Complex Graphs and Networks Lecture 3: Duplication models for biological networks

Linyuan Lu

lu@math.sc.edu

University of South Carolina

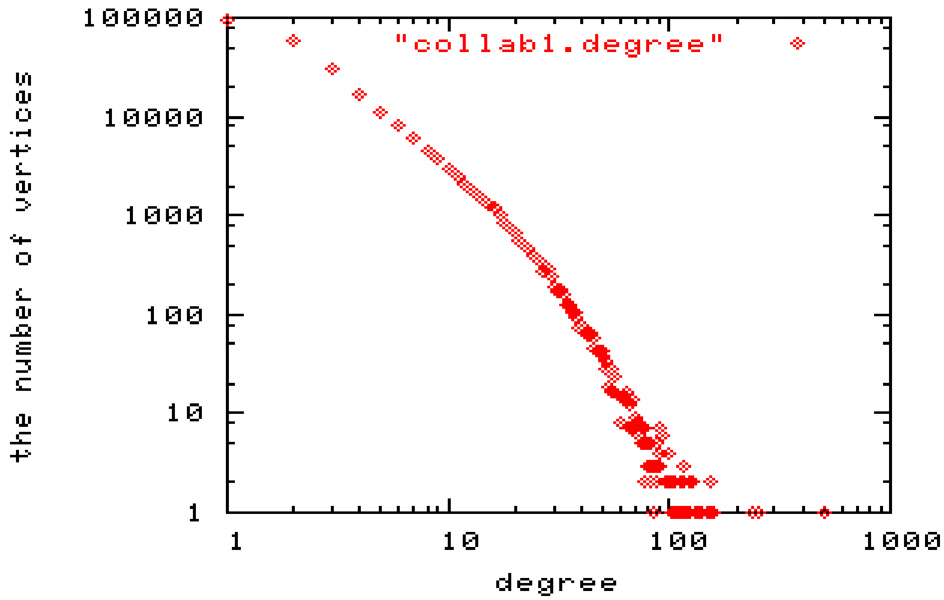
BASICS2008 SUMMER SCHOOL
July 27 – August 2, 2008

Overview of talks

- Lecture 1: Overview and outlines
- Lecture 2: Generative models - preferential attachment schemes
- Lecture 3: Duplication models for biological networks
- Lecture 4: The rise of the giant component
- Lecture 5: The small world phenomenon: average distance and diameter
- Lecture 6: Spectrum of random graphs with given degrees

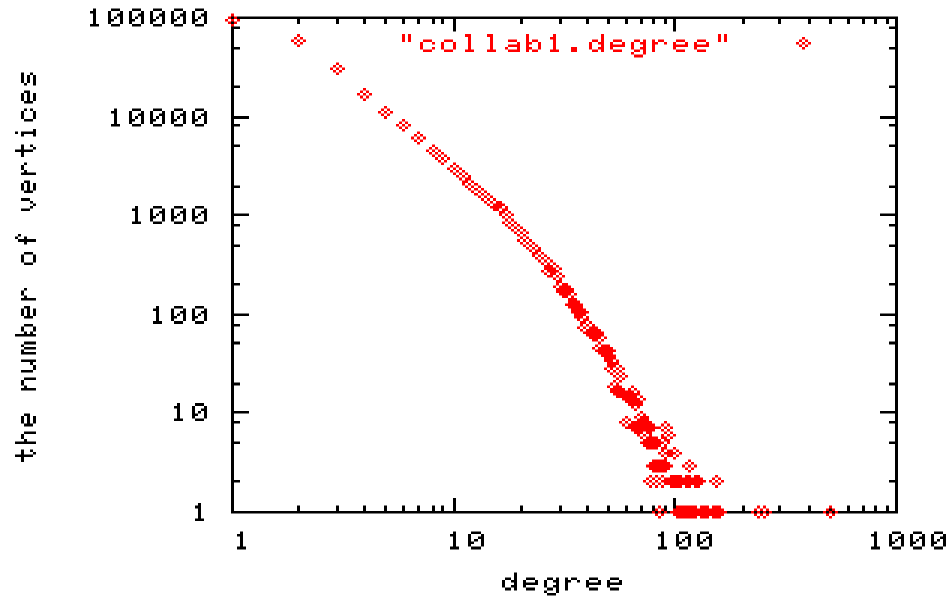
The power law

The number of vertices of degree k is approximately proportional to $k^{-\beta}$ for some positive β.



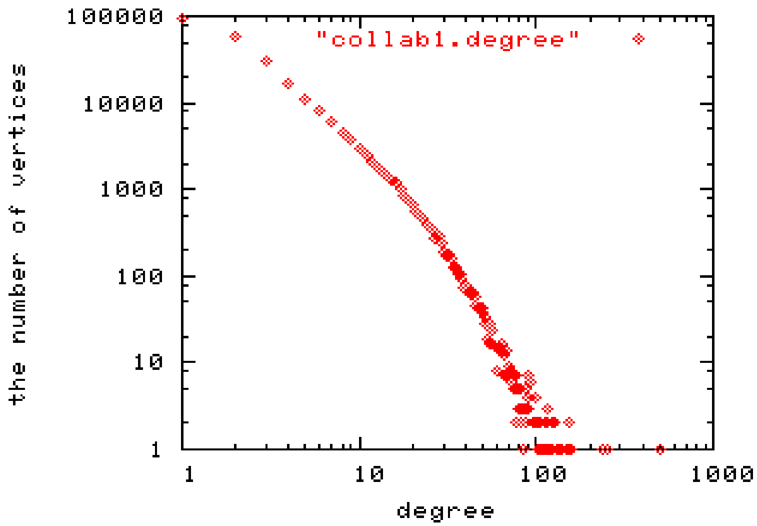
The power law

The number of vertices of degree k is approximately proportional to $k^{-\beta}$ for some positive β.



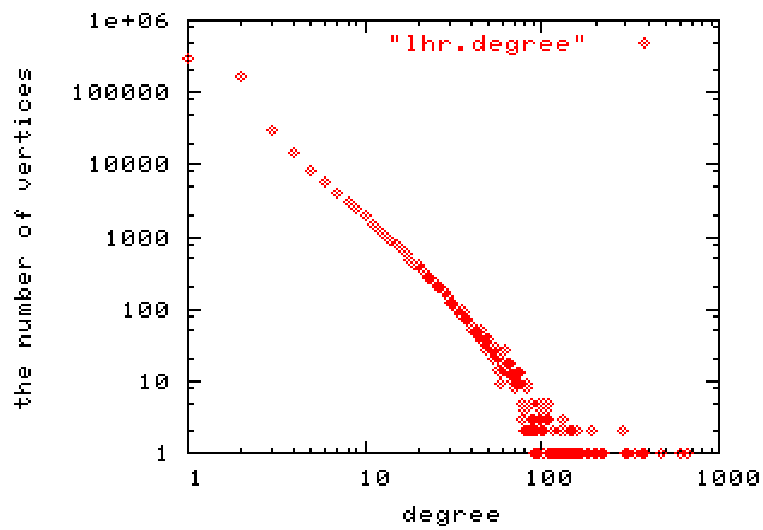
A power law graph is a graph whose degree sequence satisfies the power law.

Power law distribution

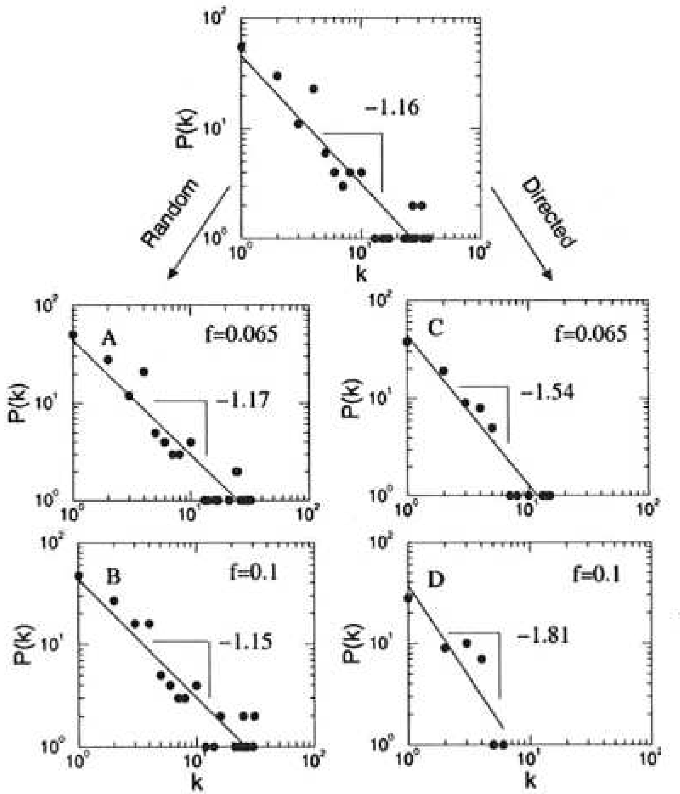


Left: The collaboration graph follows the power law degree distribution with exponent $\beta \approx 3.0$

Right: An IP graph follows the power law degree distribution with exponent $\beta \approx 2.4$



Power law in ecological networks



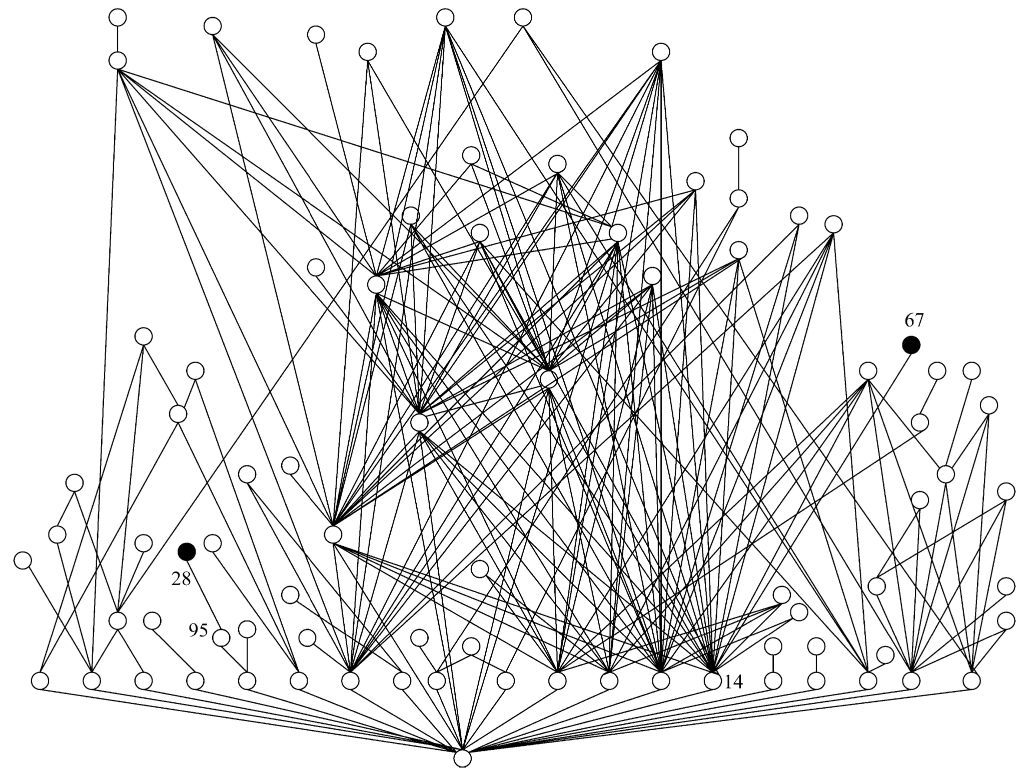
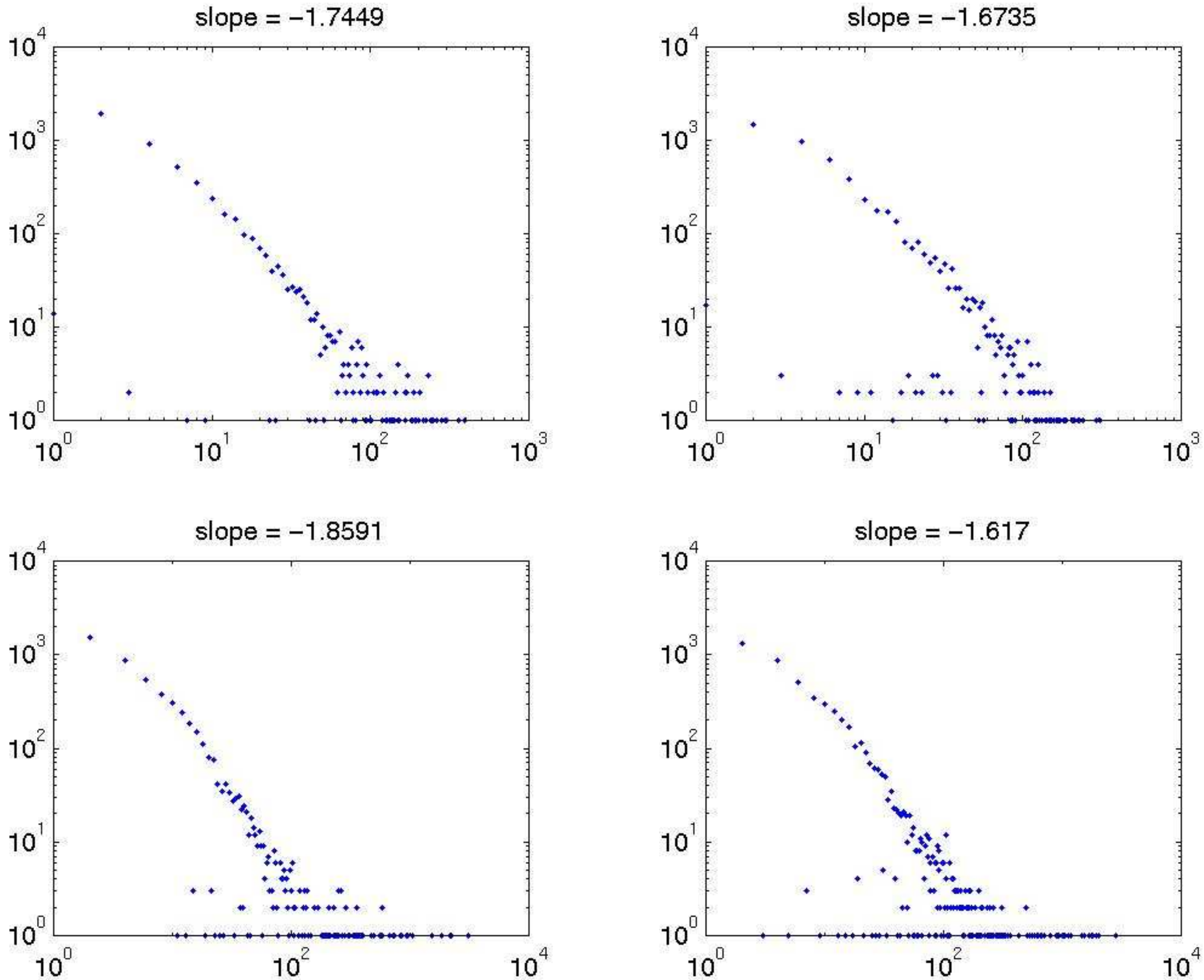


Fig. 1. The food web depicting trophic relationships between species living on broom (the single plant). Points represent trophic species having exactly the same sets of predators and prey species (the web contains 82 trophic species). Thus, this web is an

$$P(k) \sim k^{-\beta}$$

Jordan and Scheuring, Oikos, 2002

Ecological networks



Functional associations of proteins

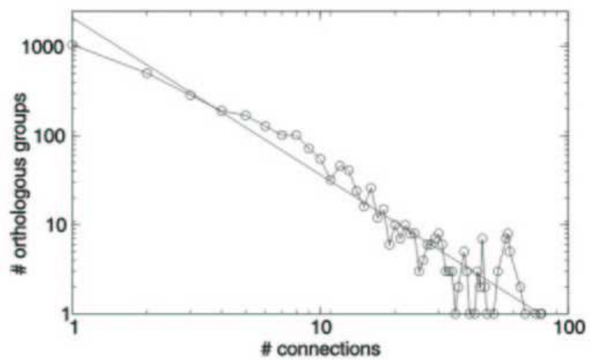
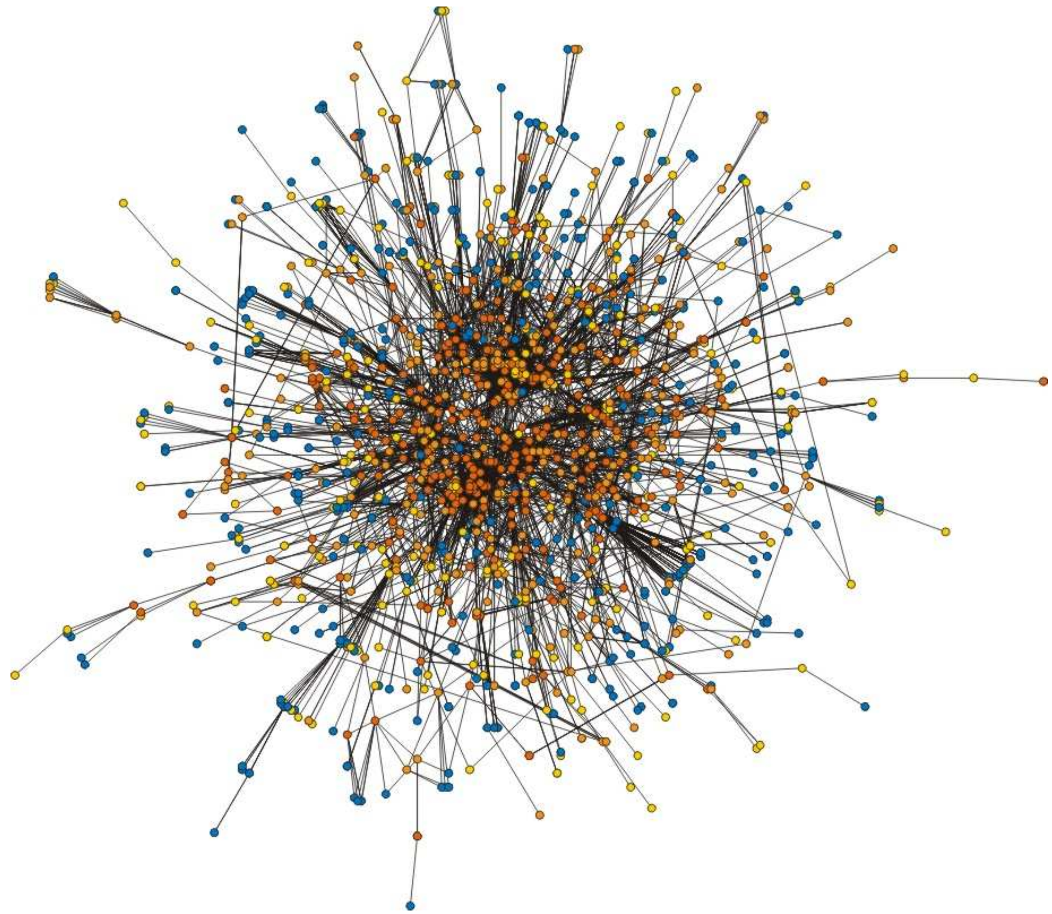


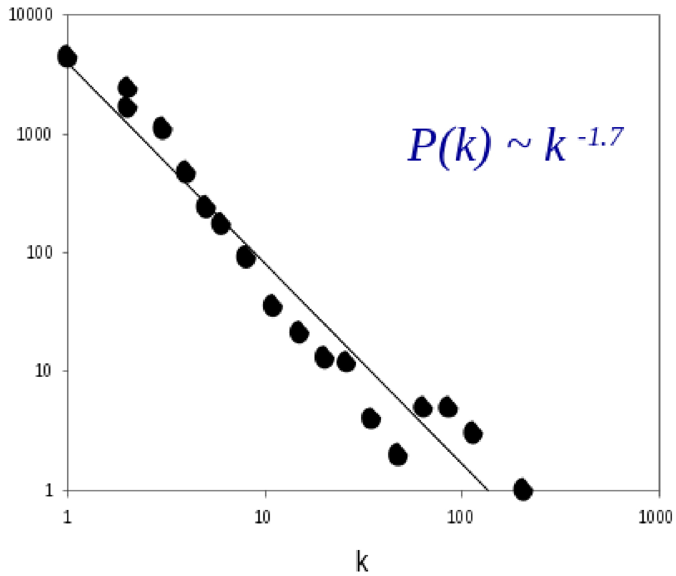
Fig. 2. Distribution of the number of associations per orthologous group. The drawn line is a power law fit to the data.

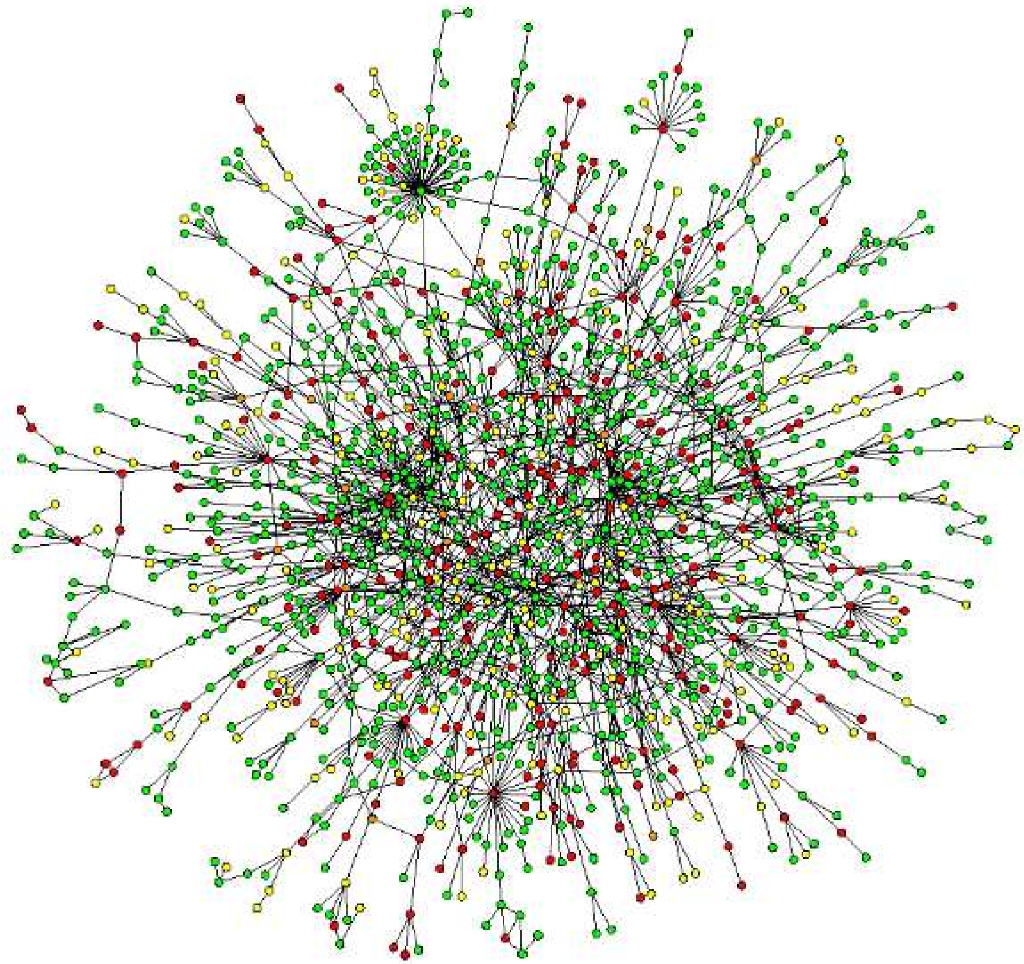


$$P(k) \sim k^{-1.6}$$

Snel, Bork & Huynen, PNAS 99, (2002)

A map of protein-protein interactions in saccharomyces cerevisiae





Jeong, Mason, Barabasi, Oltwai, Nature, 2001

Biological networks versus non-biological networks

Biological Networks	β
Yeast Gene Expression	1.5
Yeast Protein-Protein Maps	1.5, 1.7, 2.1
E. Coli Metabolic Map	1.7, 2.1
Ecology	1.7, 2.1

Other Networks	β
WWW Graphs	2.1 (in), 2.5 (out)
Collaboration Graphs	3
Call Graphs	2.2
Costars Graph of Actors	2.3

A critical threshold $\beta = 2$

Range	$1 < \beta < 2$	$2 < \beta$
Average degree	Unbounded	Bounded
Examples	Biological networks	Non-biological networks
Known evolution models	None	Many

The reference

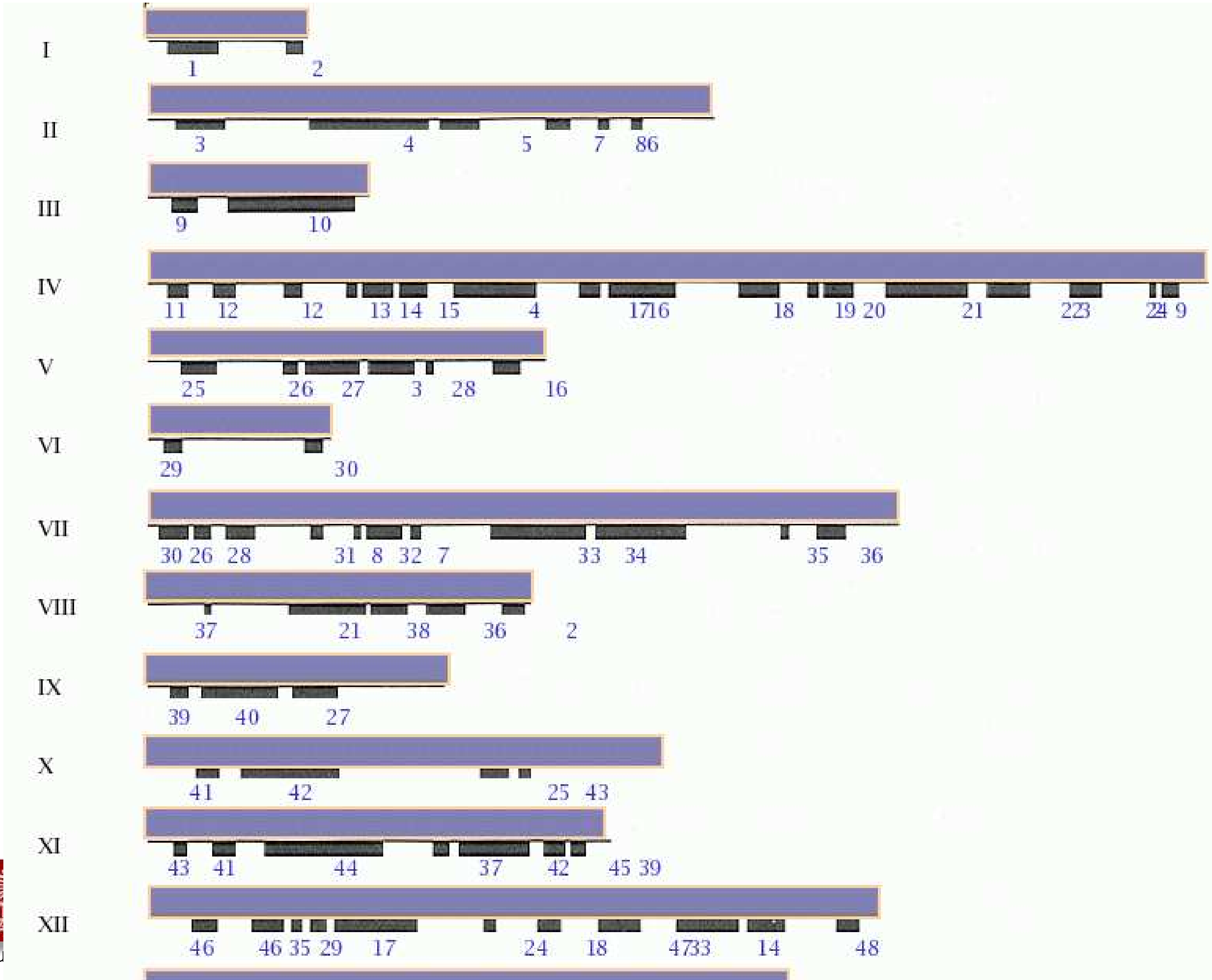
[1.] Fan Chung and Linyuan Lu, T. Gregory Dewey, and David J. Galas. Duplication models for biological networks, *Journal of Computational Biology*, **10**, No. 5, (2003), 677-688.

Genome evolution

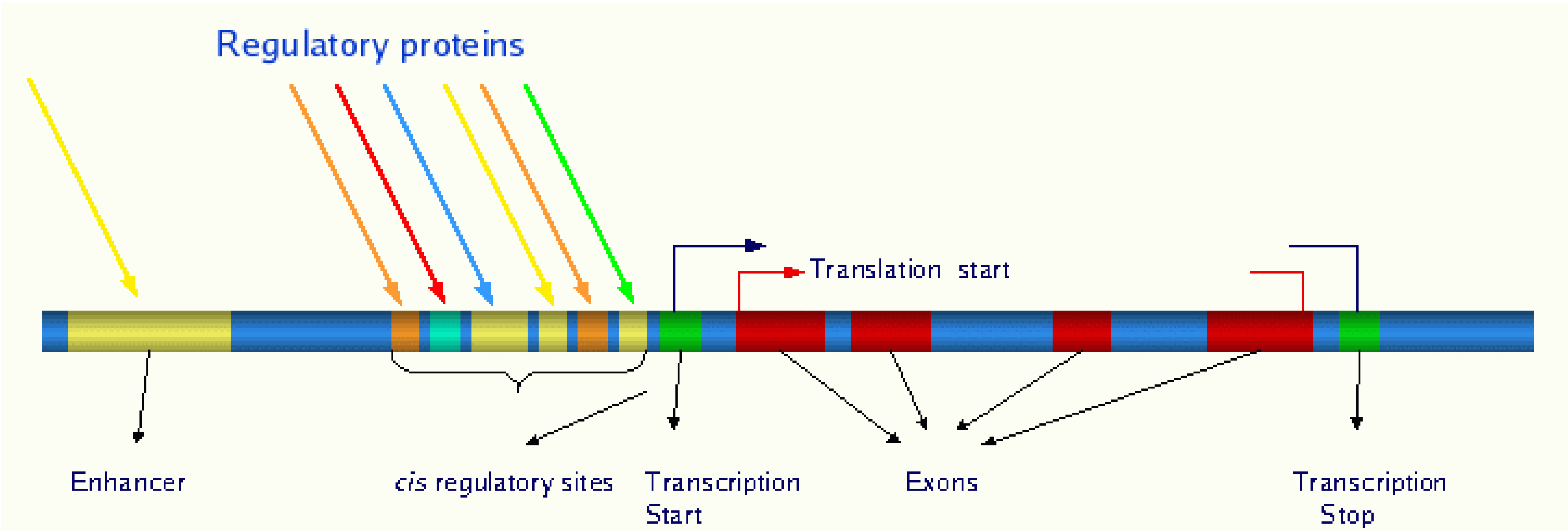
Susumu Ohno's insight

- **The best source of new genes is old genes, and that's where they come from!**
- Gene duplication can include the duplication of regulatory regions - both nodes and edges are duplicated.
- This may not be the only way to use old information for new purposes, but it's a major one.

Genomic duplications in saccharomyces cerevisiae

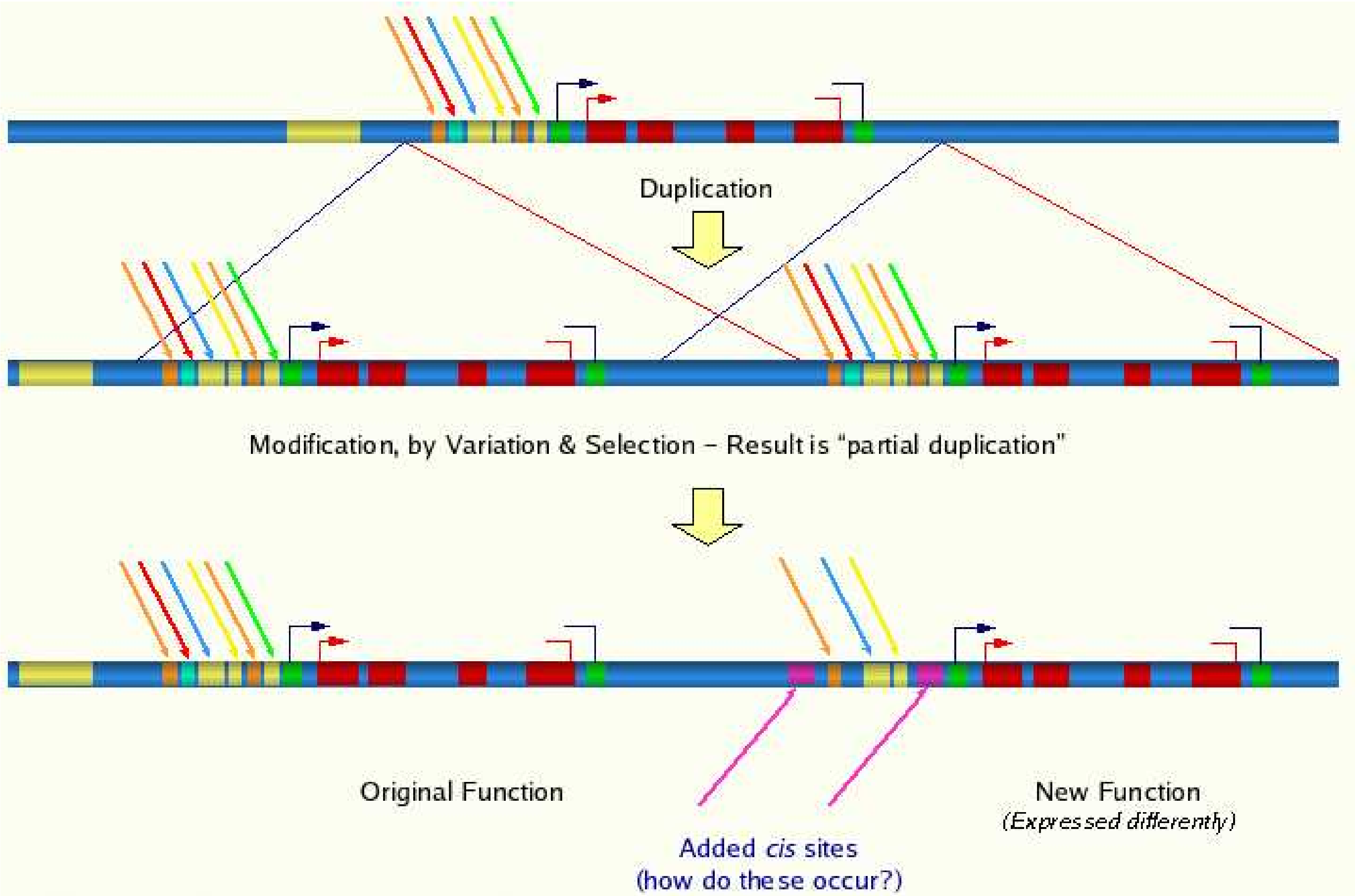


Gene regulatory graphs

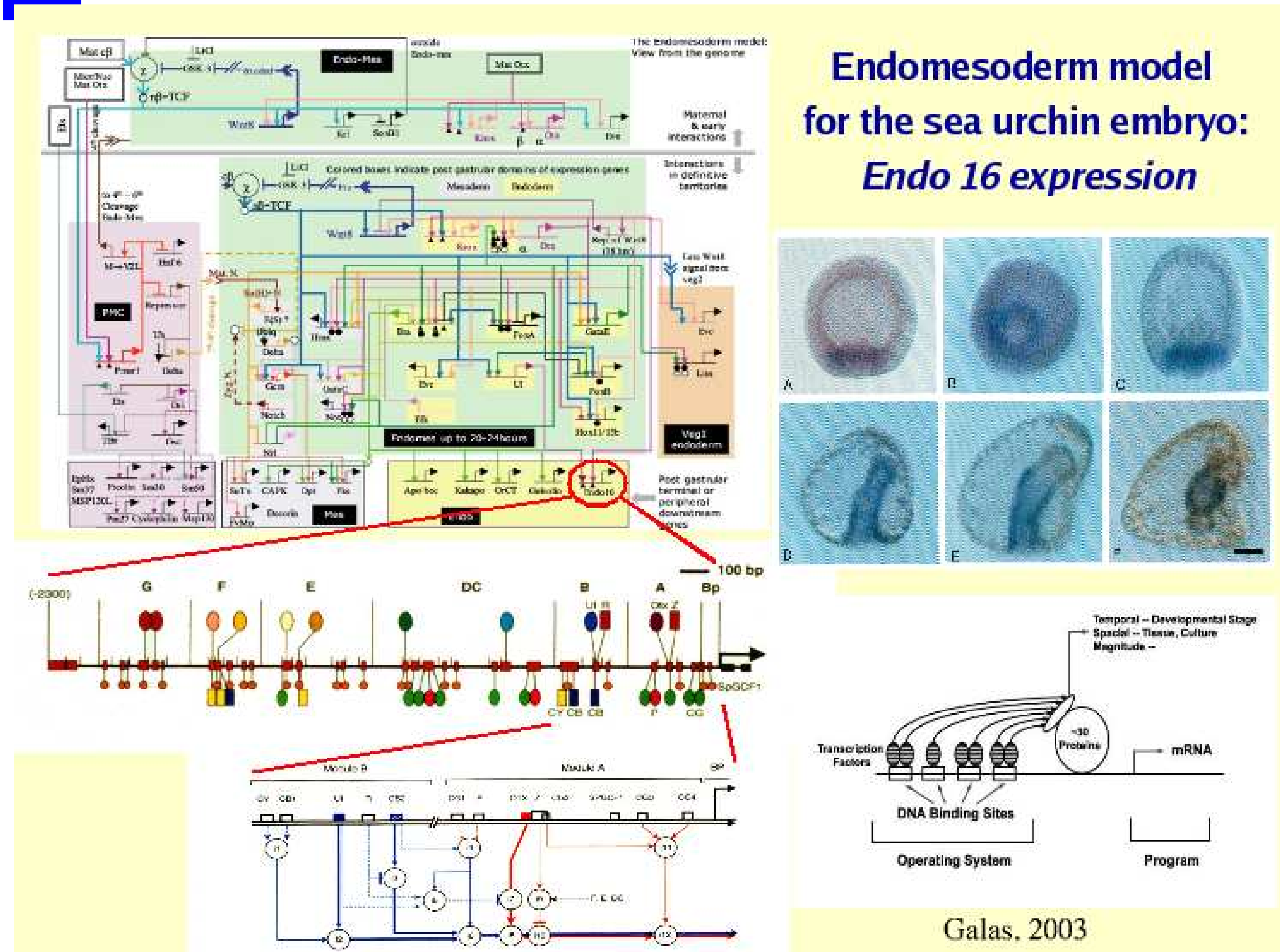


Genes $\Rightarrow$ Nodes, cis regulatory sites $\Rightarrow$ Edges.

Partial duplication



Continue



Endomesoderm model
for the sea urchin embryo:
Endo 16 expression
Galas, 2003
DNA Binding Sites
Operating System
Program
Transcription Factors
mRNA
100 bp

Partial-duplication model

Evolution of graphs

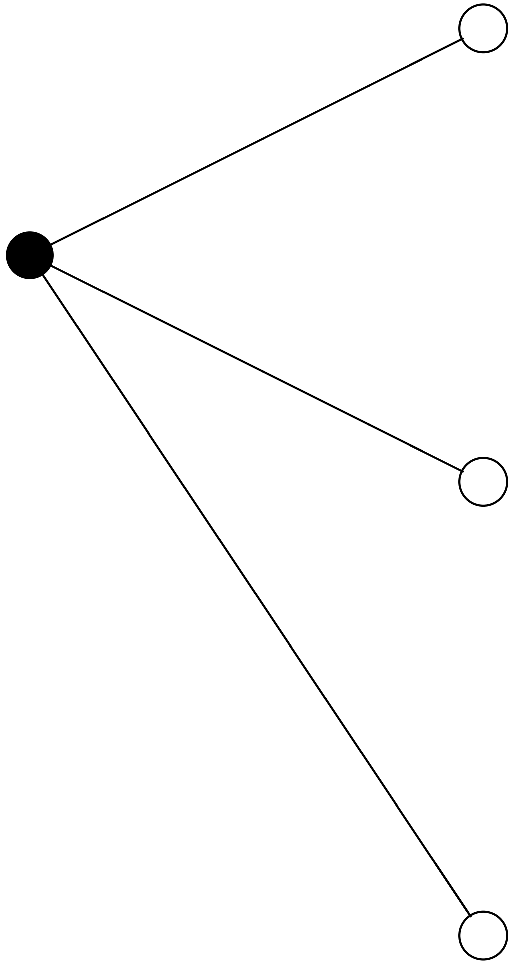
$$\cdots \subset G_{t-1} \subset G_t \subset G_{t+1} \subset \cdots$$

Construct G_{t+1} from G_t,

- Select a random vertex u of G_t uniformly.
- Add a new vertex v and the edge uv.
- For each neighbor w of u, with probability p, add an edge wv independently.

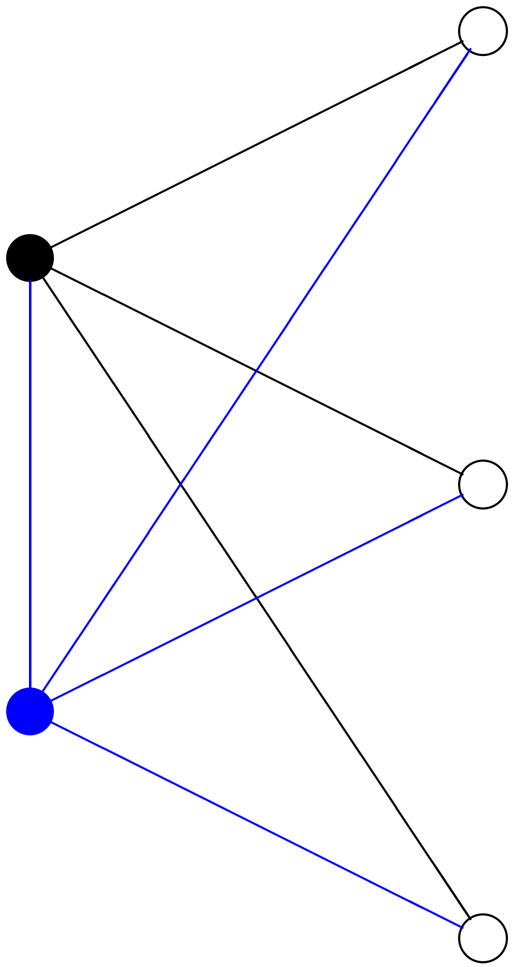
Full duplication verse partial duplication

Full duplication



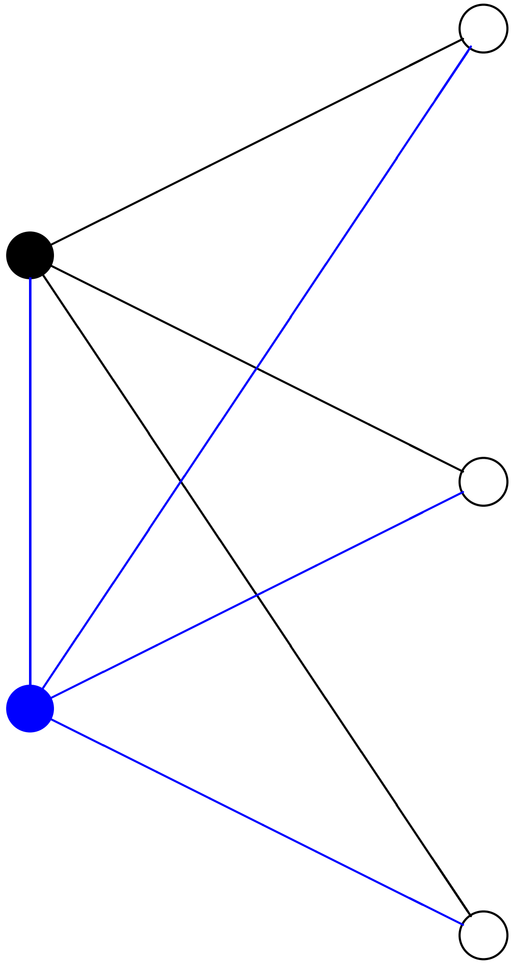
Full duplication verse partial duplication

Full duplication

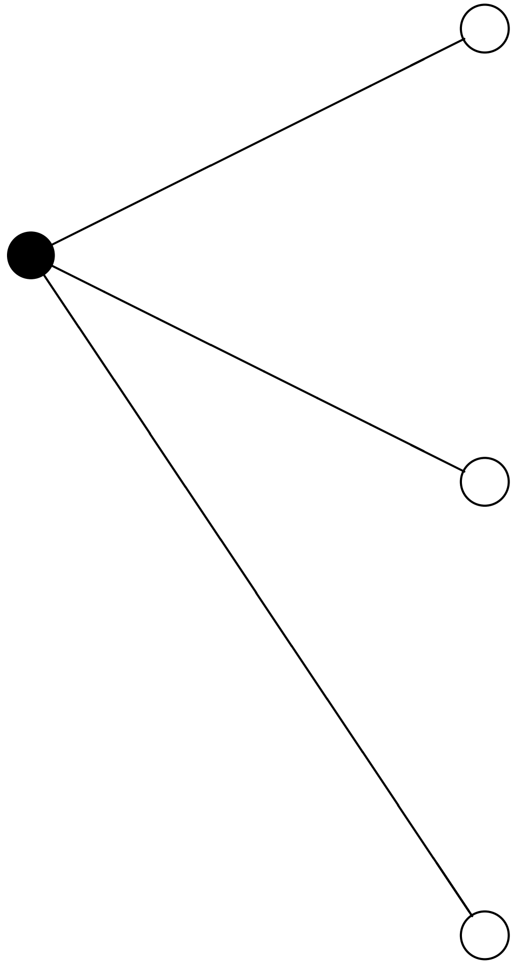


Full duplication verse partial duplication

Full duplication

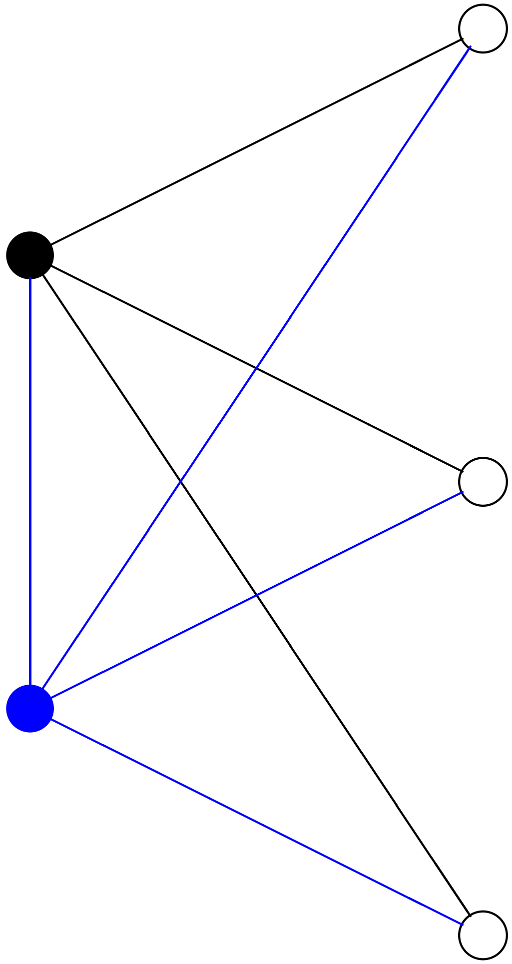


Partial duplication

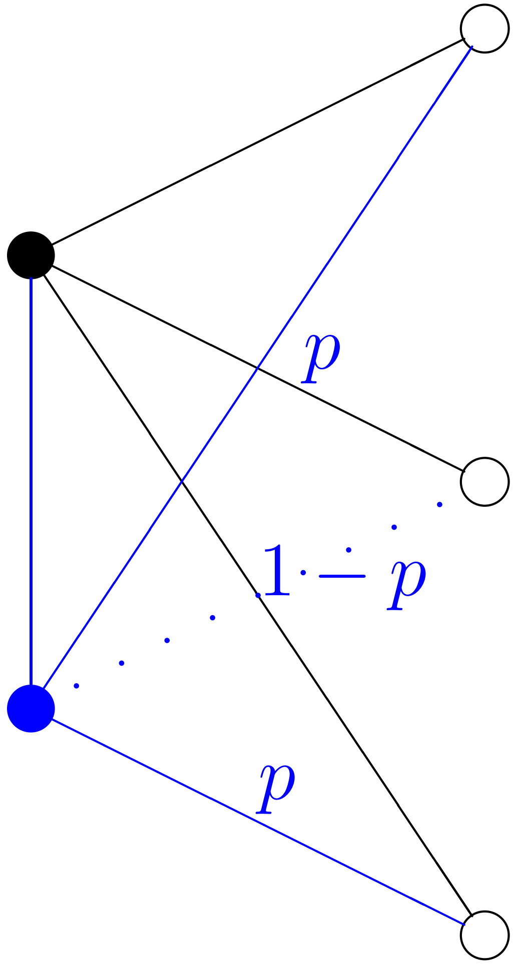


Full duplication verse partial duplication

Full duplication



Partial duplication



Recurrence formula

$f(i,t)$— the number of vertices with degree i at time t.

Type	Probability	Value
$d_u^t = i \to d_u^{t+1} = i+1$	$\frac{1+pi}{t}$	-1
$d_u^t = i-1 \to d_u^{t+1} = i$	$\frac{1+(i-1)p}{t}$	1
$d_u^t = j \to d_v^{t+1} = i$	$\binom{j}{i-1} p^{i-1}(1-p)^{j-i+1}$	1

Recurrence formula

$f(i,t)$— the number of vertices with degree i at time t.

Type	Probability	Value
$d_u^t = i \rightarrow d_u^{t+1} = i+1$	$\frac{1+pi}{t}$	-1
$d_u^t = i-1 \rightarrow d_u^{t+1} = i$	$\frac{1+(i-1)p}{t}$	1
$d_u^t = j \rightarrow d_v^{t+1} = i$	$\binom{j}{i-1} p^{i-1}(1-p)^{j-i+1}$	1

$$
\begin{aligned}
E(f(i,t+1)) &= (1-\frac{1+pi}{t})E(f(i,t)) \\
&\quad + \frac{1+(i-1)p}{t}E(f(i-1,t)) \\
&\quad + \sum_{j\geq i-1} \binom{j}{i-1} p^{i-1}(1-p)^{j-i+1}\frac{1}{t}E(f(j,t)).
\end{aligned}
$$

Heuristic analysis

Substitute $E(f(i,t))$ by $a_i t$.

$$\begin{aligned} a_i(t+1) &= (1-\frac{1+pi}{t})a_i t \\ &+\frac{1+(i-1)p}{t}a_{i-1}t \\ &+\sum_{j\geq i-1}\binom{j}{i-1}p^{i-1}(1-p)^{j-i+1}\frac{1}{t}a_j t. \end{aligned}$$

Let $t \to \infty$, we have

$$(2+ip)a_i = (1+p(i-1))a_{i-1}+\sum_{j\geq i-1} a_j \binom{j}{i-1} p^{i-1}(1-p)^{j-i+1}.$$

Recurrence formula for a_i

Replace i by $i+1$. We get

$$(2+(i+1)p)a_{i+1} = (1+pi)a_i + \sum_{j \geq i} a_j \binom{j}{i} p^i (1-p)^{j-i}.$$

Here $a_0 = 0$. (No isolated vertex.)

Lemma 1

Lemma: *For fixed k and large x, we have*

$$\frac{\binom{x-c}{k}}{\binom{x}{k}} = (1 + O(\frac{1}{x-k}))(1 - \frac{k}{x})^c.$$

Lemma 1

Lemma: *For fixed k and large x, we have*

$$\frac{\binom{x-c}{k}}{\binom{x}{k}} = (1 + O(\frac{1}{x-k}))(1 - \frac{k}{x})^c.$$

Proof:

$$\begin{aligned}
\frac{\binom{x-c}{k}}{\binom{x}{k}} &= \frac{\Gamma(x-c+1)/\Gamma(x-c+1-k)}{\Gamma(x+1)/\Gamma(x+1-k)} \\
&= (1 + O(\frac{1}{x-k}))\frac{(x+1)^{-c}}{(x+1-k)^{-c}} \\
&= (1 + O(\frac{1}{x-k}))(1 - \frac{k}{x})^c.
\end{aligned}$$

Lemma 2

Lemma: For large i, we have

$$\sum_{j \geq i} (\frac{j}{i})^{-\beta} \binom{j}{i} p^i (1-p)^{j-i} = p^{\beta-1} + O(\frac{1}{i}).$$

Lemma 2

Lemma: For large i, we have

$$\sum_{j\geq i}(\frac{j}{i})^{-\beta}\binom{j}{i}p^i(1-p)^{j-i} = p^{\beta-1} + O(\frac{1}{i}).$$

Proof:

$$\sum_{j\geq i}(\frac{j}{i})^{-\beta}\binom{j}{i}p^i(1-p)^{j-i} \;=\; \sum_{k=0}^{\infty}(1+\frac{k}{i})^{-\beta}\binom{i+k}{k}p^i(1-p)^k$$

Lemma 2

Lemma: For large i, we have

$$\sum_{j\geq i}(\frac{j}{i})^{-\beta}\binom{j}{i}p^i(1-p)^{j-i} = p^{\beta-1} + O(\frac{1}{i}).$$

Proof:

$$\begin{aligned}\sum_{j\geq i}(\frac{j}{i})^{-\beta}\binom{j}{i}p^i(1-p)^{j-i} &= \sum_{k=0}^{\infty}(1+\frac{k}{i})^{-\beta}\binom{i+k}{k}p^i(1-p)^k \\ &= (1+O(\frac{1}{i}))\sum_{k=0}^{\infty}\binom{i+k-\beta}{k}p^i(1-p)^k\end{aligned}$$

Lemma 2

$$= \quad (1+O(\frac{1}{i}))p^i \sum_{k=0}^{\infty} \binom{i+k-\beta}{k} (1-p)^k$$

Lemma 2

$$
\begin{aligned}
&= (1+O(\frac{1}{i}))p^i \sum_{k=0}^{\infty} \binom{i+k-\beta}{k} (1-p)^k \\
&= (1+O(\frac{1}{i}))p^i \sum_{k=0}^{\infty} \binom{\beta-i-1}{k} (-1)^k (1-p)^k
\end{aligned}
$$

Lemma 2

$$
\begin{aligned}
&= (1+O(\frac{1}{i}))p^i \sum_{k=0}^{\infty} \binom{i+k-\beta}{k} (1-p)^k \\
&= (1+O(\frac{1}{i}))p^i \sum_{k=0}^{\infty} \binom{\beta-i-1}{k} (-1)^k (1-p)^k \\
&= (1+O(\frac{1}{i}))p^i (1-(1-p))^{\beta-i-1} \\
&= (1+O(\frac{1}{i}))p^{\beta-1}.
\end{aligned}
$$

Heuristic analysis II

Heuristically, we let $a_i \approx C i^{-\beta}$ for large i. we have

$$\frac{a_{i+1}}{a_i} \approx (1 - \frac{\beta}{i})$$

$$\frac{a_j}{a_i} \approx (\frac{j}{i})^{-\beta}.$$

Heuristic analysis II

Heuristically, we let $a_i \approx Ci^{-\beta}$ for large i. we have

$$\frac{a_{i+1}}{a_i} \approx (1 - \frac{\beta}{i})$$

$$\frac{a_j}{a_i} \approx (\frac{j}{i})^{-\beta}.$$

$$(2+(i+1)p)(1-\frac{\beta}{i}) = (1+ip) + \sum_{j \geq i} (\frac{j}{i})^{-\beta} \binom{j}{i} p^i (1-p)^{j-i}.$$

Apply Lemma 2 and simplify it. We get

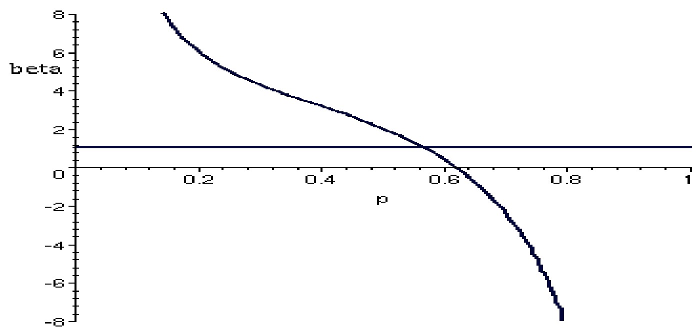
$$1 + p = p\beta + p^{\beta - 1}.$$

Results

Theorem (Chung, Dewey, Galas, Lu, 2002) *Almost surely, the partial duplication model with selection probability p generates power law graphs with the exponent β satisfying*

$$p(\beta - 1) = 1 - p^{\beta-1}.$$

In particular, if $\frac{1}{2} < p < 1$ then $\beta < 2$.



What we need to do

- Show the limit $\lim_{t\to\infty} \frac{f(t,i)}{t} = a_i$ exits.

What we need to do

- Show the limit $\lim_{t\to\infty} \frac{f(t,i)}{t} = a_i$ exits.
- Show a_i satisfies $\sum_{i=1}^{\infty} a_i = 1$ and

$$(2 + (i+1)p)a_{i+1} = (1 + pi)a_i + \sum_{j\geq i} a_j \binom{j}{i} p^i (1-p)^{j-i}.$$

What we need to do

- Show the limit $\lim_{t\to\infty} \frac{f(t,i)}{t} = a_i$ exits.
- Show a_i satisfies $\sum_{i=1}^{\infty} a_i = 1$ and

$$(2+(i+1)p)a_{i+1} = (1+pi)a_i + \sum_{j\geq i} a_j \binom{j}{i} p^i (1-p)^{j-i}.$$

- Show $a_i \approx ci^{-\beta}$ and β satisfies

$$p(\beta-1) = 1 - p^{\beta-1}.$$

A tricky way

Because the ratio $\frac{f(t,i)}{a_i t}$ oscillates a lot, it is very hard to prove $\lim_{i \to \infty} \frac{f(t,i)}{a_i t} = 1$ directly.

A tricky way

Because the ratio $\frac{f(t,i)}{a_i t}$ oscillates a lot, it is very hard to prove $\lim_{i\to\infty} \frac{f(t,i)}{a_i t} = 1$ directly.

Let $g(t,i) = \frac{1}{t}\sum_{k=1}^{i} \mathrm{E}(f(t,i))$ be the expected number of vertices with degree at most i at time t.

A tricky way

Because the ratio $\frac{f(t,i)}{a_i t}$ oscillates a lot, it is very hard to prove $\lim_{i\to\infty} \frac{f(t,i)}{a_i t} = 1$ directly.

Let $g(t,i) = \frac{1}{t}\sum_{k=1}^{i} \mathrm{E}(f(t,i))$ be the expected number of vertices with degree at most i at time t.

By definition, for fixed t, $g(t,i)$ is an increasing function of i. In particular, $g(t,i) = 1$, for all $i \geq t \geq 0$.

A tricky way

Because the ratio $\frac{f(t,i)}{a_i t}$ oscillates a lot, it is very hard to prove $\lim_{i\to\infty} \frac{f(t,i)}{a_i t} = 1$ directly.

Let $g(t,i) = \frac{1}{t}\sum_{k=1}^{i} \mathrm{E}(f(t,i))$ be the expected number of vertices with degree at most i at time t.

By definition, for fixed t, $g(t,i)$ is an increasing function of i. In particular, $g(t,i) = 1$, for all $i \geq t \geq 0$.

We will show $\lim_{t\to\infty} g(t,i) = \sum_{k=0}^{i} a_i$.

Recurrence formula for $g(t, i)$

Lemma For $i \geq 1$ and $t \geq n_0$, $g(t, i)$ satisfies the following recurrence formula.

$$\begin{aligned} g(t+1, i) &= (1 - \frac{2 + pi}{t+1}) g(t, i) + \frac{1 + pi}{t+1} g(t, i-1) \\ &+ \frac{1}{t+1} \sum_{j \geq i} g(t, j) F(j, i-1, p). \end{aligned}$$

Here $g(t, 0) = 0$ and
$F(j, i, p) = \sum_{k=0}^{i} \binom{j}{k} p^k (1-p)^{j-k} - \sum_{k=0}^{i} \binom{j+1}{k} p^k (1-p)^{j+1-k}$.

Remark

The lemma can be proved by induction on i and the following tool.

Abel summation identity

$$\sum_{j=1}^{\infty}(c_j - c_{j-1})d_j = \sum_{j=1}^{\infty} c_j(d_j - d_{j+1}) - c_0 d_1.$$

The result

Lemma: *For all* i*, the limit* $\lim_{t\to\infty} g(t,i)$ *exists. We have*

$$\lim_{t\to\infty} g(t,i) = \sum_{k=1}^{i} a_k.$$

The result

Lemma: *For all i, the limit $\lim_{t\to\infty} g(t,i)$ exists. We have*

$$\lim_{t\to\infty} g(t,i) = \sum_{k=1}^{i} a_k.$$

Proof: Let $s_i = \sum_{k=1}^{i} a_k$ and $h(t) = \sup\{\frac{g(t,i)}{s_i}\}_{i\geq 1}$. We claim that

- $h(t) \geq 1$, for all $t \geq 1$.
- $h(t)$ is a decreasing function of t.
- $\lim_{t\to\infty} h(t) = 1$.

Proof

The first item is from the following observation:

$$h(t) \geq \frac{g(t,t)}{s_t} = \frac{1}{s_t} \geq 1.$$

Proof

The first item is from the following observation:

$$h(t) \geq \frac{g(t,t)}{s_t} = \frac{1}{s_t} \geq 1.$$

To show $h(t+1) \leq h(t)$, it is sufficient to show

$$g(t+1, i) \leq h(t) s_i, \qquad \text{for all } i.$$

For $i \geq 1$, we have

Proof

$$
\begin{aligned}
g(t+1,i) &= \left(1-\frac{2+pi}{t+1}\right)g(t,i)+\frac{1+ip}{t+1}g(t,i-1) \\
&\quad +\frac{1}{t+1}\sum_{j\geq i} g(t,j)F(j,i-1,p) \\
&\leq \left(1-\frac{2+pi}{t+1}\right)h(t)s_i+\frac{1+ip}{t+1}h(t)s_{i-1} \\
&\quad +\frac{1}{t+1}\sum_{j\geq i} h(t)s_jF(j,i-1,p) \\
&= h(t)s_i.
\end{aligned}
$$

Proof

$$\begin{aligned} g(t+1,i) &= (1-\frac{2+pi}{t+1})g(t,i) + \frac{1+ip}{t+1}g(t,i-1) \\ &\quad + \frac{1}{t+1}\sum_{j\geq i} g(t,j)F(j,i-1,p) \\ &\leq (1-\frac{2+pi}{t+1})h(t)s_i + \frac{1+ip}{t+1}h(t)s_{i-1} \\ &\quad + \frac{1}{t+1}\sum_{j\geq i} h(t)s_jF(j,i-1,p) \\ &= h(t)s_i. \end{aligned}$$

Thus, $\quad h(t+1) = \sup\{\frac{g(t+1,i)}{s_i}\}_{i\geq 1} \leq h(t).$

Proof

The function $h(t)$ monotone decreases and lower-bounded by 1. Therefore, the limit $\lim_{t\to\infty} h(t)$ exists. We denote the limit by c. We have

$$c = \lim_{t\to\infty} h(t) \geq 1.$$

Proof

The function $h(t)$ monotone decreases and lower-bounded by 1. Therefore, the limit $\lim_{t\to\infty} h(t)$ exists. We denote the limit by c. We have

$$c = \lim_{t\to\infty} h(t) \geq 1.$$

If $c = 1$, then $\lim_{t\to\infty} g(i,t) = s_i$. We are done.

Proof by contradiction

Suppose, to the contrary, that $c > 1$. We note

$$\lim_{i \to \infty} s_i = \sum_{i=0}^{\infty} a_i = 1.$$

Proof by contradiction

Suppose, to the contrary, that $c > 1$. We note

$$\lim_{i \to \infty} s_i = \sum_{i=0}^{\infty} a_i = 1.$$

There exists an index i_0 so that

$$s_i \geq \frac{2}{1+c} \quad \text{for } i \geq i_0.$$

Proof by contradiction

Suppose, to the contrary, that $c > 1$. We note

$$\lim_{i\to\infty} s_i = \sum_{i=0}^{\infty} a_i = 1.$$

There exists an index i_0 so that

$$s_i \geq \frac{2}{1+c} \quad \text{for } i \geq i_0.$$

For any t and $i \geq i_0$, we have

$$\frac{g(t,i)}{s_i} \leq \frac{1}{s_i} \leq \frac{1+c}{2}.$$

Proof by contradiction

Thus, the supreme $h(t)$ is always achieved at some index $i(t) < i_0$.
Let $c_i = \sum_{j \geq i_0} \frac{a_j}{a_i} F(j, i-1, p)$. We define δ to the minimum value of $c_0, c_1, \ldots, c_{i_0}$. It is clear that $\delta > 0$.
We can prove

$$\frac{g(t,i)}{s_i} \leq h(t-1) - \frac{(c-1)\delta}{2t}.$$

Proof by contradiction

Thus, the supreme $h(t)$ is always achieved at some index $i(t) < i_0$.
Let $c_i = \sum_{j \geq i_0} \frac{a_j}{a_i} F(j, i-1, p)$. We define δ to the minimum value of $c_0, c_1, \ldots, c_{i_0}$. It is clear that $\delta > 0$.
We can prove

$$\frac{g(t,i)}{s_i} \leq h(t-1) - \frac{(c-1)\delta}{2t}.$$

Since the superium $h(t)$ is always achieved at some index $i(t) < i_0$, we have

Proof by contradiction

$$
\begin{aligned}
h(t) &= \min\{\frac{g(t,i)}{s_i}, \quad 0 \le i \le i_0\} \\
&\le h(t-1) - \frac{(c-1)\delta}{2t}.
\end{aligned}
$$

Proof by contradiction

$$\begin{aligned} h(t) &= \min\{\frac{g(t,i)}{s_i}, \quad 0 \le i \le i_0\} \\ &\le h(t-1) - \frac{(c-1)\delta}{2t}. \end{aligned}$$

In particular, we have

$$h(t) \le h(0) - \frac{(c-1)\delta}{2} \sum_{k=1}^{t} \frac{1}{k}.$$

Since the harmonic sum diverges, for t large enough, $h(t) < 0$. Contradiction to $h(t) \ge 1$, for all t. □

Generating function

Lemma: *Let* $F(z) = \sum_{i=1}^{\infty} a_i z^i$. *Then*

$$(2/z - 1)F(z) - F(pz + 1 - p) + p(1 - z)F'(z) = 0.$$

Proof: We have

$$\begin{aligned}
F(pz + 1 - p) &= \sum_{j=0}^{\infty} a_j (pz + 1 - p)^j \\
&= \sum_{j=0}^{\infty} a_j \sum_{i=0}^{j} \binom{j}{i} p^i (1-p)^{j-i} z^i \\
&= \sum_{i=0}^{\infty} z^i \sum_{j=i}^{\infty} a_j \binom{j}{i} p^i (1-p)^{j-i}.
\end{aligned}$$

Generating function

$$\begin{aligned}
F(pz+1-p) &= \sum_{i=0}^{\infty} z^i \sum_{j=i}^{\infty} a_j \binom{j}{i} p^i (1-p)^{j-i} \\
&= \sum_{i=0}^{\infty} z^i \left[(2+p(i+1))a_{i+1} - (1+pi)a_i\right] \\
&= 2\sum_{i=0}^{\infty} a_{i+1} z^i + p\sum_{i=0}^{\infty} (i+1)a_{i+1} z^i \\
&\quad -\sum_{i=0}^{\infty} a_i z^i - p\sum_{i=0}^{\infty} i a_i z^i \\
&= 2F(z)/z + pF'(z) - F(z) - pzF'(z).
\end{aligned}$$

The average degree

Rewritten as

$$\frac{F(pz+1-p)-(2/z-1)F(z)}{1-z} = pF'(z).$$

Take the limit as $z \to 1$.

$$-pF'(1) + F'(1) - 2 = pF'(1).$$

we have

$$F'(1) = \frac{2}{1-2p}.$$

If $p < \frac{1}{2}$, then the expected average degree is $\frac{2}{1-2p}$.

Maximal degree

Theorem *For the duplication model, with probability at least* $1 - te^{-c}$*, the maximum degree* Δ_t *is at most*

$$\Theta(ct^p).$$

Moreover, suppose the maximum degree of the initial graph G_{t_0} *is* Δ_{t_0}*. Then with probability at least* $1 - te^{-c}$*, we have*

$$\Delta_t \leq (\Delta_{t_0} + \frac{1}{p} + \sqrt{\frac{c^2}{9} + 2c(\Delta_{t_0} + \frac{1}{p})} + \frac{c}{3}) \left(\frac{t}{t_0}\right)^p.$$

Concentration result

For the duplication model $G(p)$, for any vertex v with degree born before time t_1, the degree $d_v(t)$ at time t satisfies:

1. *With probability at least $1 - e^{-c}$, for some $c > 0$, we have*

$$d_v(t) \leq (d_v(t_1) + \frac{1}{p} + \sqrt{\frac{16c^2}{9} + 2c(d_v(t_1) + \frac{1}{p}) } + \frac{4c}{3})(\frac{t}{t_1})^p - \frac{1}{p}$$

2. *With probability at least $1 - e^{-c}$, for some $c > 0$, we have*

$$d_v(t) \geq (d_v(t_1) + \frac{1}{p} - \sqrt{\frac{c^2}{9} + 2c(d_v(t_1) + \frac{1}{p}) } - \frac{c}{3})(\frac{t}{t_1})^p - \frac{1}{p}$$

The number of edges

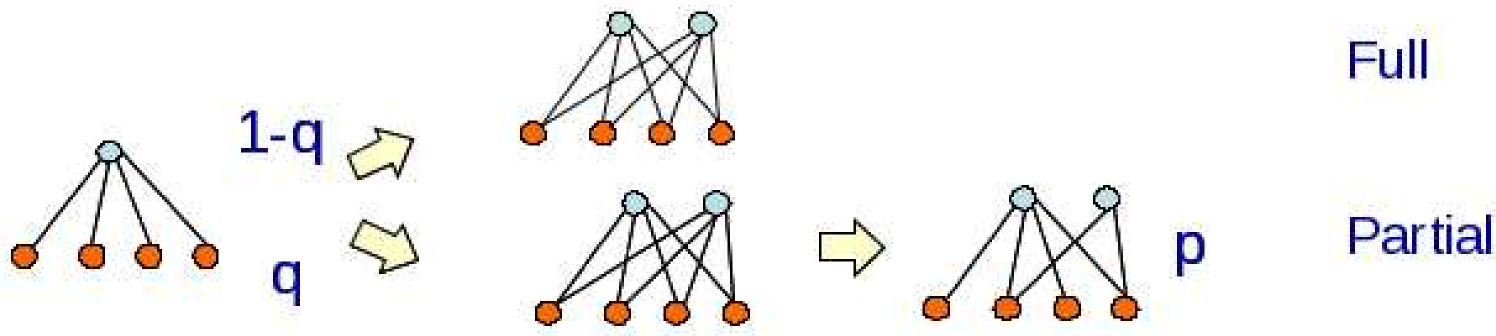
- If $p < \frac{1}{2}$, almost surely the number of edges is $\frac{t}{1-2p} + O(t^{2p} \log t)$.

The number of edges

- If $p < \frac{1}{2}$, almost surely the number of edges is $\frac{t}{1-2p} + O(t^{2p} \log t)$.
- If $p > \frac{1}{2}$, almost surely the number of edges is $O(t^{2p} \log t)$.

The mixed model

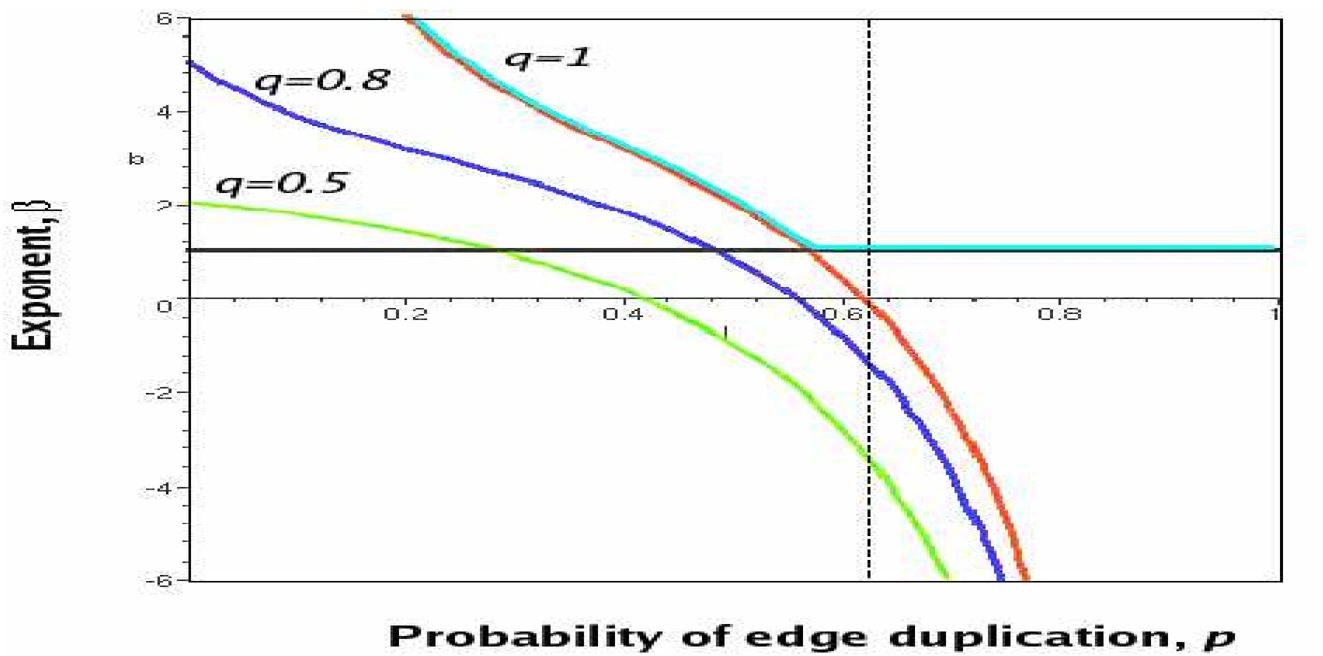
- With probability q, take a partial duplication step with the selection probability p.
- With probability $1-q$, take a full duplication step.



Result on Mixed model

Theorem 2 (Chung, Dewey, Galas, Lu, 2002)
Almost surely, the mixed model generates power law graphs with the exponent β satisfying

$$\beta(1-q) + pq(\beta - 1) = 1 - qp^{\beta-1}.$$



Summary

- The biological networks follow the power law degree distributions with exponent β in the range $(1, 2)$, which is different from non-biological networks.

Summary

- The biological networks follow the power law degree distributions with exponent β in the range $(1, 2)$, which is different from non-biological networks.
- The (partial) duplications are probably the main force shaping the biological networks.

Summary

- The biological networks follow the power law degree distributions with exponent β in the range $(1, 2)$, which is different from non-biological networks.
- The (partial) duplications are probably the main force shaping the biological networks.
- The partial duplication model is able to generate the power law graphs with exponent in the full range $(1, \infty)$.

Overview of talks

- Lecture 1: Overview and outlines
- Lecture 2: Generative models - preferential attachment schemes
- Lecture 3: Duplication models for biological networks
- Lecture 4: The rise of the giant component
- Lecture 5: The small world phenomenon: average distance and diameter
- Lecture 6: Spectrum of random graphs with given degrees