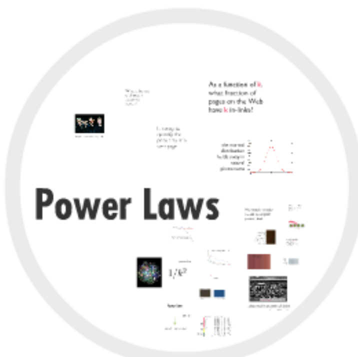
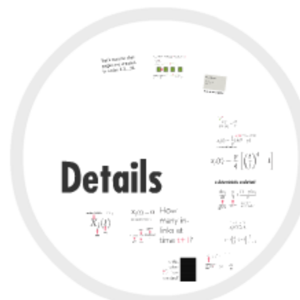
Power Laws

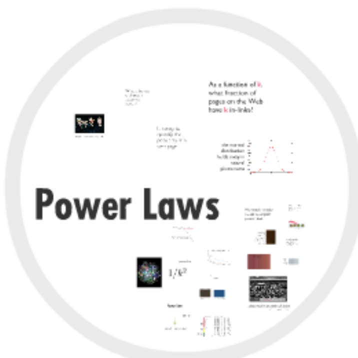
hs1@illinois.edu
hari sundaram

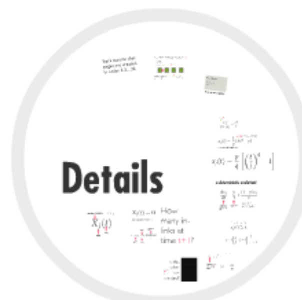




Power Laws

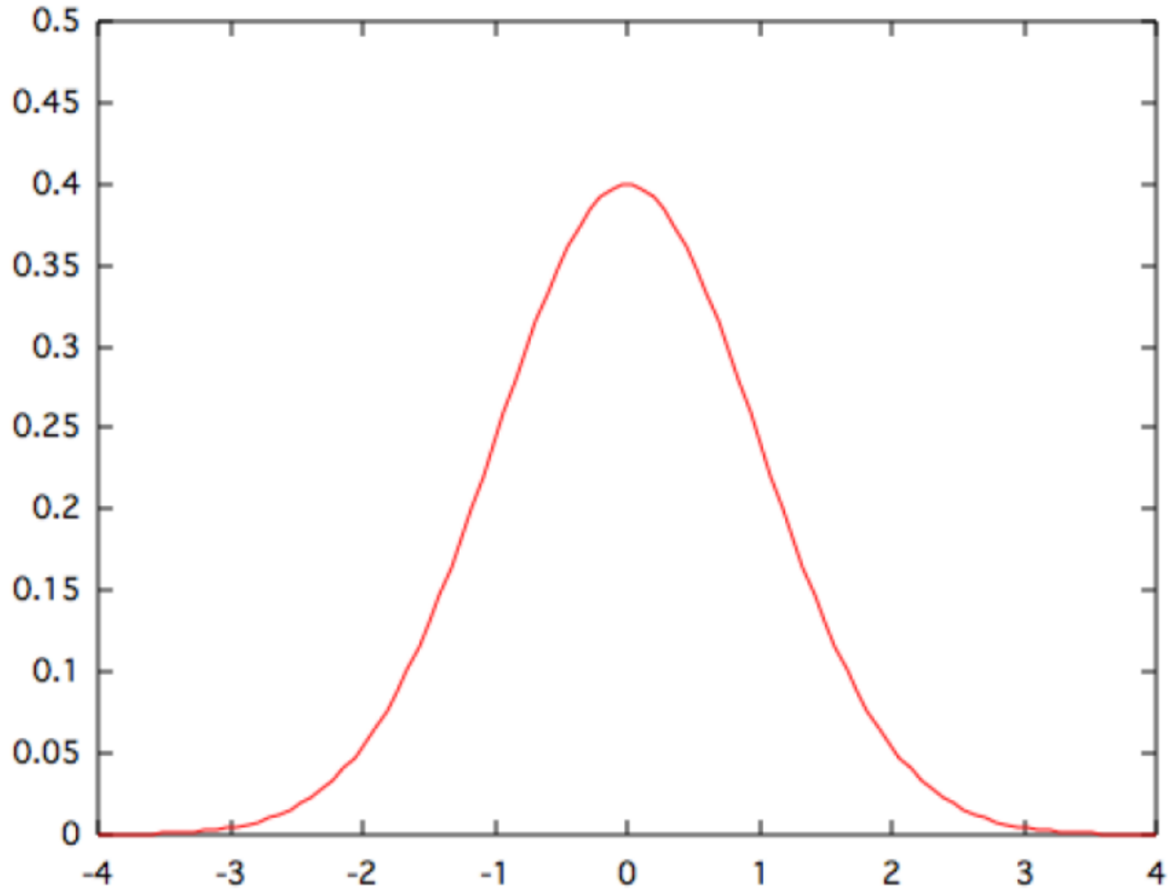
hs1@illinois.edu
hari sundaram





As a function of k, what fraction of pages on the Web have k in-links?

the normal distribution holds sway in natural phenomena



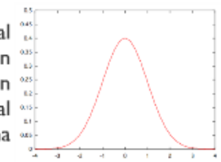
What is the role of chance in becoming popular?

Why do some bands become popular?

As a function of k, what fraction of pages on the Web have k in-links?

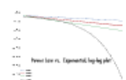
It is easy to quantify the popularity of a web page

the normal distribution holds sway in natural phenomena



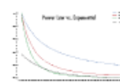
Power Laws

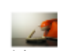
We need a simple model to explain power-laws



power law

$$1/k^2$$



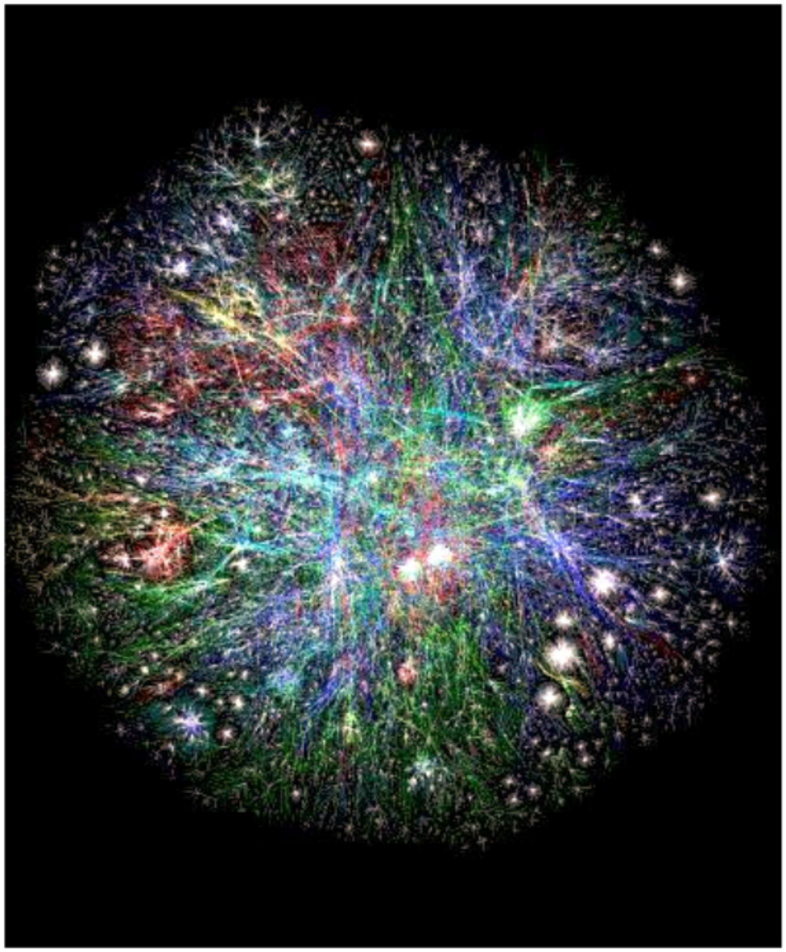


observed in growth of cities

Power Law

$$f(k) = ak^{-c}$$

$$\log f(k) = \log a - c \log k$$

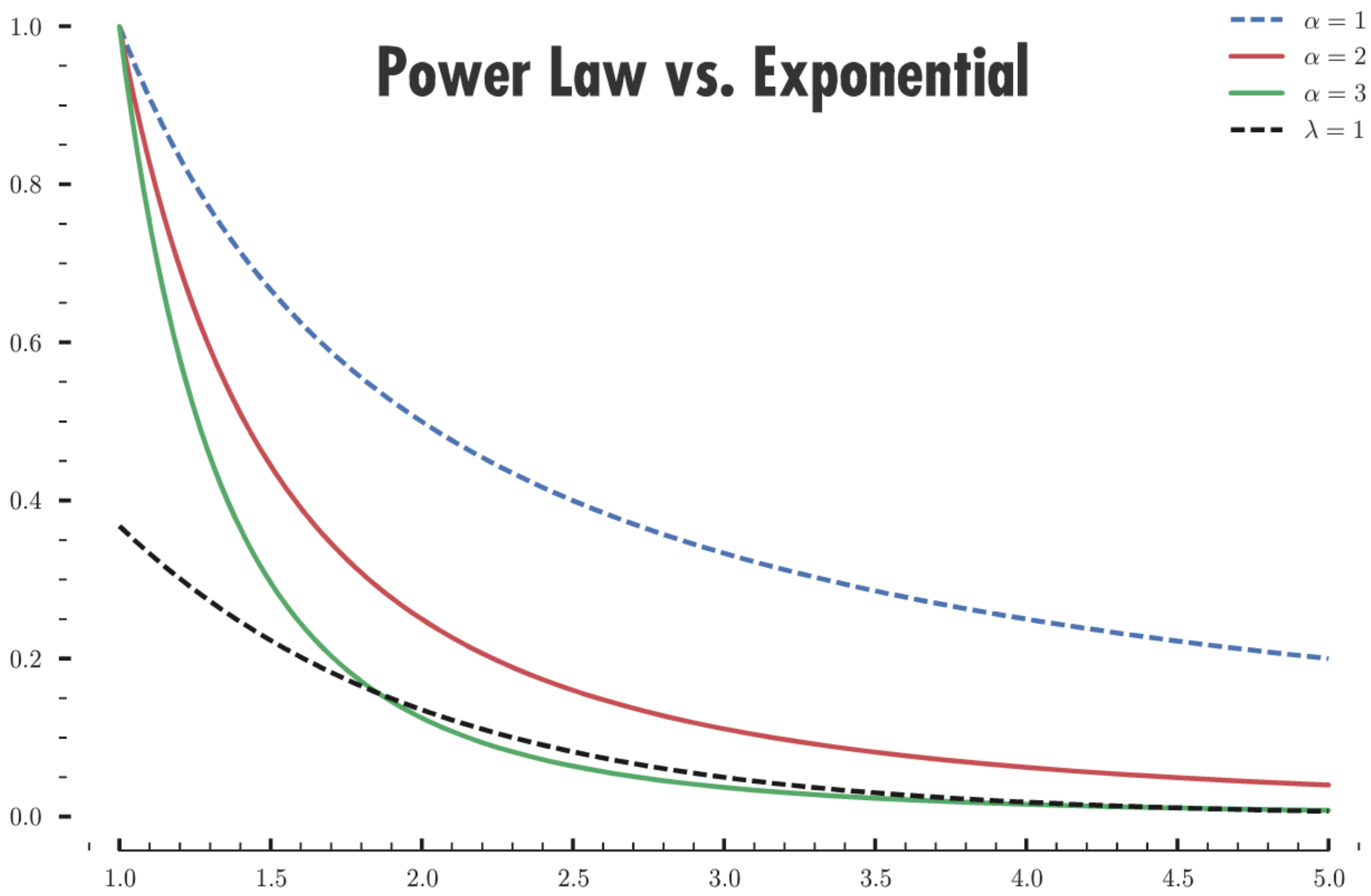


power law

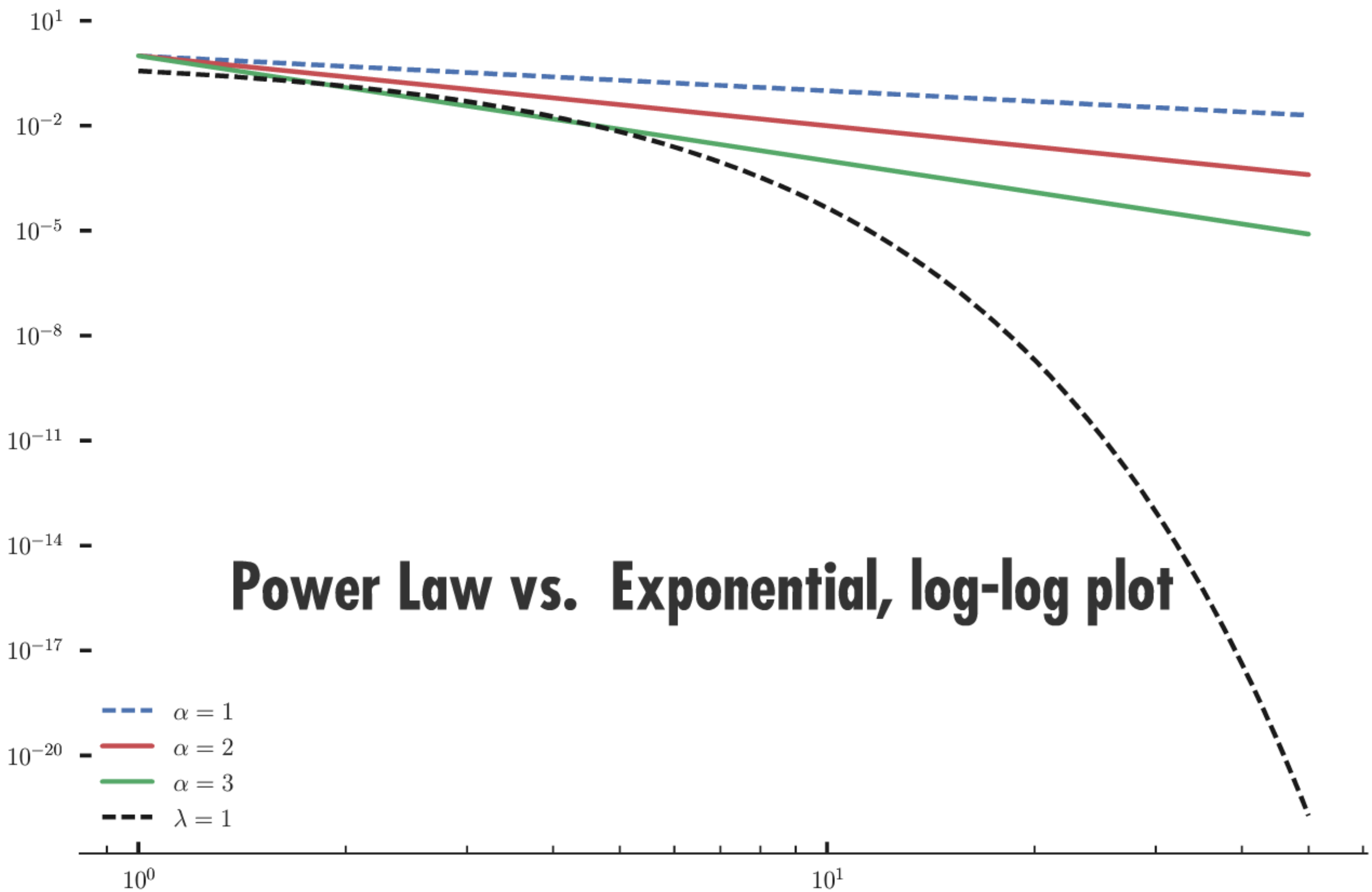
$$1/k^2$$

power law

$$/k^2$$



Power Law vs. Exponential
$\alpha = 1$
$\alpha = 2$
$\alpha = 3$
$\lambda = 1$
1.0
0.8
0.6
0.4
0.2
0.0
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0

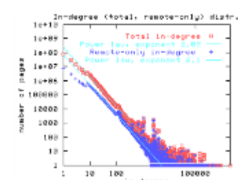

Power Law vs. Exponential, log-log plot
10^{1}
10^{-2}
10^{-5}
10^{-8}
10^{-11}
10^{-14}
10^{-17}
10^{-20}
$\alpha = 1$
$\alpha = 2$
$\alpha = 3$
$\lambda = 1$
10^{0}
10^{1}

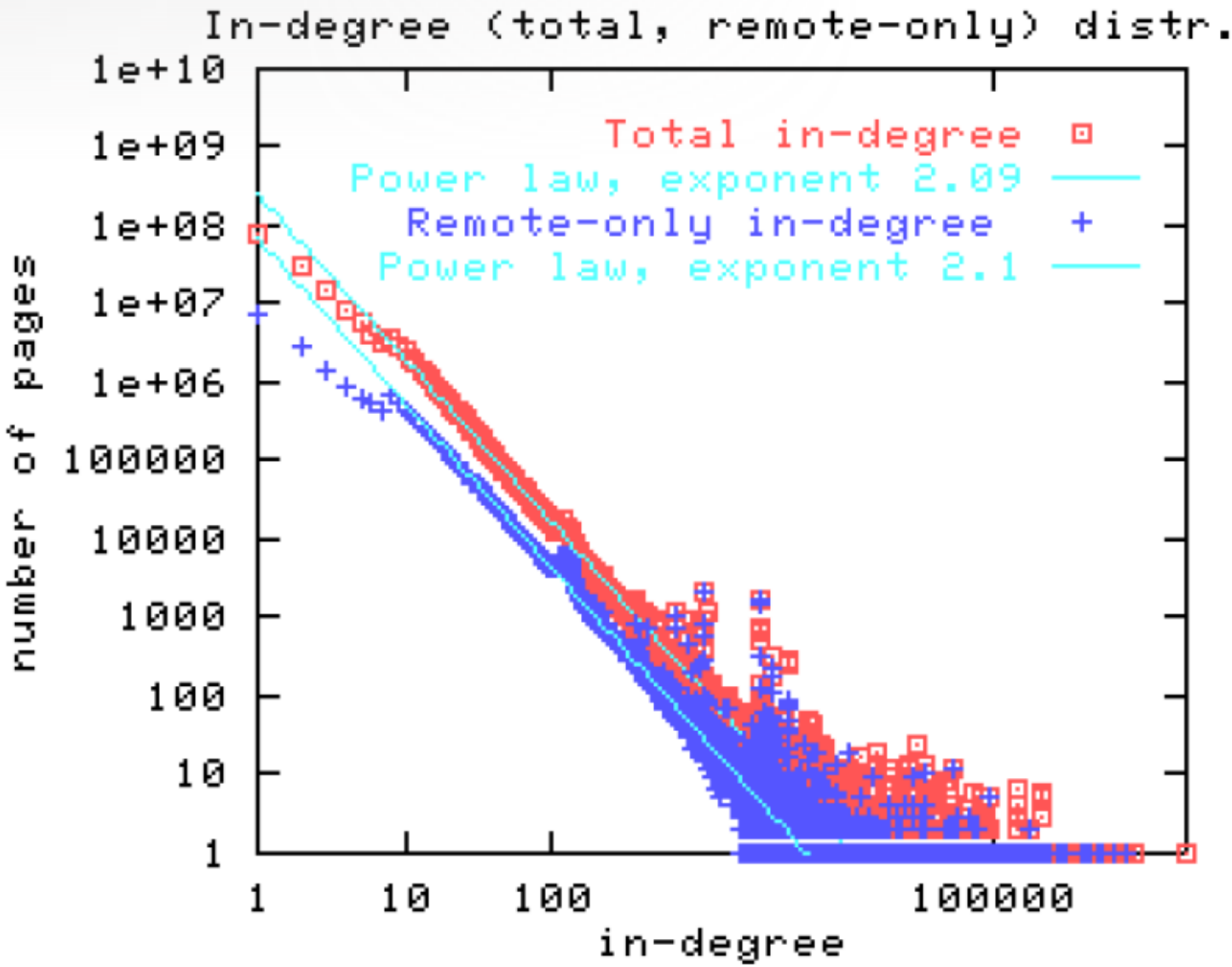
Power Law

$$f(k) = ak^{-c}$$

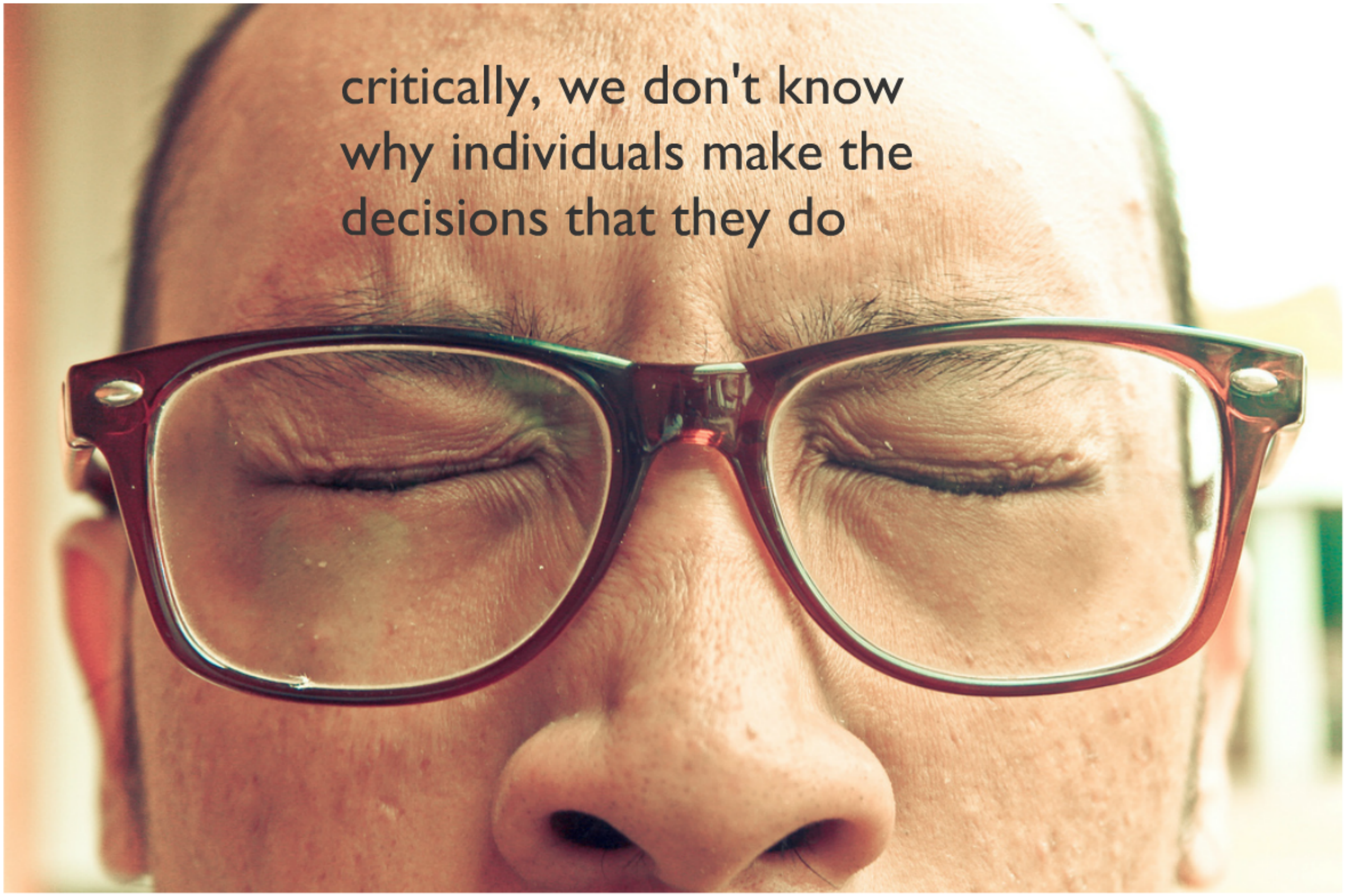
$$f(\beta k) = \beta^{-c} f(k)$$

$$\log f(k) = \log a - c \log k$$



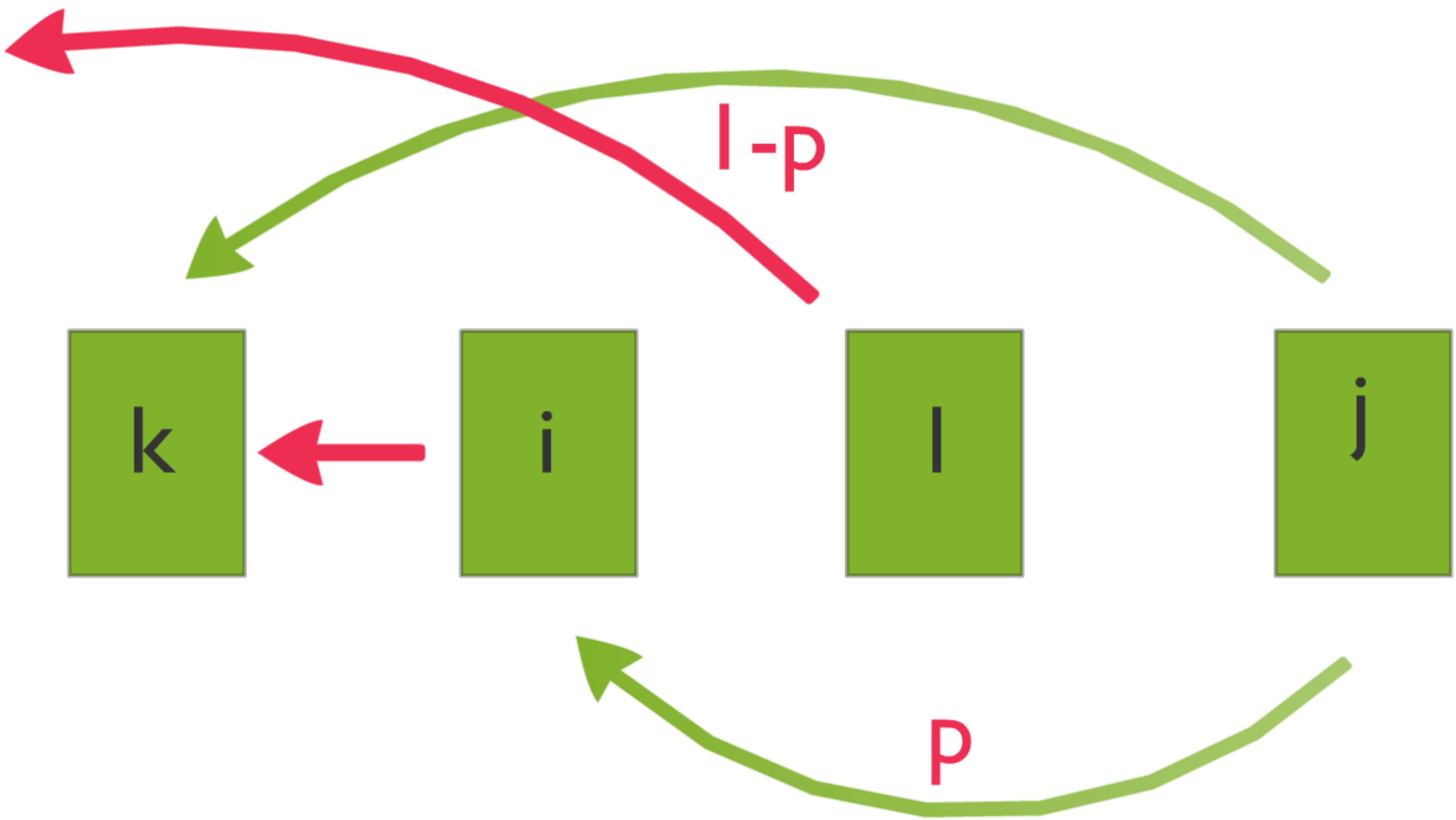

In-degree (total, remote-only) distr.
Total in-degree
Power law, exponent 2.09
Remote-only in-degree
Power law, exponent 2.1
number of pages
1e+10
1e+09
1e+08
1e+07
1e+06
100000
10000
1000
100
10
1
1
10
100
100000
in-degree

We need a simple model to explain power-laws


critically, we don't know
why individuals make the
decisions that they do

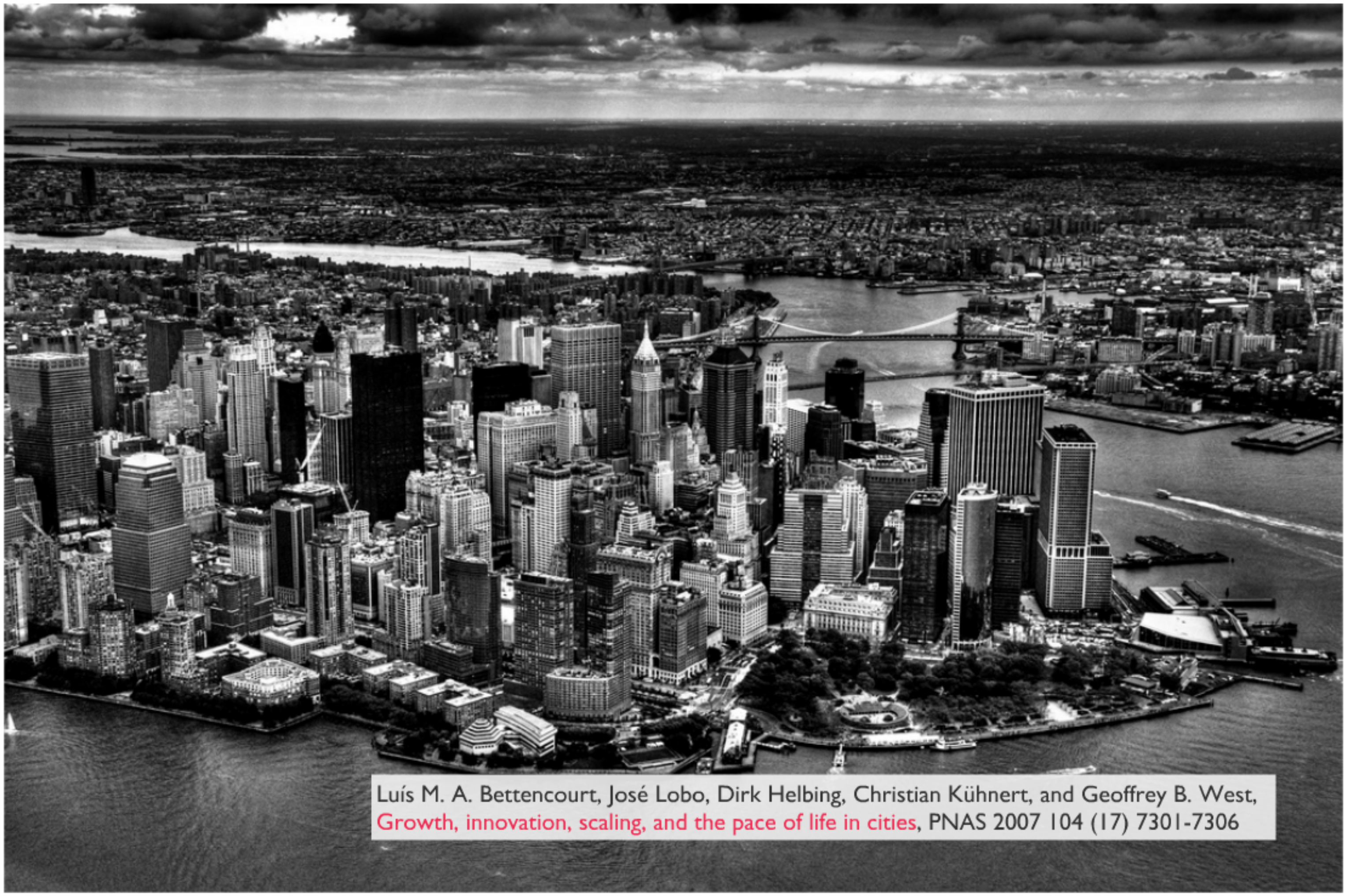
instead, we will propose a copying model: we will assume that people have a tendency to copy the decisions of people who act before them.

Let's assume that pages are created in order 1,2,...,N


l -p
k
i
l
j
p

an equivalent formulation

With probability 1-p, page j chooses a page l with probability proportional to l's current number of in-links, and creates a link to l.



Luís M. A. Bettencourt, José Lobo, Dirk Helbing, Christian Kühnert, and Geoffrey B. West, Growth, innovation, scaling, and the pace of life in cities, PNAS 2007 104 (17) 7301-7306

observed in growth of cities

metabolic rate / body mass $\frac{B}{M} \propto M^{-\frac{1}{4}}$

rate slows with size!

wealth creation, innovation as a function of size of city $\beta = 1.2$

$\beta \approx 0.8$

metabolic rate

body mass

$$\frac{B}{M} \propto M^{-\frac{1}{4}}$$

rate slows with size!

In what sense, if any, are small, medium, and large cities scaled versions of one another, thereby implying that they are manifestations of the same average idealized city?

wealth creation, innovation as a function of size of city

$$\beta = 1.2$$

power law exponent

$$\beta \approx 0.8$$

infrastructure

$$\beta \approx 0.8$$

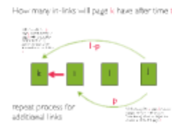
infrastructure

Table 1. Scaling exponents for urban indicators vs. city size

	Y	β	95% CI	Adj-R^2	Observations	Country–year
wealth creation	New patents	1.27	[1.25,1.29]	0.72	331	U.S. 2001
	Inventors	1.25	[1.22,1.27]	0.76	331	U.S. 2001
	Private R&D employment	1.34	[1.29,1.39]	0.92	266	U.S. 2002
	″Supercreative″ employment	1.15	[1.11,1.18]	0.89	287	U.S. 2003
	R&D establishments	1.19	[1.14,1.22]	0.77	287	U.S. 1997
	R&D employment	1.26	[1.18,1.43]	0.93	295	China 2002
	Total wages	1.12	[1.09,1.13]	0.96	361	U.S. 2002
	Total bank deposits	1.08	[1.03,1.11]	0.91	267	U.S. 1996
	GDP	1.15	[1.06,1.23]	0.96	295	China 2002
	GDP	1.26	[1.09,1.46]	0.64	196	EU 1999–2003
	GDP	1.13	[1.03,1.23]	0.94	37	Germany 2003
	Total electrical consumption	1.07	[1.03,1.11]	0.88	392	Germany 2002
	New AIDS cases	1.23	[1.18,1.29]	0.76	93	U.S. 2002–2003
	Serious crimes	1.16	[1.11, 1.18]	0.89	287	U.S. 2003
human needs	Total housing	1.00	[0.99,1.01]	0.99	316	U.S. 1990
	Total employment	1.01	[0.99,1.02]	0.98	331	U.S. 2001
	Household electrical consumption	1.00	[0.94,1.06]	0.88	377	Germany 2002
	Household electrical consumption	1.05	[0.89,1.22]	0.91	295	China 2002
	Household water consumption	1.01	[0.89,1.11]	0.96	295	China 2002
infrastructure	Gasoline stations	0.77	[0.74,0.81]	0.93	318	U.S. 2001
	Gasoline sales	0.79	[0.73,0.80]	0.94	318	U.S. 2001
	Length of electrical cables	0.87	[0.82,0.92]	0.75	380	Germany 2002
	Road surface	0.83	[0.74,0.92]	0.87	29	Germany 2002

Data sources are shown in *SI Text*. CI, confidence interval; Adj-R^2, adjusted R^2; GDP, gross domestic product.

Let's assume that pages are created in order 1,2,...,N



How can one simplify?

Details

number of in-links $t \geq j$

$$X_j(t)$$

page time

$$X_j(j) = 0$$

since a page starts with no in-links

$$X_j(t+1) = \frac{p}{t} + \frac{(1-p)X_j(t)}{t}$$

How many in-links at time t+1?

$$\frac{dx_j}{p + qx_j} = \frac{dt}{t}$$

constant

$$x_j(t) = \frac{1}{q}(At^q - p)$$

solution to the differential equation

$$x_j(t) = \frac{p}{q}\left[\left(\frac{t}{j}\right)^q - 1\right]$$

a deterministic evolution!

$$\frac{dx_j}{dt} = \frac{p}{t} + \frac{(1-p)x_j}{t}$$

change to the number of in-links

mean field approximation

for what values of j is

$$x_j(t) \geq k$$

$$j \leq t\left[\frac{q}{p}\cdot k + 1\right]^{-\frac{1}{q}}$$

is this what we wanted?

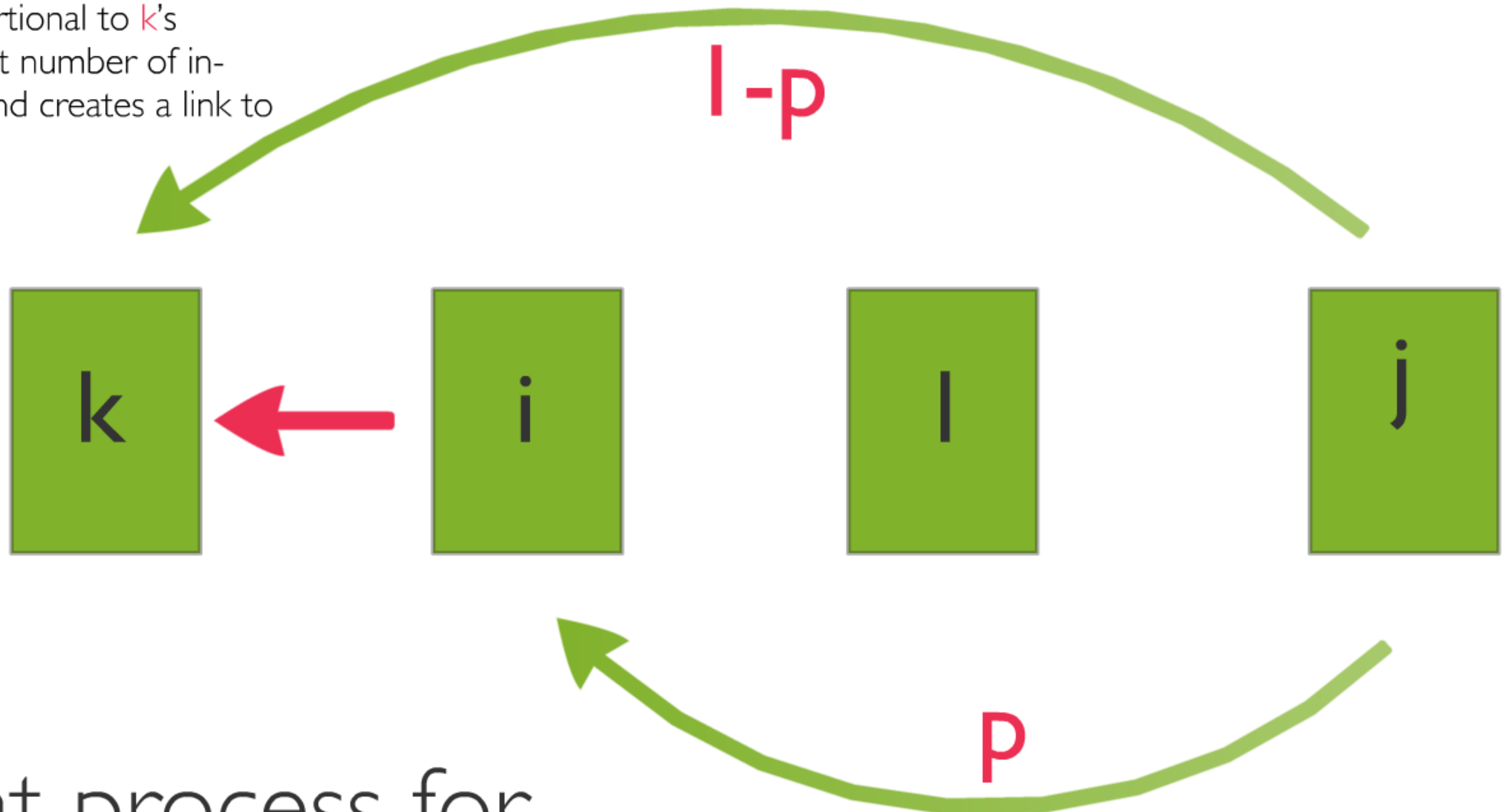
$$f(k) = \frac{1}{p}\left[\frac{1-p}{p}\cdot k + 1\right]^{-\left(1+\frac{1}{1-p}\right)}$$

fraction of nodes with exactly k inlinks

$$f(k) = -\frac{dF}{dk}$$

Let's assume that pages are created in order 1,2,...,N

With probability 1-p, page j instead chooses a page k with probability proportional to k's current number of in-links and creates a link to k.

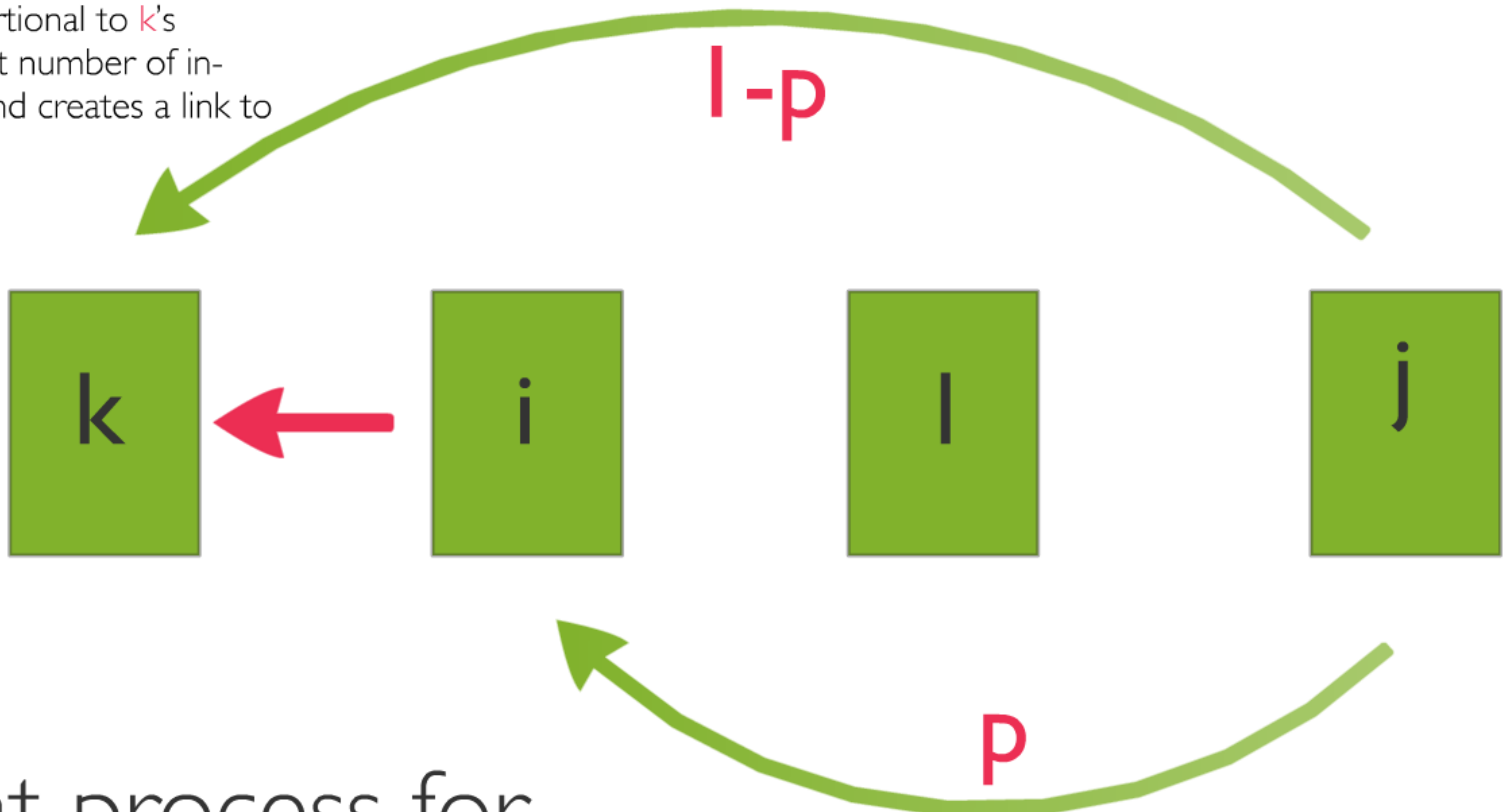


With probability p, page j chooses a page i uniformly at random from among all earlier pages and creates a link to this page i.

repeat process for additional links

How many in-links will page k have after time t?

With probability 1-p, page j instead chooses a page k with probability proportional to k's current number of in-links and creates a link to k.



With probability p, page j chooses a page i uniformly at random from among all earlier pages and creates a link to this page i.

repeat process for additional links

number of in-links

$t \geq j$

$$X_j(t)$$

page

time

$$X_j(j) = 0$$

since a page starts with no in links

choose a page proportional to its in-links with probability 1-p

choose a page at random with p

$$X_j(t+1) = \frac{p}{t} + \frac{(1-p)X_j(t)}{t}$$

change to the number of in links

there are t pages at time t+1

$$X_j(j) = 0$$

since a page starts with no in links

choose a page at random with p

choose a page proportional to its in-links with probability 1-p

$$X_j(t+1) = \frac{p}{t} + \frac{(1-p)X_j(t)}{t}$$

change to the number of in links

there are t pages at time t+1

How many in-links at time t+1?


It's complicated!
stochastic processes always appear to be

How can one simplify?

$q = 1 - p$

$$\frac{dx_j}{dt} = \frac{p}{t} + \frac{qx_j}{t}$$

A deterministic, continuous time approximation

q=1-p

$$\frac{\mathrm{d}x_j}{\mathrm{d}t} = \frac{p}{t} + \frac{qx_j}{t}$$

A deterministic, continuous time approximation

$$\frac{\mathrm{d}x_j}{p + qx_j} = \frac{\mathrm{d}t}{t}$$

q=1-p

$$\frac{\mathrm{d}x_j}{\mathrm{d}t} = \frac{p}{t} + \frac{qx_j}{t}$$

$$\frac{\mathrm{d}x_j}{p + qx_j} = \frac{\mathrm{d}t}{t}$$

constant Use $x_j(j) = 0$ to determine A

$$x_j(t) = \frac{1}{q}(At^q - p)$$

solution to the differential equation

q=1-p

$$\frac{dx_j}{dt} = \frac{p}{t} + \frac{qx_j}{t}$$

$$\frac{dx_j}{p + qx_j} = \frac{dt}{t}$$

constant

Use $x_j(j) = 0$ to determine A

$$x_j(t) = \frac{1}{q}(At^q - p)$$

solution to the differential equation

$$x_j(t) = \frac{p}{q}\left[\left(\frac{t}{j}\right)^q - 1\right]$$

for what values of j is

$$x_j(t) \geq k$$

for what values of j is

$$x_j(t) \geq k$$

$$j \leq t\left[\frac{q}{p} \cdot k + 1\right]^{-\frac{1}{q}}$$

$$x_j(t) = \frac{p}{q}\left[\left(\frac{t}{j}\right)^q - 1\right]$$

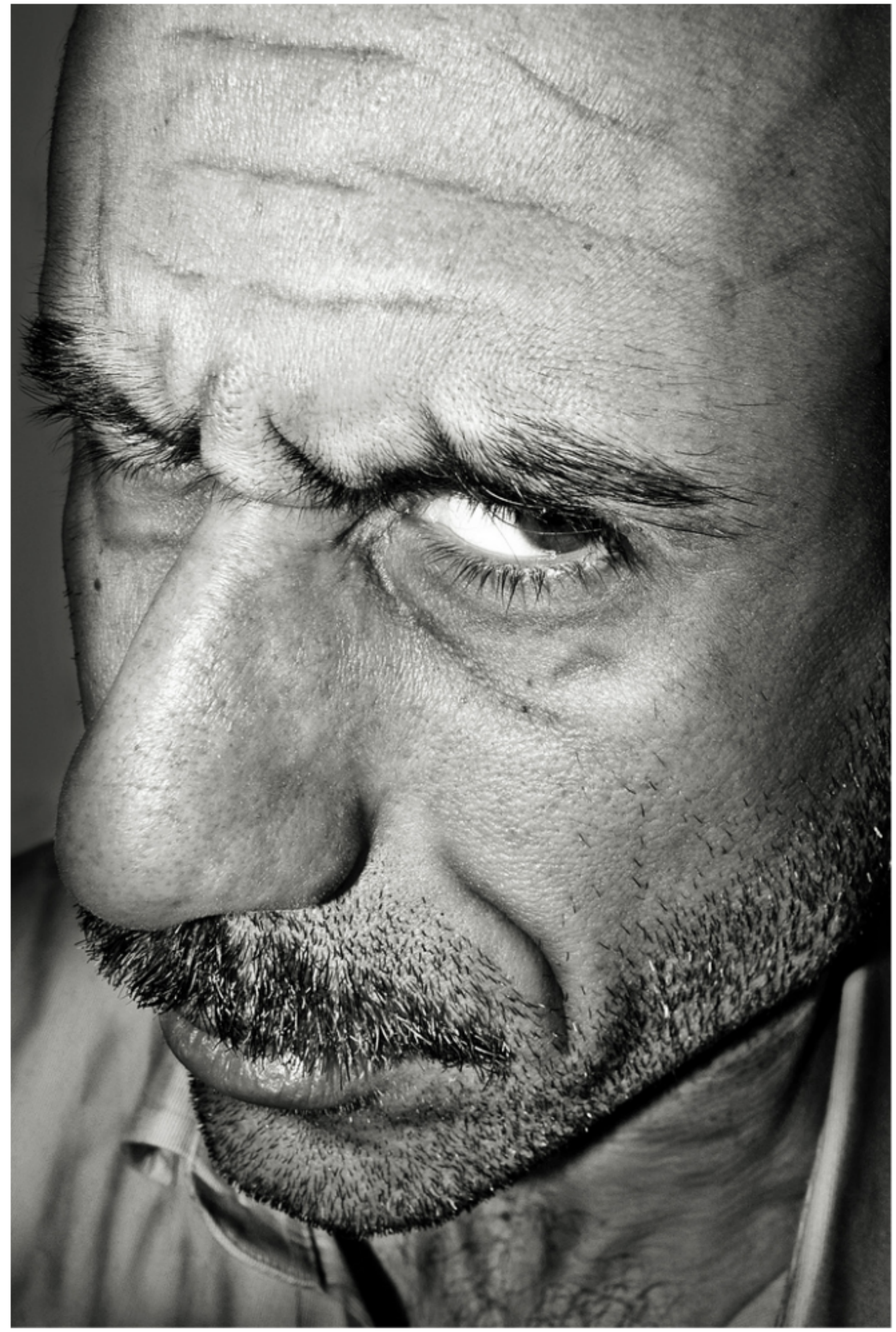
fraction of nodes having at least k links at time t

$$\frac{j}{t} \leq \left[\frac{q}{p} \cdot k + 1\right]^{-\frac{1}{q}}$$

is this what we wanted?

fraction of nodes having at least **k** links at time **t**

$$\frac{j}{t} \leq \left[\frac{q}{p} \cdot k + 1\right]^{-\frac{1}{q}}$$



$$f(k) = \frac{1}{p}\left[\frac{1-p}{p}\cdot k + 1\right]^{-\left(1+\frac{1}{1-p}\right)}$$

fraction of nodes with exactly k inlinks

$$f(k) = -\frac{\mathrm{d}F}{\mathrm{d}k}$$