Directed Scale-free Graphs

B. Bollobás, C. Borgs, J. Chayes, O. Riordan

April 11

Outline

Outline

1. Description of the model
2. Analysis of the model
3. Results
4. Other models

Outline

Outline

Description of the model

- Graph $G(t_0) = G_0$ with n_0 vertices and t_0 edges and constants $\alpha, \beta, \gamma, \delta_{in}$ and δ_{out} s.t. $\alpha + \beta + \gamma = 1$
- At timestep t, perform one of the following
  - w.p. α add a new vertex v and an edge (v, w) to an existing vertex w
  - w.p. β add an edge (v, w) to the existing graph
  - w.p. γ add a new vertex v and an edge (w, v) from an existing vertex w

Analysis of the model

- $G(t)$- Graph at time t
- $n(t) - \#$ of vertices at time t
- $x_i(t) - \#$ of nodes with in-degree i at time t
- $y_i(t) - \#$ of nodes with out-degree i at time t

Analysis of the model(contd.)

Theorem: Let $i \geq 0$ be fixed. There are constants p_i and q_i such that $x_i(t) = p_i t + o(t)$ and $y_i(t) = q_i t + o(t)$ hold with probability 1. Furthermore, if $\alpha\delta_{in} + \gamma > 0$ and $\gamma < 1$, then as $i \to \infty$ we have

$$p_i \sim C_{IN} i^{-X_{IN}}.$$

If $\gamma\delta_{out} + \alpha > 0$ and $\alpha < 1$, then as $i \to \infty$ we have

$$q_i \sim C_{OUT} i^{-X_{OUT}}.$$

Analysis of the model(contd.)

Observation: Number of vertices, $n(t)$ is n_0 plus a Binomial distribution with mean $(\alpha + \gamma)(t - t_0)$.

Chernoff Bound

Let $X_1, X_2, ..., X_n$ be independent Bernoulli trials such that, for $1 \leq i \leq n$, $\mathbf{Pr}[X_i = 1] = p$, where $0 < p < 1$. Then, for $X = \sum_{i=1}^{n} X_i, \mu = E[X] = np$, and any $0 < \delta < 1$,

$$\mathbf{Pr}[X > (1+\delta)\mu] < \left[\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right]^{\mu}$$

$$\mathbf{Pr}[X < (1-\delta)\mu] < e^{-\mu\delta^2/2}$$

Analysis of the model(contd.)

By the Chernoff bound, it follows that

$$\mathbf{Pr}[|n(t) - (\alpha + \gamma)t| \geq t^{1/2} \log t] \leq e^{-c(logt)^2}.$$

In particular, the probability above is $o(t^{-1})$ as $t \to \infty$.

Analysis of the model(contd.)

- Now, we will estimate $x_i(t)$ – # of vertices of in-degree i at time t.
- In going from $G(t)$ to $G(t+1)$, what is the chance of destroying a vertex of degree i?
- The answer

$$(\alpha+\beta)x_i(t)\frac{i+\delta_{in}}{t+\delta_{in}n(t)}.$$

Analysis of the model(contd.)

- Now, we will estimate $x_i(t) - \#$ of vertices of in-degree i at time t.
- In going from $G(t)$ to $G(t+1)$, what is the chance of destroying a vertex of degree i?
- The answer

$$(\alpha + \beta)x_i(t)\frac{i + \delta_{in}}{t + \delta_{in}n(t)}.$$

Analysis of the model(contd.)

- Now, we will estimate $x_i(t) - \#$ of vertices of in-degree i at time t.
- In going from $G(t)$ to $G(t+1)$, what is the chance of destroying a vertex of degree i?
- The answer

$$(\alpha + \beta) x_i(t) \frac{i + \delta_{in}}{t + \delta_{in} n(t)}.$$

Analysis of the model(contd.)

Putting everything together, we get

$$
\begin{aligned}
\mathbf{E}[x_i(t+1)|G(t)] = & x_i(t) \quad + \\
& \frac{(\alpha+\beta)}{t+\delta_{in}n(t)}\Big((i-1+\delta_{in})x_{i-1}(t)-(i+\delta_{in})x_i(t)\Big)+ \\
& \alpha\mathbf{1}_{\{i=0\}}+\gamma\mathbf{1}_{\{i=1\}}.
\end{aligned}
$$

$$x_{-1}(t)=0.$$

Analysis of the model(contd.)

Putting everything together, we get

$$\mathbf{E}[x_i(t+1)|G(t)] = x_i(t) \quad + \frac{(\alpha+\beta)}{t+\delta_{in}n(t)}\Big((i-1+\delta_{in})x_{i-1}(t) - (i+\delta_{in})x_i(t)\Big) + \alpha\mathbf{1}_{\{i=0\}} + \gamma\mathbf{1}_{\{i=1\}}.$$

$x_{-1}(t) = 0.$

Analysis of the model(contd.)

Putting everything together, we get

$$\begin{aligned}\mathbf{E}[x_i(t+1)|G(t)] = & x_i(t) \quad + \\ & \frac{(\alpha+\beta)}{t+\delta_{in}n(t)}\Big((i-1+\delta_{in})x_{i-1}(t)-(i+\delta_{in})x_i(t)\Big)+ \\ & \alpha\mathbf{1}_{\{i=0\}}+\gamma\mathbf{1}_{\{i=1\}}.\end{aligned}$$

$x_{-1}(t) = 0.$

Analysis of the model(contd.)

By the Chernoff bound, it follows that

$$\mathbf{Pr}[|n(t) - (\alpha + \gamma)t| \geq t^{1/2} \log t] \leq e^{-c(logt)^2}.$$

In particular, the probability above is $o(t^{-1})$ as $t \to \infty$.

We could assume that w.p. $1 - o(t^{-1})$,

$$|n(t) - (\alpha + \gamma)t| = o(t^{3/5}).$$

$$(\alpha + \beta)\frac{i + \delta_{in}}{t + \delta_{in} n(t)} x_i(t) = O(1) \quad \text{true for } n(t) \geq 0$$

Analysis of the model(contd.)

By the Chernoff bound, it follows that

$$\mathbf{Pr}[|n(t) - (\alpha + \gamma)t| \geq t^{1/2} \log t] \leq e^{-c(logt)^2}.$$

In particular, the probability above is $o(t^{-1})$ as $t \to \infty$.

We could assume that w.p. $1 - o(t^{-1})$,

$$|n(t) - (\alpha + \gamma)t| = o(t^{3/5}).$$

$$(\alpha + \beta)\frac{i + \delta_{in}}{t + \delta_{in} n(t)} x_i(t) = O(1) \quad \text{true for } n(t) \geq 0$$

Analysis of the model(contd.)

By the Chernoff bound, it follows that

$$\mathbf{Pr}[|n(t) - (\alpha + \gamma)t| \geq t^{1/2} \log t] \leq e^{-c(log t)^2}.$$

In particular, the probability above is $o(t^{-1})$ as $t \to \infty$.

We could assume that w.p. $1 - o(t^{-1})$,

$$|n(t) - (\alpha + \gamma)t| = o(t^{3/5}).$$

$$(\alpha + \beta)\frac{i + \delta_{in}}{t + \delta_{in} n(t)} x_i(t) = O(1) \quad \text{true for } n(t) \geq 0$$

Analysis of the model(contd.)

$$\mathbf{E}\left[\frac{\alpha+\beta}{t+\delta_{in}n(t)}(i+\delta_{in})x_i(t)\right]$$
$$=\frac{\alpha+\beta}{t+\delta_{in}(\alpha+\beta)t}(i+\delta_{in})\mathbf{E}x_i(t)\left(1+o(t^{-2/5})\right)$$
$$=\frac{\alpha+\beta}{t+\delta_{in}(\alpha+\beta)t}(i+\delta_{in})\mathbf{E}x_i(t)+o(t^{-2/5})$$

.

Analysis of the model(contd.)

$$\begin{aligned}
&\mathbf{E}\left[\frac{\alpha+\beta}{t+\delta_{in}n(t)}(i+\delta_{in})x_i(t)\right] \\
&\qquad = \frac{\alpha+\beta}{t+\delta_{in}(\alpha+\beta)t}(i+\delta_{in})\mathbf{E}x_i(t)\left(1+o(t^{-2/5})\right) \\
&\qquad = \frac{\alpha+\beta}{t+\delta_{in}(\alpha+\beta)t}(i+\delta_{in})\mathbf{E}x_i(t)+o(t^{-2/5})
\end{aligned}$$

.

Analysis of the model(contd.)

$$\mathbf{E}[x_i(t+1)|G(t)] = x_i(t) \quad + \frac{(\alpha+\beta)}{t+\delta_{in}n(t)}\Big((i-1+\delta_{in})x_{i-1}(t)-(i+\delta_{in})x_i(t)\Big) + \alpha\mathbf{1}_{\{i=0\}}+\gamma\mathbf{1}_{\{i=1\}}.$$

Taking expectation of both sides,

$$\mathbf{E}x_i(t+1) = \mathbf{E}x_i(t) \quad + \frac{(\alpha+\beta)}{t+\delta_{in}(\alpha+\beta)t}\Big((i-1+\delta_{in})\mathbf{E}x_{i-1}(t)-(i+\delta_{in})\mathbf{E}x_i(t)\Big) + \alpha\mathbf{1}_{\{i=0\}}+\gamma\mathbf{1}_{\{i=1\}}+o(t^{-2/5}).$$

Analysis of the model(contd.)

$$\mathbf{E}[x_i(t+1)|G(t)] = x_i(t) \quad + \\ \frac{(\alpha+\beta)}{t+\delta_{in}n(t)}\Big((i-1+\delta_{in})x_{i-1}(t)-(i+\delta_{in})x_i(t)\Big)+ \\ \alpha\mathbf{1}_{\{i=0\}}+\gamma\mathbf{1}_{\{i=1\}}.$$

Taking expectation of both sides,

$$\mathbf{E}x_i(t+1) = \mathbf{E}x_i(t) \quad + \\ \frac{(\alpha+\beta)}{t+\delta_{in}(\alpha+\beta)t}\Big((i-1+\delta_{in})\mathbf{E}x_{i-1}(t)-(i+\delta_{in})\mathbf{E}x_i(t)\Big) \\ +\alpha\mathbf{1}_{\{i=0\}}+\gamma\mathbf{1}_{\{i=1\}}+o(t^{-2/5}).$$

Analysis of the model(contd.)

As in the Preferential attachment model, we will assume

$$\mathbf{E}x_i(t) \sim p_i t$$

Now we obtain a recurrence relation in p_i's.

$$\begin{aligned} p_i(t+1) = & p_i t \quad + \\ & \frac{(\alpha+\beta)}{t+\delta_{in}(\alpha+\beta)t}\Big((i-1+\delta_{in})p_{i-1}t-(i+\delta_{in})p_i t\Big) \\ & +\alpha \mathbf{1}_{\{i=0\}}+\gamma \mathbf{1}_{\{i=1\}}. \end{aligned}$$

Analysis of the model(contd.)

As in the Preferential attachment model, we will assume

$$\mathbf{E}x_i(t) \sim p_i t$$

Now we obtain a recurrence relation in p_i's.

$$\begin{aligned} p_i(t+1) = & p_i t \quad + \\ & \frac{(\alpha+\beta)}{t+\delta_{in}(\alpha+\beta)t}\Big((i-1+\delta_{in})p_{i-1}t-(i+\delta_{in})p_i t\Big) \\ & + \alpha \mathbf{1}_{\{i=0\}} + \gamma \mathbf{1}_{\{i=1\}}. \end{aligned}$$

Analysis of the model(contd.)

Let

$$c_1 = \frac{(\alpha + \beta)}{1 + \delta_{in}(\alpha + \beta)}.$$

Let $p_{-1} = 0$ and for $i \geq 0$,

$$p_i = c_1\Big((i - 1 + \delta_{in})p_{i-1} - (i + \delta_{in})p_i\Big) + \alpha\mathbf{1}_{\{i=0\}} + \gamma\mathbf{1}_{\{i=1\}}.$$

Analysis of the model(contd.)

Solving the recurrence, we get $p_0 = \alpha/(1 + c_1\delta_m)$,

$$p_1 = (1 + \delta_m + c_1^{-1})^{-1} \left(\frac{\alpha\delta_i n}{1 + c_1\delta_{in}} + \frac{\gamma}{c_1} \right),$$

and for $i \geq 1$,

$$\begin{aligned} p_i &= \frac{(i - 1 + \delta_{in})_{i-1}}{(i + \delta_{in} + c_1^{-1})_{i-1}} p_1 \\ &= \frac{(i - 1 + \delta_{in})!}{(i + \delta_{in} + c_1^{-1})!} \frac{(1 + \delta_{in} + c_1^{-1})!}{\delta_{in}!} p_1 . \end{aligned}$$

Analysis of the model(contd.)

Wanted to show that

$$p_i \sim C_{IN} i^{-X_{IN}}$$

$$X_{IN} = (\delta_{in} + c_1^{-1}) - (\delta_{in} + c_1^{-1}) = 1 + c_1^{-1}$$

$$X_{OUT} = (\delta_{out} + c_2^{-1}) - (\delta_{out} + c_2^{-1}) = 1 + c_2^{-1}$$

where $c_1 = \frac{\alpha+\beta}{1+\delta_{in}(\alpha+\gamma)}$ and $c_2 = \frac{\beta+\gamma}{1+\delta_{out}(\alpha+\gamma)}$.

Analysis of the model(contd.)

Wanted to show that

$$p_i \sim C_{IN} i^{-X_{IN}}$$

$$X_{IN} = (\delta_{in} + c_1^{-1}) - (\delta_{in} + c_1^{-1}) = 1 + c_1^{-1}$$

$$X_{OUT} = (\delta_{out} + c_2^{-1}) - (\delta_{out} + c_2^{-1}) = 1 + c_2^{-1}$$

where $c_1 = \frac{\alpha+\beta}{1+\delta_{in}(\alpha+\gamma)}$ and $c_2 = \frac{\beta+\gamma}{1+\delta_{out}(\alpha+\gamma)}$.

Analysis of the model(contd.)

Wanted to show that

$$p_i \sim C_{IN} i^{-X_{IN}}$$

$$X_{IN} = (\delta_{in} + c_1^{-1}) - (\delta_{in} + c_1^{-1}) = 1 + c_1^{-1}$$

$$X_{OUT} = (\delta_{out} + c_2^{-1}) - (\delta_{out} + c_2^{-1}) = 1 + c_2^{-1}$$

where $c_1 = \frac{\alpha+\beta}{1+\delta_{in}(\alpha+\gamma)}$ and $c_2 = \frac{\beta+\gamma}{1+\delta_{out}(\alpha+\gamma)}$.

Analysis of the model(contd.)

Theorem:

$$|\mathbf{E}x_i(t) - p_i t| = O(t^{3/5}). \tag{1}$$

Proof: Recurrence for expectation

$$\begin{aligned}\mathbf{E}x_i(t+1) = &\mathbf{E}x_i(t) \quad + \\ &\frac{(\alpha+\beta)}{t+\delta_{in}(\alpha+\beta)t}\Big((i-1+\delta_{in})\mathbf{E}x_{i-1}(t) - (i+\delta_{in})\mathbf{E}x_i(t)\Big) \\ &+ \alpha\mathbf{1}_{\{i=0\}} + \gamma\mathbf{1}_{\{i=1\}} + o(t^{-2/5}).\end{aligned}$$

Writing $\mathbf{E}x_i(t) = p_i t + \epsilon_i(t)$,

$$\epsilon_i(t+1) = \left(\frac{c_1(i-1+\delta_{in})}{t}\right)\epsilon_{i-1}(t) + \left(1 - \frac{c_1(i+\delta_{in})}{t}\right)\epsilon_i(t) + o(t^{-2/5}).$$

Analysis of the model(contd.)

Theorem:

$$|\mathbf{E}x_i(t) - p_i t| = O(t^{3/5}). \tag{1}$$

Proof: Recurrence for expectation

$$\begin{aligned}\mathbf{E}x_i(t+1) = &\mathbf{E}x_i(t) \quad + \\ &\frac{(\alpha+\beta)}{t+\delta_{in}(\alpha+\beta)t}\Big((i-1+\delta_{in})\mathbf{E}x_{i-1}(t) - (i+\delta_{in})\mathbf{E}x_i(t)\Big) \\ &+ \alpha\mathbf{1}_{\{i=0\}} + \gamma\mathbf{1}_{\{i=1\}} + o(t^{-2/5}).\end{aligned}$$

Writing $\mathbf{E}x_i(t) = p_i t + \epsilon_i(t)$,

$$\epsilon_i(t+1) = \Big(\frac{c_1(i-1+\delta_{in})}{t}\Big)\epsilon_{i-1}(t) + \Big(1 - \frac{c_1(i+\delta_{in})}{t}\Big)\epsilon_i(t) + o(t^{-2/5}).$$

Analysis of the model(contd.)

Theorem:

$$|\mathbf{E}x_i(t) - p_i t| = O(t^{3/5}). \qquad (1)$$

Proof: Recurrence for expectation

$$\begin{aligned}\mathbf{E}x_i(t+1) = &\mathbf{E}x_i(t) \quad + \\ &\frac{(\alpha+\beta)}{t+\delta_{in}(\alpha+\beta)t}\Big((i-1+\delta_{in})\mathbf{E}x_{i-1}(t) - (i+\delta_{in})\mathbf{E}x_i(t)\Big) \\ &+ \alpha\mathbf{1}_{\{i=0\}} + \gamma\mathbf{1}_{\{i=1\}} + o(t^{-2/5}).\end{aligned}$$

Writing $\mathbf{E}x_i(t) = p_i t + \epsilon_i(t)$,

$$\epsilon_i(t+1) = \Big(\frac{c_1(i-1+\delta_{in})}{t}\Big)\epsilon_{i-1}(t) + \Big(1 - \frac{c_1(i+\delta_{in})}{t}\Big)\epsilon_i(t) + o(t^{-2/5}).$$

Analysis of the model(contd.)

Proof contd: We induct on t.
Assume $|\epsilon_i(t)| \leq At^{3/5} \; \forall \; i \geq 0$

$$|\epsilon_i(t+1)| \leq$$

$$\left(\frac{c_1(i-1+\delta_{in})}{t}\right)|\epsilon_{i-1}(t)| + \left(1 - \frac{c_1(i+\delta_{in})}{t}\right)|\epsilon_i(t)| + |o(t^{-2/5})|$$

$$\left(\frac{c_1(i-1+\delta_{in})}{t}\right)At^{3/5} + \left(1 - \frac{c_1(i+\delta_{in})}{t}\right)At^{3/5} + |o(t^{-2/5})|$$

$$\left(1 - \frac{c_1}{t}\right)At^{3/5} + |o(t^{-2/5})|$$

Analysis of the model(contd.)

Proof contd: For large t, we approximate

$$(t+1)^{3/5} = t^{3/5}(1+1/t)^{3/5} \geq t^{3/5}(1+\frac{2}{5}t^{-1}).$$

$$\begin{aligned}
|\epsilon_i(t+1)| &\leq \left(1-\frac{c_1}{t}\right)At^{3/5} + |o(t^{-2/5})| \\
&\leq At^{3/5} - c_1 t^{-2/5} + |o(t^{-2/5})| \\
&\leq A(t+1)^{3/5} - \frac{2}{5}t^{-2/5} - c_1 t^{-2/5} + |o(t^{-2/5})| \\
&\leq A(t+1)^{3/5}
\end{aligned}$$

Azuma-Hoeffding inequality

Suppose that $X_0, X_1, ..., X_n$ is a martingale w.r.t $A_0, A_1, ..., A_n$, and $a_i \le X_{i+1} - X_i \le b_i \;\; i = 1, 2, ..., n-1$, then for any $t \ge 0$, we have

$$\mathbf{Pr}[|X_n - X_0| \ge t] \le 2e^{-t^2/\sum_i (b_i - a_i)^2}.$$

We use Azuma-Hoeffding inequality to show concentration of $x_i(t)$.

Azuma-Hoeffding inequality

Suppose that $X_0, X_1, ..., X_n$ is a martingale w.r.t $A_0, A_1, ..., A_n$, and $a_i \leq X_{i+1} - X_i \leq b_i \quad i = 1, 2, ..., n-1$, then for any $t \geq 0$, we have

$$\mathbf{Pr}[|X_n - X_0| \geq t] \leq 2e^{-t^2/\sum_i (b_i - a_i)^2}.$$

We use Azuma-Hoeffding inequality to show concentration of $x_i(t)$.

Restriction in the sequence of choices

- A-H inequality can be applied only to those sequences where the number of vertices introduced is roughly the mean.
- Denote the set of such sequences by $\mathcal{A}$
- **Pr**(a sequence $C \in \mathcal{A}) \geq 1 - o(t^{-1})$

It is easy to see that

$$\mathbf{E}(x_i(t)|\mathcal{A}) = p_i t + o(t) \tag{2}$$

Concentration

Given $\mathcal{A}$,

- $\#$ of old vertices involved in determining $G(t)$ at most $2t$
- Changing a choice from v to v' only affects the degrees of v and v'
- $x_i(t)$ changes by at most 2.

Applying A-H inequality,

$$\mathbf{Pr}\left[|x_i(t) - \mathbf{E}(x_i(t)|\mathcal{A})| \geq t^{3/4}|\mathcal{A}\right] \leq 2e^{-2t^{3/2}/((2t)2^2)} = 2e^{-\sqrt{t}/4}. \tag{3}$$

Theorem: Assume $\alpha, \gamma < 1$, and that $\alpha\delta_{in} + \gamma\delta_{out} > 0$. Let $i, j \geq 0$ be fixed. Let $n_{i,j} - \#$ of vertices with in-degree i and out-degree j. Then there is a constant $f_{i,j}$ such that $n_{ij}(t) = f_{ij}t + o(t)$ holds with probability 1. Furthermore, for $j \geq 1$ fixed and $i \to \infty$,

$$f_{ij} \sim C_j i^{-X'_{IN}}$$

while for $i \geq 1$ fixed and $j \to \infty$,

$$f_{ij} \sim D_i j^{-X'_{OUT}}.$$

where the C_j and D_i are positive constants.

Results

- In the world wide web, $X_{IN} = 2.1$ and $X_{OUT} = 2.7$.
- For $c_2 = 0.59, \alpha = 0.41, c_1 = 1/1.1$ and

$$\delta_{in} = \frac{1.1(\alpha + \beta) - 1}{1 - \beta},$$

the model achieves the measured exponents.

Other Models

Nodes with *fitness*:

- Every node has an in-fitness and out-fitness associated with it, denoted by λ_v and μ_v.
- λ_v and μ_v are drawn from distributions D_{IN} and D_{OUT} on the non-negative reals.