

Let $d_k(t)$ denote the expected number of vertices of degree k at time t.

$$d_k(t+1) = d_k(t) + m\frac{(k-1)d_{k-1}(t)}{2mt} - m\frac{kd_k(t)}{2mt} + 1_{k=m} + error\ terms.$$

Let $d_k(t)$ denote the expected number of vertices of degree k at time t.

$$d_k(t+1) = d_k(t) + m\frac{(k-1)d_{k-1}(t)}{2mt} - m\frac{kd_k(t)}{2mt} + 1_{k=m} + error\ terms.$$

Assume that $d_k(t) \sim d_k t$. Then

$$d_k\left(\frac{k}{2}+1\right) \sim d_{k-1}\frac{k-1}{2} + 1_{k=m}$$

Let $d_k(t)$ denote the expected number of vertices of degree k at time t.

$$d_k(t+1) = d_k(t) + m\frac{(k-1)d_{k-1}(t)}{2mt} - m\frac{kd_k(t)}{2mt} + 1_{k=m} + error\ terms.$$

Assume that $d_k(t) \sim d_k t$. Then

$$d_k\left(\frac{k}{2}+1\right) \sim d_{k-1}\frac{k-1}{2} + 1_{k=m}$$

$$d_k \sim \frac{2m(m+1)}{(k+2)(k+1)k}t \qquad for\ k \geq m.$$

Bollobás,Riordan,Spencer,Tusnady

Diameter: Bollobás,Riordan.

$$Diameter \sim \frac{\log t}{\log \log t}$$

Diameter: Bollobás,Riordan.

$$Diameter \sim \frac{\log t}{\log \log t}$$

Expected degree of vertex v_s at time t is $\sim m\sqrt{t/s}$.

Diameter: Bollobás,Riordan.

$$Diameter \sim \frac{\log t}{\log \log t}$$

Expected degree of vertex v_s at time t is $\sim m\sqrt{t/s}$.
If $i < j$,

$$\mathbf{Pr}(Edge\ (v_i, v_j)\ exists) \leq \omega \cdot \frac{\sqrt{j/i}}{2mj} \leq \frac{\omega}{\sqrt{ij}}$$

where $\omega \to \infty$ slowly.

Diameter: Bollobás,Riordan.

$$Diameter \sim \frac{\log t}{\log\log t}$$

Let $k = (1-\epsilon)\log t / \log\log t$

$$\begin{aligned}
\mathbf{Pr}(\exists Path\ length\ k, t \to t-1) &\leq \sum_{t_0=t,\ldots,t_k=t-1} \prod_{i=1}^{k} \frac{\omega}{\sqrt{t_{i-1}t_i}} \\
&\leq \omega^k \frac{1}{t(t-1)} \left(\sum_{i=1}^{k} \frac{1}{i} \right)^k \\
&= o(1).
\end{aligned}$$

Diameter: Bollobás,Riordan.

$$Diameter \sim \frac{\log t}{\log \log t}$$

Upper bound is harder

Copying Model

Communities: A large dense bipartite sub-graph of the WWW indicates a “community”. Experiments indicate a larger number of communities than you would get say from the simple model PAM.

The next model does give many. It is due to Kumar,Raghavan,Sivakumar,Upfal

Copying Model

Communities: A large dense bipartite sub-graph of the WWW indicates a “community”. Experiments indicate a larger number of communities than you would get say from the simple model PAM.

The next model does give many. It is due to Kumar,Raghavan,Sivakumar,Upfal

The edges of these bipartite cliques are oriented the same way.

As in PAM, at each stage we add a new vertex v_t and give it m edges.
Its construction rests on a parameter $0 < \alpha, 1$.

As in PAM, at each stage we add a new vertex v_t and give it m edges.
Its construction rests on a parameter $0 < \alpha, 1$.

A vertex u is chosen uniformly at random from $\{v_1, \ldots, v_{t-1}\}$ and then for $i = 1, \ldots, m$ we

As in PAM, at each stage we add a new vertex v_t and give it m edges.
Its construction rests on a parameter $0 < \alpha, 1$.

A vertex u is chosen uniformly at random from $\{v_1, \ldots, v_{t-1}\}$ and then for $i = 1, \ldots, m$ we

1. With probability α we create edge (v_t, x) where x is chosen uniformly at random.

As in PAM, at each stage we add a new vertex v_t and give it m edges.
Its construction rests on a parameter $0 < \alpha, 1$.

A vertex u is chosen uniformly at random from $\{v_1, \ldots, v_{t-1}\}$ and then for $i = 1, \ldots, m$ we

1. With probability α we create edge (v_t, x) where x is chosen uniformly at random.

2. With probability $1 - \alpha$ we create edge (v_t, y) where y is the ith choice of u.

Whp the degree sequence has a power law with exponent $\frac{2-\alpha}{1-\alpha}$.

Whp the degree sequence has a power law with exponent $\frac{2-\alpha}{1-\alpha}$.

Whp the number of copies of $K_{r,r}, r \leq m$ is $\Omega(te^{-r})$

Whp the degree sequence has a power law with exponent $\frac{2-\alpha}{1-\alpha}$.

Whp the number of copies of $K_{r,r}, r \leq m$ is $\Omega(te^{-r})$

This contrasts with PAM which has $O(1)$ in expectation.

Deletions: Bollobás and Riordan

Robustness: Suppose we build our PAM graph and then delete the first ct vertices. What remains has a giant ($\Omega(t)$ size) component iff

$$c < \frac{m-1}{m+1}$$

.

Deletions: Bollobás and Riordan

Robustness: Suppose we build our PAM graph and then delete the first ct vertices. What remains has a giant ($\Omega(t)$ size) component iff

$$c < \frac{m-1}{m+1}$$

.

If vertices are deleted randomly with probability $p < 1$ constant. Then there is a giant component for any p.

Deletions: Bollobás and Riordan

Robustness: Suppose we build our PAM graph and then delete the first ct vertices. What remains has a giant ($\Omega(t)$ size) component iff

$$c < \frac{m-1}{m+1}$$

.

If vertices are deleted randomly with probability $p < 1$ constant. Then there is a giant component for any p.

What about deletions during the growing phase?

Random deletion in a scale free random graph

We consider a model where edges are added using preferential attachment and vertices/edges are deleted randomly. Flaxman, Frieze, Vera

Random deletion in a scale free random graph

We consider a model where edges are added using preferential attachment and vertices/edges are deleted randomly. Flaxman, Frieze, Vera

Same model considered by Chung and Lu.

$G_t = (V_t, E_t)$ denotes the graph at time t.

- **Initialisation:** Start with G_1 with vertices x_1 and no edges.
  Step $t \geq 2$

$G_t = (V_t, E_t)$ denotes the graph at time t.

- **Initialisation:** Start with G_1 with vertices x_1 and no edges.
  Step $t \geq 2$
- A Probability $1 - \alpha - \alpha_0$: delete a randomly chosen vertex.

$G_t = (V_t, E_t)$ denotes the graph at time t.

- **Initialisation:** Start with G_1 with vertices x_1 and no edges. Step $t \geq 2$
- A Probability $1 - \alpha - \alpha_0$: delete a randomly chosen vertex.
- B Probability α_0 we delete m randomly chosen edges.

$G_t = (V_t, E_t)$ denotes the graph at time t.

- **Initialisation:** Start with G_1 with vertices x_1 and no edges. Step $t \geq 2$
- A Probability $1 - \alpha - \alpha_0$: delete a randomly chosen vertex.
- B Probability α_0 we delete m randomly chosen edges.
- C Probability α_1: add new vertex x_t and m random neighbours $w_1, \ldots, w_m$.

$$\mathbf{Pr}(w_i = w) = \frac{d(w, t-1)}{2e_{t-1}}. \tag{-3}$$

$G_t = (V_t, E_t)$ denotes the graph at time t.

- **Initialisation:** Start with G_1 with vertices x_1 and no edges. Step $t \geq 2$
- A Probability $1 - \alpha - \alpha_0$: delete a randomly chosen vertex.
- B Probability α_0 we delete m randomly chosen edges.
- C Probability α_1: add new vertex x_t and m random neighbours $w_1, \ldots, w_m$.

$$\mathbf{Pr}(w_i = w) = \frac{d(w, t-1)}{2e_{t-1}}. \qquad \text{(-3)}$$

- D Probability $\alpha - \alpha_1$: Add m random edges with endpoints chosen independently as in (-3).

We skip over details of what to do if there are no vertices to delete or what to do with multiple edges etc.

$D_k(t)$ is the number of vertices of degree k in G_t and $\overline{D}_k(t) = \mathbf{E}(D_k(t))$.

$$\beta = \frac{2(\alpha - \alpha_0)}{3\alpha - 1 - \alpha_1 - \alpha_0}$$

Theorem

Under natural restrictions on the parameters, there exists a constant $C = C(m, \alpha, \alpha_0, \alpha_1)$ such that for $k \geq 1$,

$$\left| \frac{\overline{D}_k(t)}{t} - Ck^{-1-\beta} \right| = O\left(t^{-\epsilon}\right) + O\left(k^{-2-\beta}\right).$$

Suppose that v_t, e_t denote the number of vertices and edges in G_t.

Suppose that v_t, e_t denote the number of vertices and edges in G_t.

$$\overline{D}_k(t+1) = \overline{D}_k(t) +$$

$$(2\alpha - \alpha_1) m \mathsf{E}\left(-\frac{k D_k(t)}{2e_t} + \frac{(k-1) D_{k-1}(t)}{2e_t} \;\middle|\; e_t > 0 \right) \mathbf{Pr}(e_t > 0)$$

$$+ (1-\alpha)(k+1) \mathsf{E}\left(\frac{D_{k+1}(t) - D_k(t)}{v_t} \;\middle|\; e_t > 0 \right) \mathbf{Pr}(e_t > 0)$$

$$+ \alpha_1 1_{k=m} + \text{ error terms}.$$

Let

$$\nu = \alpha + \alpha_0 + \alpha_1 - 1 > 0 \text{ and } \eta = \frac{m(\alpha - \alpha_0)\nu}{1 + \alpha_1 - \alpha - \alpha_0}.$$

We show

Let

$$\nu = \alpha + \alpha_0 + \alpha_1 - 1 > 0 \text{ and } \eta = \frac{m(\alpha - \alpha_0)\nu}{1 + \alpha_1 - \alpha - \alpha_0}.$$

We show

$$|v_t - \nu t| \leq t^{1/2} \log t, \qquad \mathbf{qs}.$$

Let

$$\nu = \alpha + \alpha_0 + \alpha_1 - 1 > 0 \text{ and } \eta = \frac{m(\alpha - \alpha_0)\nu}{1 + \alpha_1 - \alpha - \alpha_0}.$$

We show

$$|v_t - \nu t| \leq t^{1/2} \log t, \qquad \mathbf{qs}.$$

$$\mathbf{Pr}(|e_t - \eta t| \geq t^{1-\epsilon}) = O(t^{-\epsilon}).$$

Let

$$\nu = \alpha + \alpha_0 + \alpha_1 - 1 > 0 \text{ and } \eta = \frac{m(\alpha - \alpha_0)\nu}{1 + \alpha_1 - \alpha - \alpha_0}.$$

We show

$$|v_t - \nu t| \leq t^{1/2} \log t, \qquad \mathbf{qs}.$$

$$\mathbf{Pr}(|e_t - \eta t| \geq t^{1-\epsilon}) = O(t^{-\epsilon}).$$

We used Chebychef to handle e_t. Chung and Lu modify Azuma's inequality, avoids constraint on edge deletion prob.

As a consequence we can write

$$\overline{D}_k(t+1) = \overline{D}_k(t) + (A_2(k+1) + B_2)\frac{\overline{D}_{k+1}(t)}{t} + (A_1 k + B_1 + 1)\frac{\overline{D}_k(t)}{t} + (A_0(k-1) + B_0)\frac{\overline{D}_{k-1}(t)}{t} + \alpha_1 1_{k=m} + O(t^{-\epsilon}).$$

$$A_2 = \frac{1-\alpha-\alpha_0}{\nu} + \frac{m\alpha_0}{\eta} \qquad B_2 = 0$$

$$A_1 = -\frac{(2\alpha-\alpha_1+2\alpha_0)m}{2\eta} - \frac{1-\alpha-\alpha_0}{\nu} \qquad B_1 = -1 - \frac{1-\alpha-\alpha_0}{\nu}$$

$$A_0 = \frac{(2\alpha-\alpha_1)m}{2\eta} \qquad B_0 = 0$$

Assume $\overline{D}_k(t) \sim d_k t$. $d_{-1} = 0$ and for $k \geq -1$,

$$(A_2(k+2) + B_2)d_{k+2} + (A_1(k+1) + B_1)d_{k+1} + (A_0 k + B_0)d_k = -\alpha_1 1_{k=m-1}. \qquad \text{(-2)}$$

Assume $\overline{D}_k(t) \sim d_k t$. $d_{-1} = 0$ and for $k \geq -1$,

$$(A_2(k+2) + B_2)d_{k+2} + (A_1(k+1) + B_1)d_{k+1} + (A_0 k + B_0)d_k = -\alpha_1 1_{k=m-1}. \qquad \text{(-2)}$$

We first tackle homogeneous equation: for $k \geq 1$

$$(A_2(k+2) + B_2)e_{k+2} + (A_1(k+1) + B_1)e_{k+1} + (A_0 k + B_0)e_k = 0,$$

Assume $\overline{D}_k(t) \sim d_k t$. $d_{-1} = 0$ and for $k \geq -1$,

$$(A_2(k+2) + B_2)d_{k+2} + (A_1(k+1) + B_1)d_{k+1} + (A_0 k + B_0)d_k = -\alpha_1 1_{k=m-1}. \qquad \text{(-2)}$$

We first tackle homogeneous equation: for $k \geq 1$

$$(A_2(k+2) + B_2)e_{k+2} + (A_1(k+1) + B_1)e_{k+1} + (A_0 k + B_0)e_k = 0,$$

We use Laplace's method and make the substitution

$$e_k = \int_{t=a}^{t=b} t^{k-1} v(t) dt$$

for $a, b, v(t)$ to be determined.

Integrating by parts

$$ke_k = [t^k v(t)]_a^b - \int_{t=a}^{t=b} t^k v'(t) dt$$

Integrating by parts

$$k e_k = [t^k v(t)]_a^b - \int_{t=a}^{t=b} t^k v'(t) dt$$

Let

$$\phi_1(t) = A_2 t^2 + A_1 t + A_0, \ \phi_0(t) = B_2 t^2 + B_1 t + B_0$$

Integrating by parts

$$ke_k = [t^k v(t)]_a^b - \int_{t=a}^{t=b} t^k v'(t)dt$$

Let

$$\phi_1(t) = A_2t^2 + A_1t + A_0,\ \phi_0(t) = B_2t^2 + B_1t + B_0$$

Substituting gives

$$[t^k\phi_1(t)v(t)]_a^b - \int_a^b t^k\phi_1(t)v'(t)dt + \int_a^b t^{k-1}\phi_0(t)v(t)dt = 0.$$

So, $v(t)$ will give a solution to the homogeneous equation if

$$[t^k v(t)\phi_1(t)]_a^b = 0 \qquad \text{and} \qquad \frac{v'(t)}{v(t)} = \frac{\phi_0(t)}{t\phi_1(t)}.$$

So, $v(t)$ will give a solution to the homogeneous equation if

$$[t^k v(t)\phi_1(t)]_a^b = 0 \qquad \text{and} \qquad \frac{v'(t)}{v(t)} = \frac{\phi_0(t)}{t\phi_1(t)}.$$

The differential equation is homogeneous and can be integrated to give,

$$v(t) = C_0(t-1)^{\beta}\,(t-A)^{-\beta}$$

where $A > 1$ and $C_0 \neq 0$.

We take $a = 0, b = 1$ and satisfy $[t^k v(t)\phi_1(t)]_a^b = 0$.

We take $a = 0, b = 1$ and satisfy $[t^k v(t)\phi_1(t)]_a^b = 0$.

Substituting we get the following solution to the homogeneous equation: valid for $k \geq 1$,

$$u_1(k) = \int_0^1 t^{k-1} \left(\frac{1-t}{1-\gamma t} \right)^\beta dt$$

where

$$\gamma < 1.$$

$$
\begin{aligned}
u_1(k) &= \int_0^1 t^{k-1}(1-t)^\beta \frac{1}{(1-\gamma t)^\beta} dt \\
&= \int_0^1 t^{k-1}(1-t)^\beta \sum_{j=0}^{\infty} \binom{\beta+j-1}{j} (\gamma t)^j dt \\
&= \sum_{j=0}^{\infty} \binom{\beta+j-1}{j} \gamma^j \int_0^1 t^{k+j-1}(1-t)^\beta dt \\
&= \sum_{j=0}^{\infty} \binom{\beta+j-1}{j} \gamma^j \frac{\Gamma(k+j)\Gamma(\beta+1)}{\Gamma(k+j+\beta+1)} \\
&= \sum_{j=0}^{\infty} \gamma^j \frac{\Gamma(\beta+j)}{\Gamma(j+1)\Gamma(\beta)} \frac{\Gamma(k+j)\Gamma(\beta+1)}{\Gamma(k+j+\beta+1)}
\end{aligned}
$$

assuming k is large, using Stirling for $\Gamma(k+j), \Gamma(k+j+\beta+1)$, we get

$$
\begin{aligned}
&= (1+O(k^{-1}))\beta \sum_{j=0}^{\infty} \gamma^j \frac{\Gamma(j+\beta)}{\Gamma(j+1)} (k+\beta+j)^{-\beta-1} \\
&= (1+O(k^{-1}))C_1 k^{-1-\beta}
\end{aligned}
$$

We need to consider the non-homogeneous equation.

We need to consider the non-homogeneous equation.

$d_{-1} = 0$ and for $k \geq -1$,

$$\begin{aligned}(A_2(k+2)+B_2)d_{k+2} + (A_1(k+1)+B_1)d_{k+1} + (A_0k+B_0)d_k \\ = -\alpha_1 1_{k=m-1}. \qquad \text{(-2)}\end{aligned}$$

We need to consider the non-homogeneous equation.

$d_{-1} = 0$ and for $k \geq -1$,

$$(A_2(k+2) + B_2)d_{k+2} + (A_1(k+1) + B_1)d_{k+1} + (A_0 k + B_0)d_k = -\alpha_1 1_{k=m-1}. \qquad \text{(-2)}$$

We show there is a solution such that

$$d_k = C u_1(k)$$

for $k \geq m$.

Adversarial deletion in a scale free random graph:
Flaxman, Frieze, Vera

Adversarial deletion in a scale free random graph:
Flaxman, Frieze, Vera

What if an adversary is allowed to delete ϵn vertices as the the graph grows.

Adversarial deletion in a scale free random graph:
Flaxman, Frieze, Vera

What if an adversary is allowed to delete ϵn vertices as the the graph grows.

We show that if m is sufficiently large then the final graph has a giant component (size $\Theta(n)$) **whp.**

Adversarial deletion in a scale free random graph: Flaxman, Frieze, Vera

What if an adversary is allowed to delete ϵn vertices as the the graph grows.

We show that if m is sufficiently large then the final graph has a giant component (size $\Theta(n)$) **whp.**

Bollobás and Riordan allow the adversary to delete vertices at the end only.

Adversarial deletion in a scale free random graph: Flaxman, Frieze, Vera

What if an adversary is allowed to delete ϵn vertices as the the graph grows.

We show that if m is sufficiently large then the final graph has a giant component (size $\Theta(n)$) **whp.**

Bollobás and Riordan allow the adversary to delete vertices at the end only.

On the other hand, we use their idea of coupling with $G_{n,p}$.

A geometric scale free random graph:
Flaxman, Frieze, Vera.

A geometric scale free random graph:
Flaxman, Frieze, Vera.

$V_n = \{X_1, X_2, \ldots, X_n\}$ are chosen uniformly from the surface of the unit sphere in $\boldsymbol{R}^3$

A geometric scale free random graph:
Flaxman, Frieze, Vera.

$V_n = \{X_1, X_2, \ldots, X_n\}$ are chosen uniformly from the surface of the unit sphere in $\boldsymbol{R}^3$

Step t: Add m random edges from

$$X_t \text{ to } Y_1, Y_2, \ldots, Y_m \in \{X_1, X_2, \ldots, X_{t-1}\}$$

where

$$|Y_i - X_t| \leq r = n^{\epsilon - 1/2}$$

and the Y_i's are chosen via preferential attachment.

Theorem

- If m is a sufficiently large constant then there exists a constant $c > 0$ such that

$$d_k(n) \sim \frac{cn}{k(k+1)(k+2)}$$

where $d_k(n)$ is the number of vertices of degree k at time n.

Theorem

- If m is a sufficiently large constant then there exists a constant $c > 0$ such that

$$d_k(n) \sim \frac{cn}{k(k+1)(k+2)}$$

 where $d_k(n)$ is the number of vertices of degree k at time n.

- **Whp** V_n can be partitioned into $T, \bar{T}$ such that $|T|, |\bar{T}| \sim n/2$, and there are at most $4\sqrt{\pi r}nm$ edges between T and $\bar{T}$.

Theorem

- If m is a sufficiently large constant then there exists a constant $c > 0$ such that

$$d_k(n) \sim \frac{cn}{k(k+1)(k+2)}$$

 where $d_k(n)$ is the number of vertices of degree k at time n.

- **Whp** V_n can be partitioned into $T, \bar{T}$ such that $|T|, |\bar{T}| \sim n/2$, and there are at most $4\sqrt{\pi r n m}$ edges between T and $\bar{T}$.
- If $m \geq K \log n$ and K is sufficiently large then **whp** G_n is connected.

Heuristically Optimized Trade-offs

Carlson and Doyle; Fabrikant, Koutsoupias and Papadimitriou

Heuristically Optimized Trade-offs

Carlson and Doyle; Fabrikant, Koutsoupias and Papadimitriou

$X_1, X_2, \ldots, X_n$ are chosen uniformly at random in the unit square $[0, 1]^2$.

Heuristically Optimized Trade-offs

Carlson and Doyle; Fabrikant, Koutsoupias and Papadimitriou

$X_1, X_2, \ldots, X_n$ are chosen uniformly at random in the unit square $[0,1]^2$.

After $X_1, \ldots, X_i$ have beeen generated there is a tree T_i on them.

Heuristically Optimized Trade-offs

Carlson and Doyle; Fabrikant, Koutsoupias and Papadimitriou

$X_1, X_2, \ldots, X_n$ are chosen uniformly at random in the unit square $[0,1]^2$.

After $X_1, \ldots, X_i$ have beeen generated there is a tree T_i on them.

Then at step $i+1$, X_{i+1} is joined by an edge to the vertex X_j which minimises

$$\alpha|X_{i+1} - X_j| + h_j$$

where h_j is the number of edges from X_j to X_1 in T_i.

Theorem

If $\alpha < 1/\sqrt{2}$ then T_n is a star.

Theorem

If $\alpha < 1/\sqrt{2}$ then T_n is a star.

If $\alpha = o(n^{1/2}/(\log n)^2)$ then **whp** T_n has $n - o(n)$ leaves.

Theorem

If $\alpha < 1/\sqrt{2}$ then T_n is a star.

If $\alpha = o(n^{1/2}/(\log n)^2)$ then **whp** T_n has $n - o(n)$ leaves.

If $\alpha/(n^{1/2}\log n) \to \infty$ then $A_1 e^{-c_1 k} \leq \rho_k \leq A_2 e^{-c-2k}$ where ρ_k is the proportion of vertices of degree k

Theorem

If $\alpha < 1/\sqrt{2}$ then T_n is a star.

If $\alpha = o(n^{1/2}/(\log n)^2)$ then **whp** T_n has $n - o(n)$ leaves.

If $\alpha/(n^{1/2}\log n) \to \infty$ then $A_1 e^{-c_1 k} \le \rho_k \le A_2 e^{-c-2k}$ where ρ_k is the proportion of vertices of degree k

If $\alpha \ge 4$ and $\alpha = o(n^{1/2})$ then **E**$(\rho_{\ge k}) \ge c(k/n)^{\beta}$.

Theorem

If $\alpha < 1/\sqrt{2}$ then T_n is a star.

If $\alpha = o(n^{1/2}/(\log n)^2)$ then **whp** T_n has $n - o(n)$ leaves.

If $\alpha/(n^{1/2}\log n) \to \infty$ then $A_1 e^{-c_1 k} \leq \rho_k \leq A_2 e^{-c-2k}$ where ρ_k is the proportion of vertices of degree k

If $\alpha \geq 4$ and $\alpha = o(n^{1/2})$ then $\mathbf{E}(\rho_{\geq k}) \geq c(k/n)^{\beta}$.

Some of these results are from Berger,Bollobás,Borgs,Chayes,Riordan:
Also Berger,Borgs,Chayes,D'Souza,Kleinberg

Crawling on web graphs

Cooper and Frieze

Crawling on web graphs
Cooper and Frieze

Sequence of random graphs $G(t) : G(t) = G(t-1)$ plus vertex t and m random edges $\{t, v_i\}, i = 1, 2, \ldots, m$.

Crawling on web graphs
Cooper and Frieze

Sequence of random graphs $G(t) : G(t) = G(t-1)$ plus vertex t and m random edges $\{t, v_i\}, i = 1, 2, \ldots, m$.

Model 1: The vertices $v_1, v_2, \ldots, v_m$ are chosen uniformly with replacement from $[t-1]$.

Crawling on web graphs
Cooper and Frieze

Sequence of random graphs $G(t) : G(t) = G(t-1)$ plus vertex t and m random edges $\{t, v_i\}, i = 1, 2, \ldots, m$.

Model 1: The vertices $v_1, v_2, \ldots, v_m$ are chosen uniformly with replacement from $[t-1]$.

Model 2 The vertices $v_1, v_2, \ldots, v_m$ are chosen proportional to their degree after step $t-1$.

Crawling on web graphs
Cooper and Frieze

Sequence of random graphs $G(t) : G(t) = G(t-1)$ plus vertex t and m random edges $\{t, v_i\}, i = 1, 2, \ldots, m$.

Model 1: The vertices $v_1, v_2, \ldots, v_m$ are chosen uniformly with replacement from $[t-1]$.

Model 2 The vertices $v_1, v_2, \ldots, v_m$ are chosen proportional to their degree after step $t-1$.

Spider $\mathcal{S}$ sits at vertex X_{t-1} of $G(t-1)$. After the addition of vertex t, and before step $t+1$, spider makes a random walk of length ℓ

$\nu_{\ell,m}(t)$ is the expected number of vertices **not** visited by $\mathcal{S}$ at the end of step t.

$\nu_{\ell,m}(t)$ is the expected number of vertices **not** visited by $\mathcal{S}$ at the end of step t.

Theorem

In either model, if m is sufficiently large then,

$$\nu_{\ell,m}(t) \sim \mathbf{E} \sum_{s=1}^{t} \prod_{\tau=s}^{t} \left(1 - \frac{d(s,\tau)}{2m\tau}\right)^{\ell}$$

where $d(s,\tau)$ denotes the degree of s in $G(\tau)$.

Error in expression of order m^{-1} omitted.

Let

$$\eta_\ell = \lim_{m\to\infty} \lim_{t\to\infty} \frac{\mathbf{E}\nu_{\ell,m}(t)}{t}.$$

Let

$$\eta_\ell = \lim_{m\to\infty} \lim_{t\to\infty} \frac{\mathbf{E}\nu_{\ell,m}(t)}{t}.$$

(a) For Model 1,

$$\eta_\ell = \sqrt{\frac{2}{\ell}} e^{(\ell+2)^2/(4\ell)} \int_{(\ell+2)/\sqrt{2\ell}}^{\infty} e^{-y^2/2}\, dy$$

where $\Psi(x)$ is the standardized Normal cumulate for the interval $(-\infty, x]$.
In particular, $\eta_1 = 0.57\cdots$ and $\eta_\ell \sim 2/\ell$ as $\ell \to \infty$.

Let

$$\eta_\ell = \lim_{m\to\infty} \lim_{t\to\infty} \frac{\mathbf{E}\nu_{\ell,m}(t)}{t}.$$

(b) For Model 2

$$\eta_\ell = e^\ell 2\ell^2 \int_\ell^\infty y^{-3} e^{-y}\, dy.$$

In particular, $\eta_1 = 0.59\cdots$. and $\eta_\ell \sim 2/\ell$ as $\ell \to \infty$.

Further work

- Try other models of a random web-graph.
- Try non-uniform random walks.
- Prove concentration of the number of vertices visited.

The proof technique is robust enough to handle other models and walks, once one has established rapid mixing. The calculations for various non-uniform walks can get tedious.
It should be possible to estimate the variance of the number of unvisited vertices and apply Chebychef. Stronger concentration seems more challenging.

Cover time of preferential attachment graph

Cover time of preferential attachment graph

$G = (V, E)$ is a connected graph. ($|V| = n, |E| = m$). For $v \in V$ let C_v be the expected time taken for a simple random walk W on G starting at v, to visit every vertex of G.

Cover time of preferential attachment graph

$G = (V, E)$ is a connected graph. ($|V| = n, |E| = m$). For $v \in V$ let C_v be the expected time taken for a simple random walk W on G starting at v, to visit every vertex of G.

The *cover time* C_G of G is defined as $C_G = \max_{v \in V} C_v$.

Cover time of preferential attachment graph

$G = (V, E)$ is a connected graph. ($|V| = n, |E| = m$). For $v \in V$ let C_v be the expected time taken for a simple random walk W on G starting at v, to visit every vertex of G.

The *cover time* C_G of G is defined as $C_G = \max_{v \in V} C_v$.

$(1 - o(1))n \log n \leq C_G \leq (1 + o(1))\frac{4}{27}n^3$: Feige

Cover time of preferential attachment graph

$G = (V, E)$ is a connected graph. ($|V| = n, |E| = m$). For $v \in V$ let C_v be the expected time taken for a simple random walk W on G starting at v, to visit every vertex of G.

The *cover time* C_G of G is defined as $C_G = \max_{v \in V} C_v$.

Cooper and Frieze If G is preferential attachment graph then **whp**

$$C_G \sim \frac{2m}{m-1} n \log n$$

Effect of search engines: Chakrabarti, Frieze, Vera

Effect of search engines: Chakrabarti, Frieze, Vera

The model has parameters p, N

Effect of search engines: Chakrabarti, Frieze, Vera
The model has parameters p, N

At time step t we add vertex v_t and m randomly chosen edges to $\{u_1, \ldots, u_m\}$. For each i, u_i is chosen preferentially from

- With probability p we choose u_i from N vertices of largest degree..
- With probability $q = 1 - p$ we choose u_i from all vertices.

Effect of search engines: Chakrabarti, Frieze, Vera
The model has parameters p, N

At time step t we add vertex v_t and m randomly chosen edges to $\{u_1, \ldots, u_m\}$. For each i, u_i is chosen preferentially from

- With probability p we choose u_i from N vertices of largest degree..
- With probability $q = 1 - p$ we choose u_i from all vertices.

Supposed to model surfer who obtains links from first N given by search engine.

Theorem

(a) For $i \leq N$ there exists constant $\alpha_i > 0$ such that

$$\mathbf{E}\left[\deg_n(x_i)\right] = \alpha_i n + O\left(n^{1/2}\right)$$

(b) There is an absolute constant A_1 such that for every

$$k \geq m, \bar{d}_k(n) = (1 + o(1))\frac{A_1 n}{k^{1+2/(1-p)}}.$$

where $\bar{d}_k(n)$ is the expected number of vertices of degree k outside the N largest vertices.