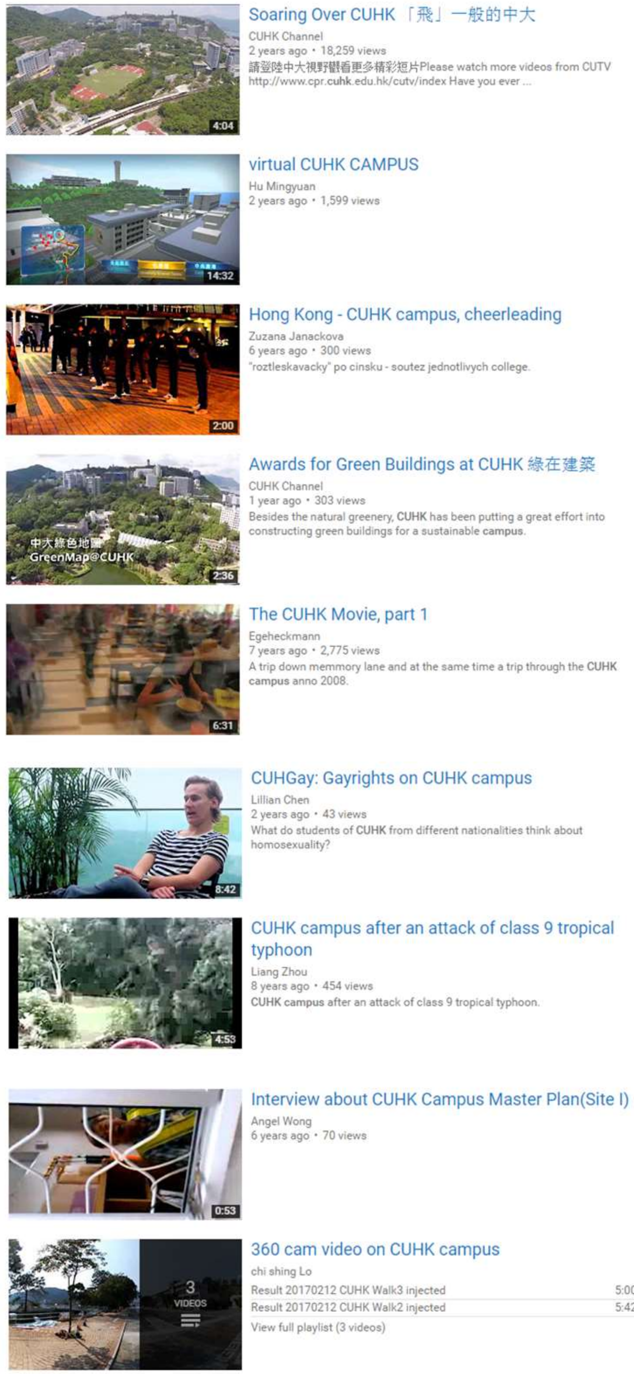
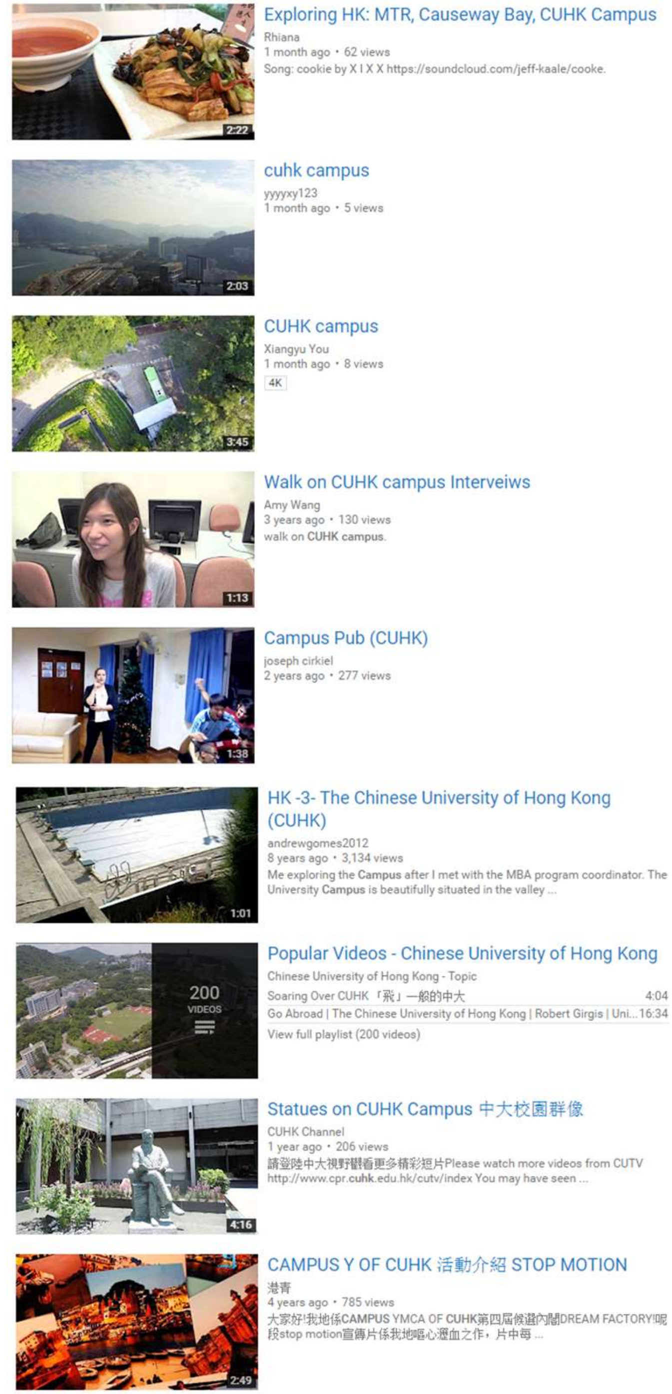
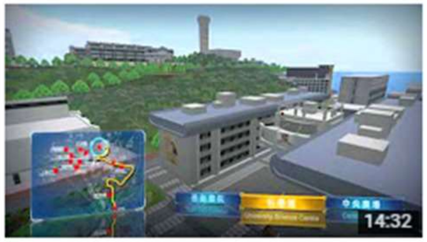
Power Laws and Rich-Get-Richer Models


Soaring Over CUHK「飛」一般的中大
CUHK Channel
2 years ago • 18,259 views
請登陸中大視野觀看更多精彩短片Please watch more videos from CUTV http://www.cpr.cuhk.edu.hk/cutv/index Have you ever ...
4:04
virtual CUHK CAMPUS
Hu Mingyuan
2 years ago • 1,599 views
14:32
Hong Kong - CUHK campus, cheerleading
Zuzana Janackova
6 years ago • 300 views
"roztleskavacky" po cinsku - soutez jednotlivych college.
2:00
Awards for Green Buildings at CUHK 綠在建築
CUHK Channel
1 year ago • 303 views
Besides the natural greenery, CUHK has been putting a great effort into constructing green buildings for a sustainable campus.
中大綠色地圖
GreenMap@CUHK
2:36
The CUHK Movie, part 1
Egeheckmann
7 years ago • 2,775 views
A trip down memmory lane and at the same time a trip through the CUHK campus anno 2008.
6:31
CUHGay: Gayrights on CUHK campus
Lillian Chen
2 years ago • 43 views
What do students of CUHK from different nationalities think about homosexuality?
8:42
CUHK campus after an attack of class 9 tropical typhoon
Liang Zhou
8 years ago • 454 views
CUHK campus after an attack of class 9 tropical typhoon.
4:53
Interview about CUHK Campus Master Plan(Site I)
Angel Wong
6 years ago • 70 views
0:53
360 cam video on CUHK campus
chi shing Lo
Result 20170212 CUHK Walk3 injected 5:00
Result 20170212 CUHK Walk2 injected 5:42
View full playlist (3 videos)
3
VIDEOS


Exploring HK: MTR, Causeway Bay, CUHK Campus
Rhiana
1 month ago • 62 views
Song: cookie by X I X X https://soundcloud.com/jeff-kaale/cooke.
2:22
cuhk campus
yyyyxy123
1 month ago • 5 views
2:03
CUHK campus
Xiangyu You
1 month ago • 8 views
4K
3:45
Walk on CUHK campus Interveiws
Amy Wang
3 years ago • 130 views
walk on CUHK campus.
1:13
Campus Pub (CUHK)
joseph cirkiel
2 years ago • 277 views
1:38
HK -3- The Chinese University of Hong Kong (CUHK)
andrewgomes2012
8 years ago • 3,134 views
Me exploring the Campus after I met with the MBA program coordinator. The University Campus is beautifully situated in the valley ...
1:01
Popular Videos - Chinese University of Hong Kong
Chinese University of Hong Kong - Topic
Soaring Over CUHK「飛」一般的中大 4:04
Go Abroad | The Chinese University of Hong Kong | Robert Girgis | Uni... 16:34
View full playlist (200 videos)
200
VIDEOS
Statues on CUHK Campus 中大校園群像
CUHK Channel
1 year ago • 206 views
請登陸中大視野觀看更多精彩短片Please watch more videos from CUTV http://www.cpr.cuhk.edu.hk/cutv/index You may have seen ...
4:16
CAMPUS Y OF CUHK 活動介紹 STOP MOTION
港青
4 years ago • 785 views
大家好!我地係CAMPUS YMCA OF CUHK第四屆候選內閣DREAM FACTORY!呢段stop motion宣傳片係我地嘔心瀝血之作，片中每 ...
2:49

Soaring Over CUHK 「飛」一般的中大

CUHK Channel
2 years ago · 18,259 views
請登陸中大視野觀看更多精彩短片Please watch more videos from CUTV http://www.cpr.cuhk.edu.hk/cutv/index Have you ever ...



virtual CUHK CAMPUS

Hu Mingyuan
2 years ago · 1,599 views

Hong Kong - CUHK campus, cheerleading

Zuzana Janackova
6 years ago · 300 views
"roztleskavacky" po cinsku - soutez jednotlivych college.

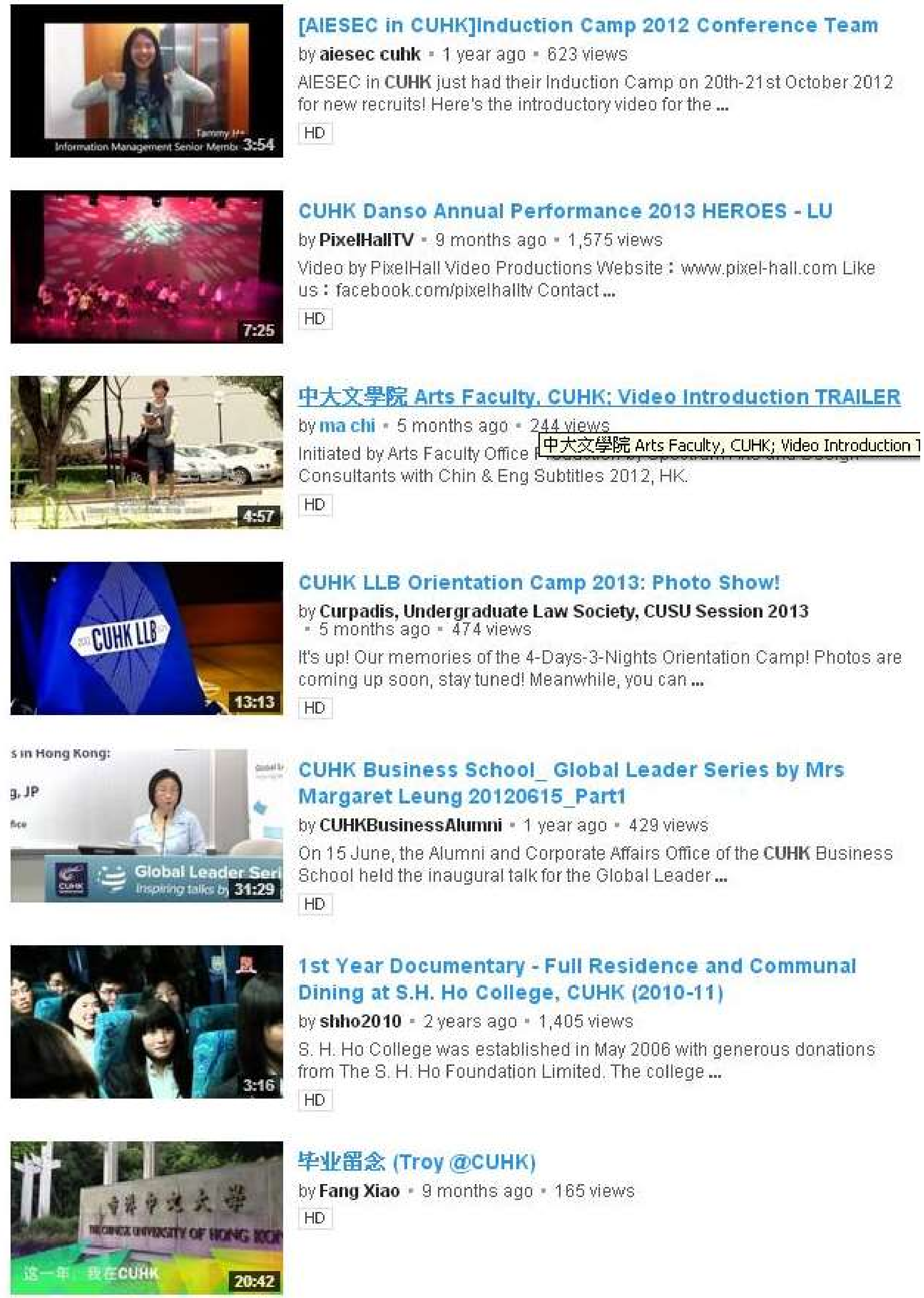
Awards for Green Buildings at CUHK 綠在建築

CUHK Channel
1 year ago · 303 views
Besides the natural greenery, **CUHK** has been putting a great effort into constructing green buildings for a sustainable **campus**.

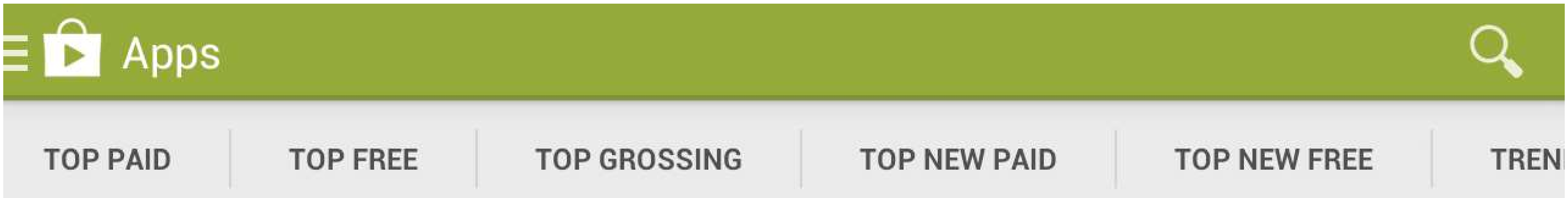
The CUHK Movie, part 1

Egeheckmann
7 years ago · 2,775 views
A trip down memmory lane and at the same time a trip through the **CUHK** **campus** anno 2008.


Filters
About 19,600 results
CUHK 2013 畢業微電影- 從這以後
by XING Agassi • 9 months ago • 4,842 views
HD
29:32
CUHK 七BA莊文廣會師
by Elim Li • 1 year ago • 3,991 views
HD
9:15
About CUHK
by CUHKchannel
14 videos
Understanding CUHK 認識中大 (廣東話) 8:09
Understanding CUHK 認識中大 (英語) 9:41
View full playlist (14 videos)
CUHK Permanent Dorm
by Jess Medeiros • 2 years ago • 2,065 views
My room in Student Hostel II High Block... Home sweet Hong Kong! Benny Hill Theme Song.
HD
2:48
校長都想知答案！ (Unique One - CUHK)
by CUCCC1972 • 1 year ago • 5,902 views
Facebook Page : Unique One - CUHK
http://www.facebook.com/YouNeedOne.
HD
0:38
Step Foot into CUHK with ISA!
by ISACUHK • 7 months ago • 178 views
The International-Student Orientation Camp will be held in CUHK from August 21st to August 25th. Join the ISA family and step ...
HD
3:57


[AIESEC in CUHK]Induction Camp 2012 Conference Team
by aiesec cuhk • 1 year ago • 623 views
AIESEC in CUHK just had their Induction Camp on 20th-21st October 2012 for new recruits! Here's the introductory video for the ...
HD
3:54
CUHK Danso Annual Performance 2013 HEROES - LU
by PixelHallTV • 9 months ago • 1,575 views
Video by PixelHall Video Productions Website：www.pixel-hall.com Like us：facebook.com/pixelhalltv Contact ...
HD
7:25
中大文學院 Arts Faculty, CUHK; Video Introduction TRAILER
by ma chi • 5 months ago • 244 views
中大文學院 Arts Faculty, CUHK; Video Introduction T
Initiated by Arts Faculty Office
Consultants with Chin & Eng Subtitles 2012, HK.
HD
4:57
CUHK LLB Orientation Camp 2013: Photo Show!
by Curpadis, Undergraduate Law Society, CUSU Session 2013
• 5 months ago • 474 views
It's up! Our memories of the 4-Days-3-Nights Orientation Camp! Photos are coming up soon, stay tuned! Meanwhile, you can ...
HD
CUHK LLB
13:13
CUHK Business School_ Global Leader Series by Mrs Margaret Leung 20120615_Part1
by CUHKBusinessAlumni • 1 year ago • 429 views
On 15 June, the Alumni and Corporate Affairs Office of the CUHK Business School held the inaugural talk for the Global Leader ...
HD
Global Leader Seri
31:29
1st Year Documentary - Full Residence and Communal Dining at S.H. Ho College, CUHK (2010-11)
by shho2010 • 2 years ago • 1,405 views
S. H. Ho College was established in May 2006 with generous donations from The S. H. Ho Foundation Limited. The college ...
HD
3:16
毕业留念 (Troy @CUHK)
by Fang Xiao • 9 months ago • 165 views
HD
20:42

Popularity

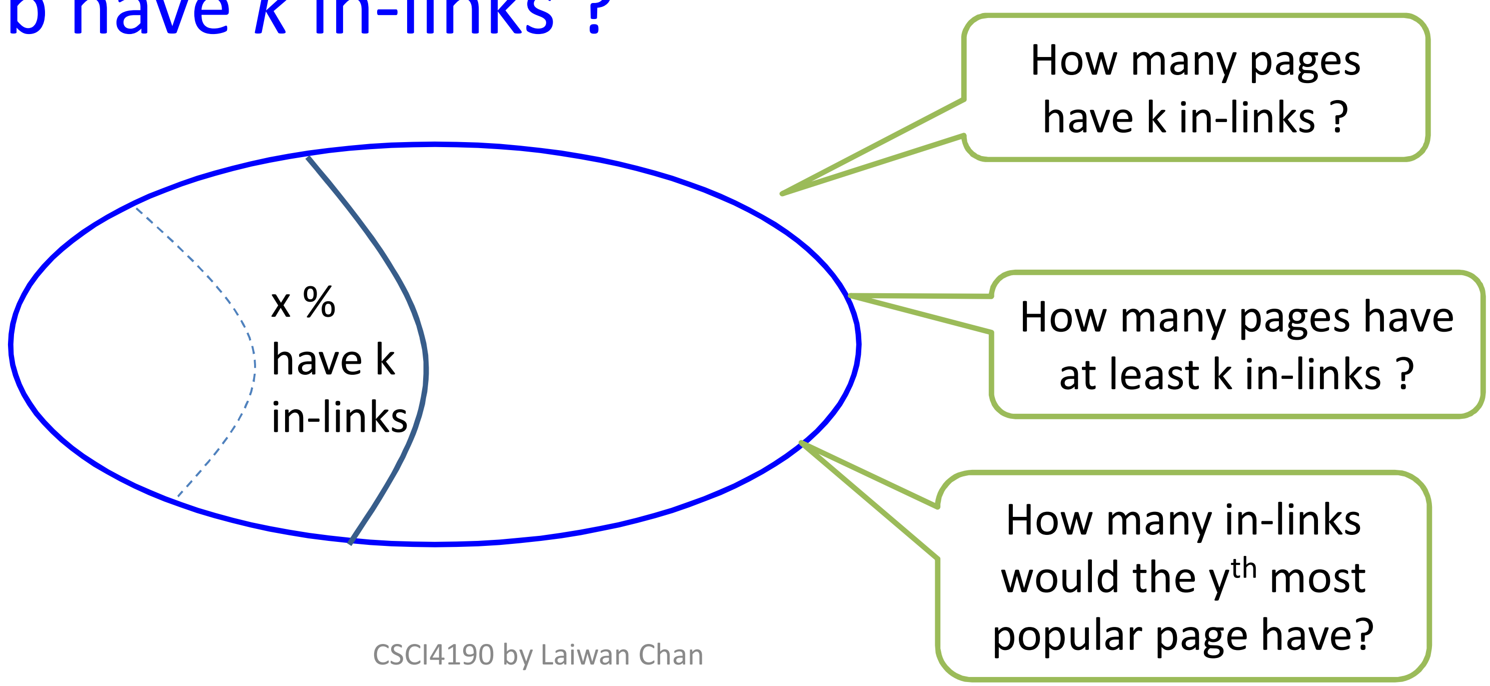


- Popularity : <u>extreme imbalanced</u> characteristics
  - How can we quantify these imbalances?
  - Why do they arise?
  - Are they somehow intrinsic to the whole idea of popularity?

Number of in-links

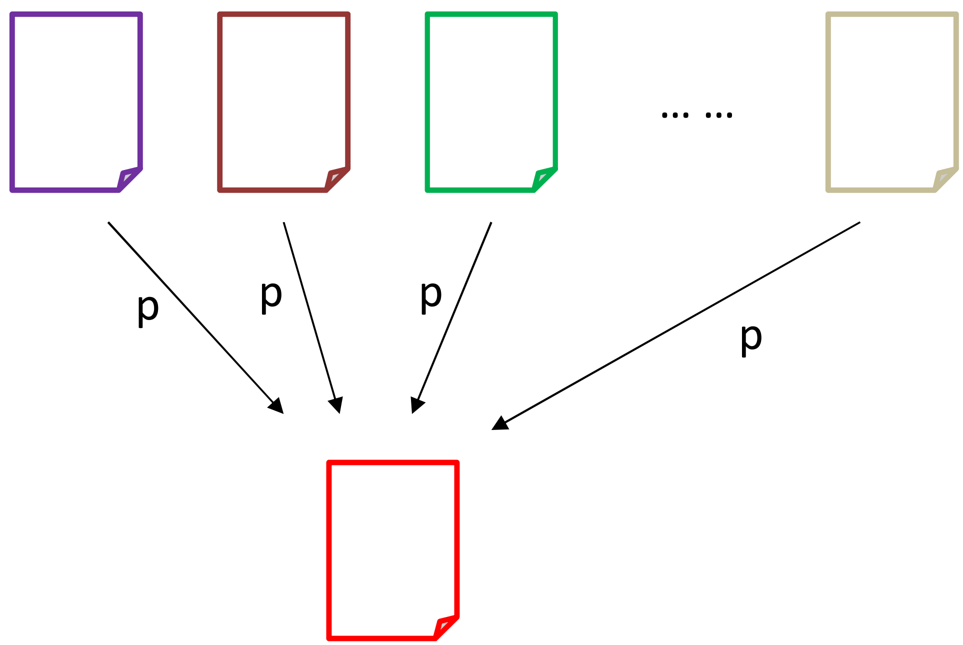
views

- Focus on web as its popularity can be measured by the number of links (in-links).
- As a function of *k*, what fraction of pages on the web have *k* in-links ?



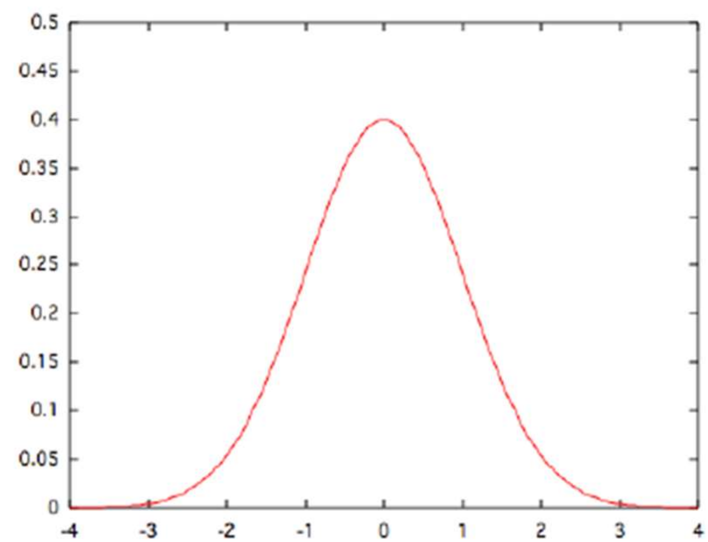
Number of in-links

– if each page decides independently at random whether to link to any other given page,



Normal Distribution

- Central Limit Theorem says that if we take any sequence of small independent random quantities, then in the limit their sum (or average) will be distributed according to the normal distribution.



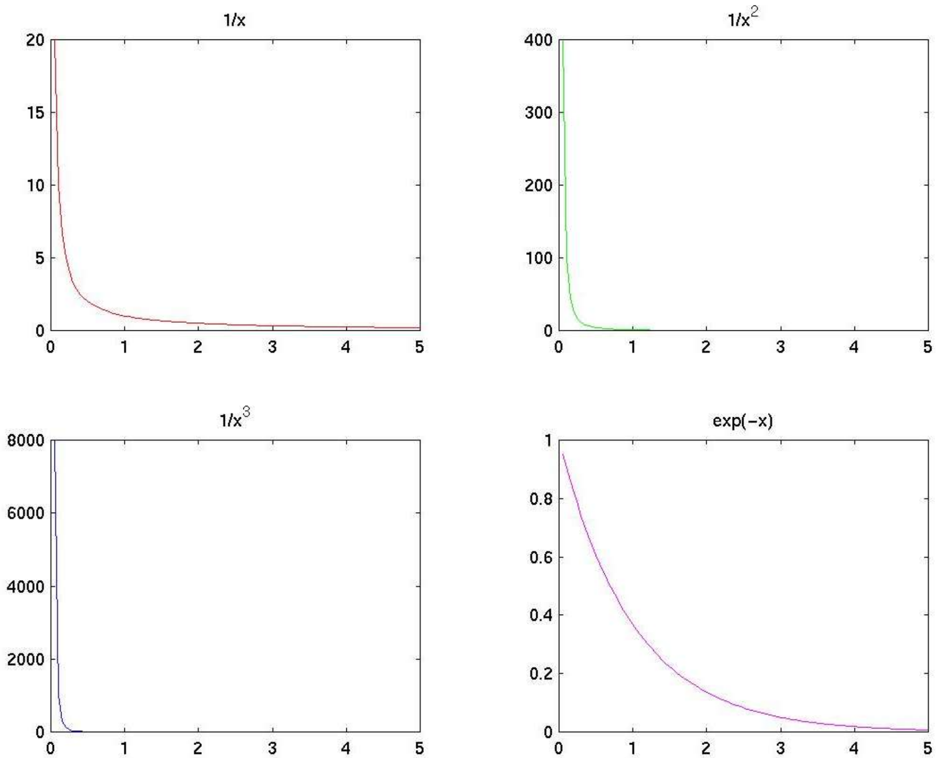
Number of in-links

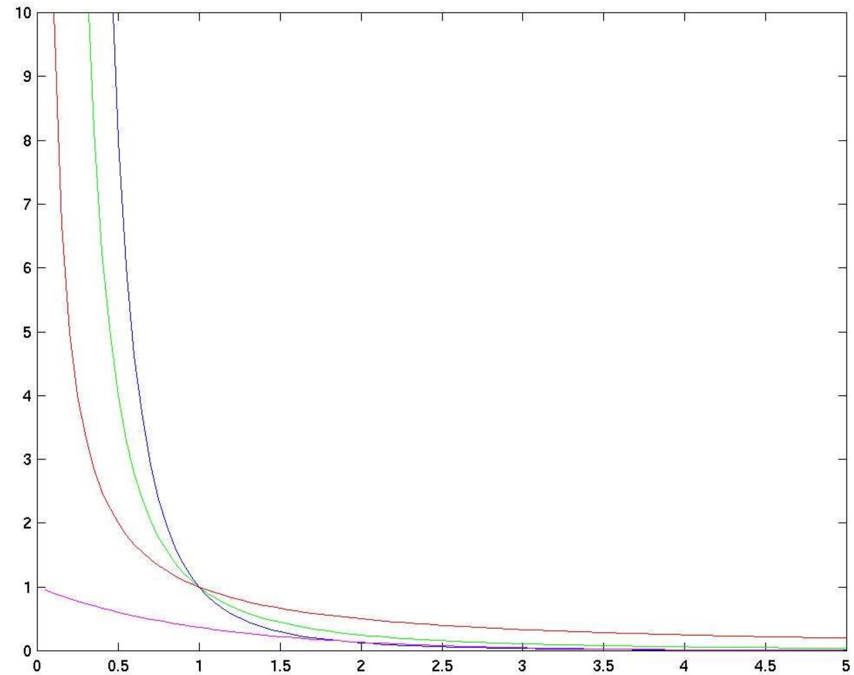
- In theory
  - if each page decides independently at random whether to link to any other given page,
  - then the number of in-links to a given page is the sum of many independent random quantities.
  - According to the Central Limit Theorem, the sum of any sequence of small independent random quantities follows the normally distribution.

Number of in-links

- In practice, the recurring finding is that the fraction of Web pages that have k in-links is approximately proportional to $1/k^2$

- A function that decreases as k to some fixed power, such as $1/k^2$ is called a power law.



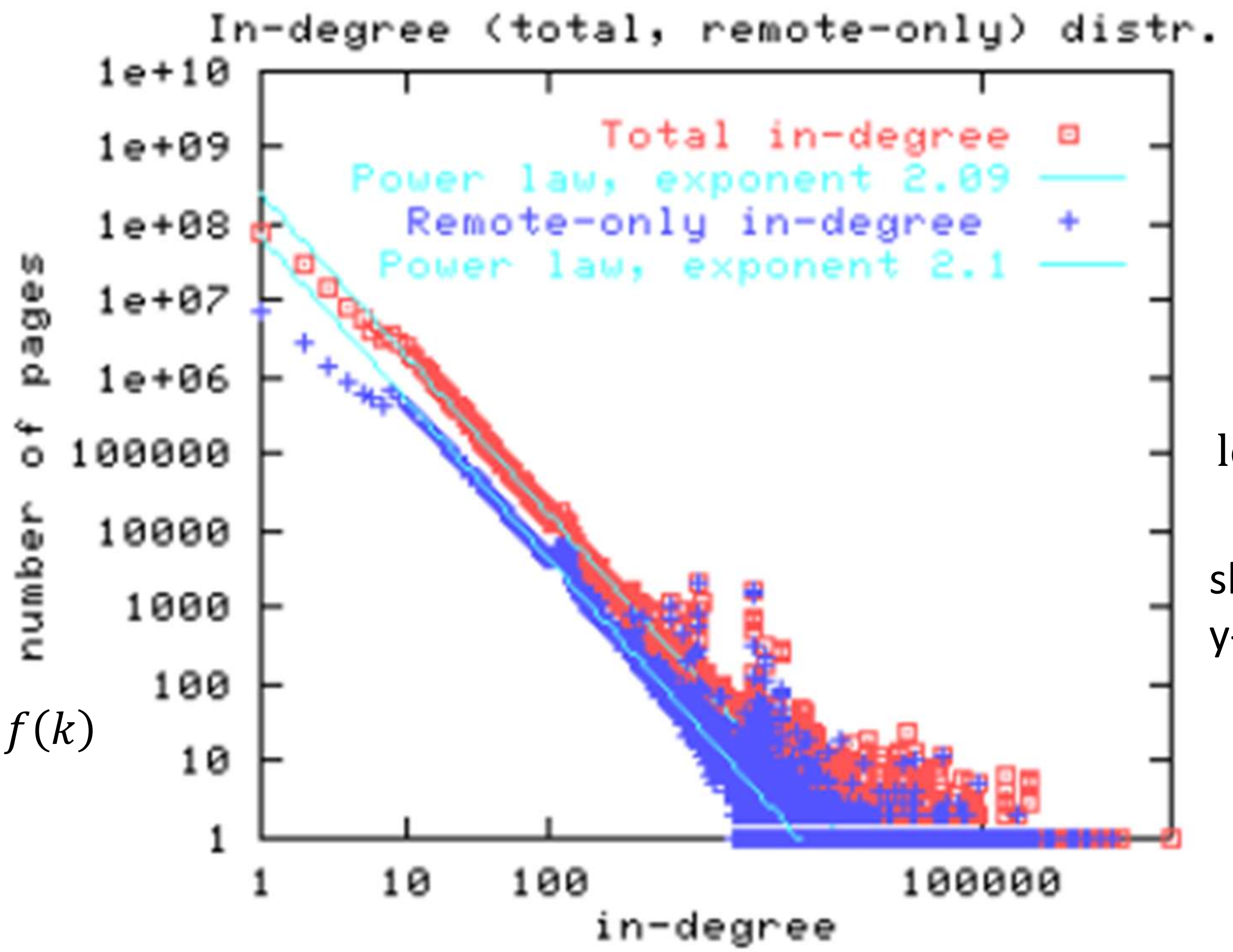


k = 1000	
$1/k^2$	10^{-6}
2^{-k}	9.3×10^{-302}

Power Laws

- Power Law $1/k^c$ for some c
  - The fraction of telephone numbers that receive k calls per day is roughly proportional to $1/k^2$.
  - The fraction of books that are bought by *k* people is roughly proportional to $1/k^3$.
  - The fraction of scientific papers that receive *k* citations in total is roughly proportional to $1/k^3$.

Number of in-links (from Broder)



$\log f(k)$

$\log k$

$$f(k) = a/k^c$$

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

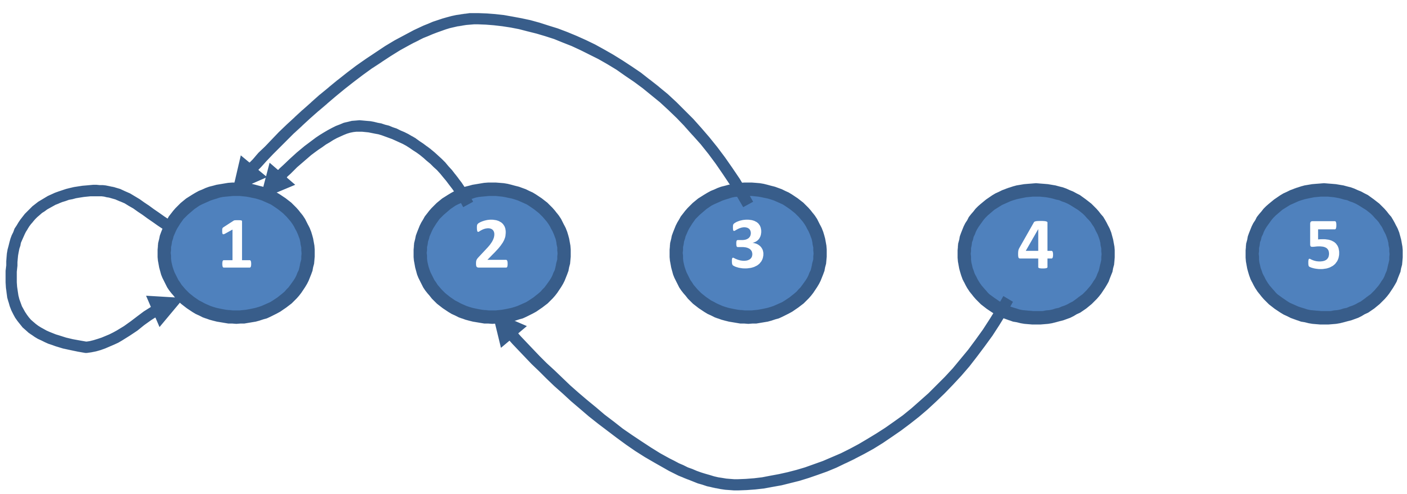
slope = -c

y-intercept = log a

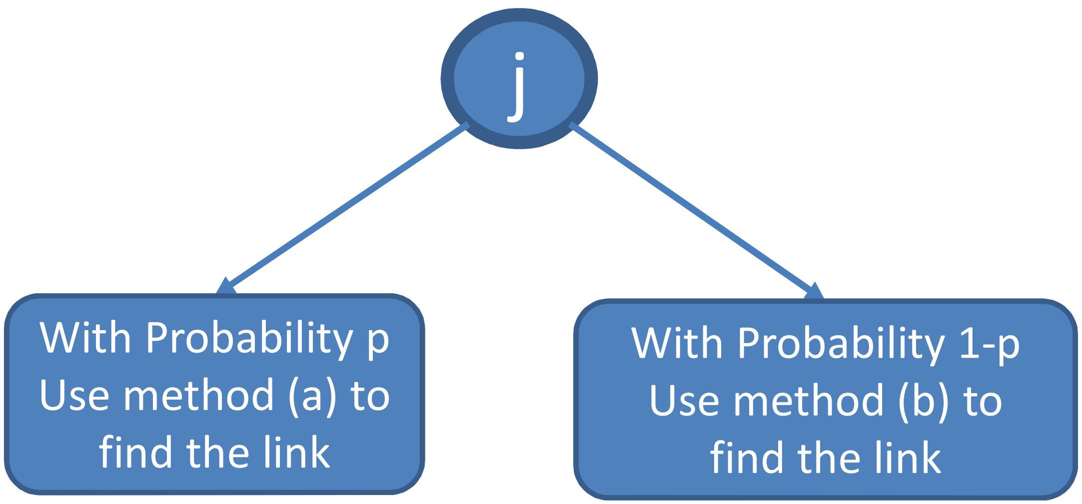
Rich-Get-Richer Models

(1) Created pages in order : 1, 2, 3, . . . , j.

(2) When j is created, it will have a link to a previous page.

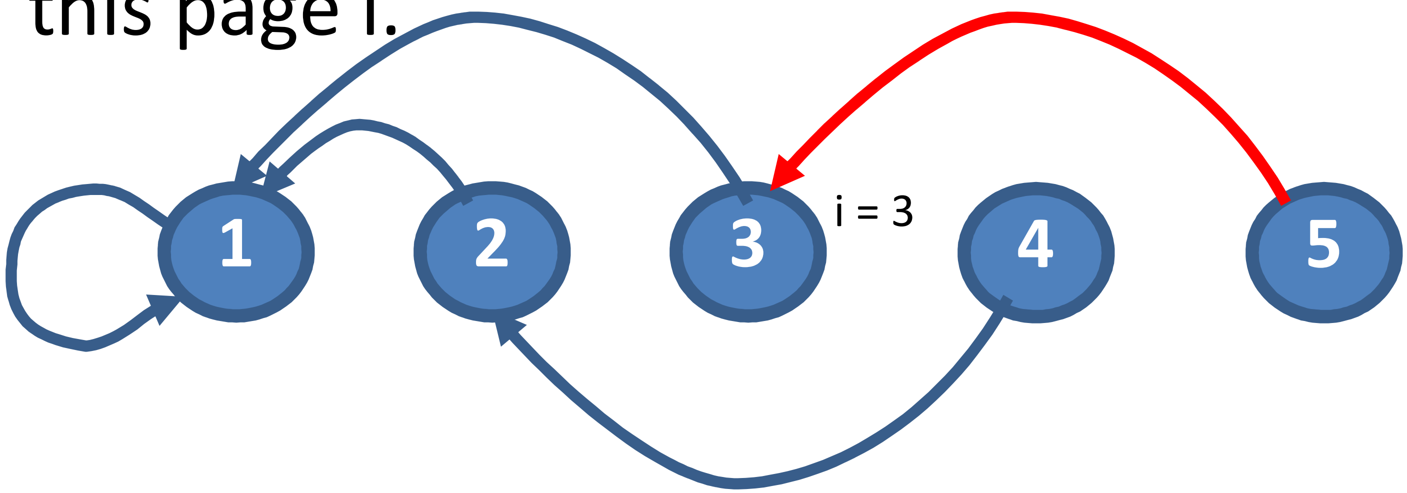


Which page will be linked by page j ?

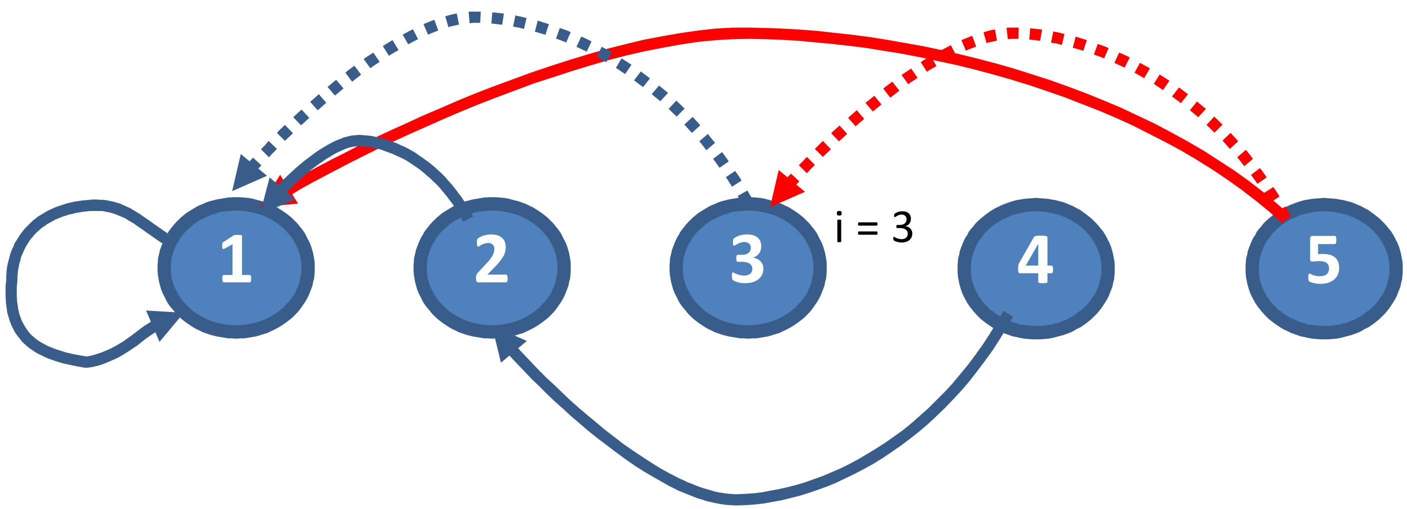


(2) When j is created, a page i is chosen uniformly at random from among all earlier pages.

(a) With probability p, page j, creates a link to this page i.



(b) With probability 1-p, page j creates a link to the page that i links to.



1. Pages are created in order, and named 1, 2, 3, . . . , N.
2. When page j is created, it produces a link to an earlier Web page according to
   a) 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.
   b) With probability 1 – p, page j instead chooses a page i uniformly at random from among all earlier pages, and creates a link to the page that i points to.
   c) This describes the creation of a single link from page j; one can repeat this process to create multiple, independently generated links from page j. (We suppose that each page creates just one outbound link.)

Re-writing 2(b)

Previous version :

2(b) With probability 1 – p, page j instead chooses a page i uniformly at random from among all earlier pages, and creates a link to the page that i points to.

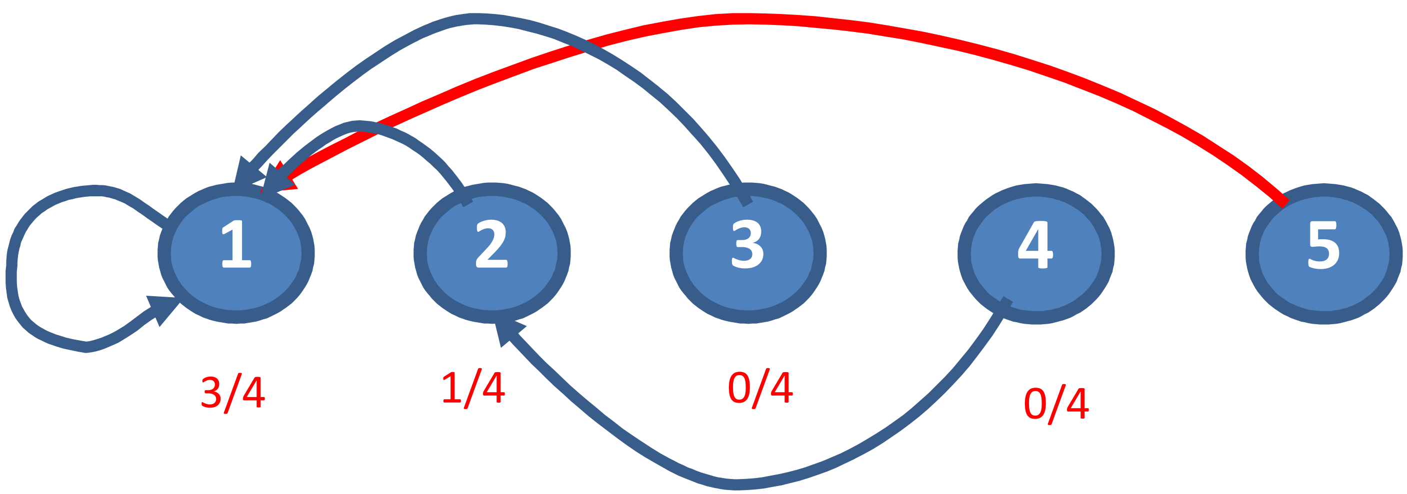
The probability that you end up linking to some page is directly proportional to the total number of pages that currently link to.

New version :

2 (b) 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.

Re-writing 2(b)

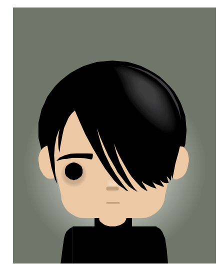
- 2 (b) With probability 1 − p, page j chooses a page ℓ with probability proportional to ℓ's current number of in-links, and creates a link to ℓ.

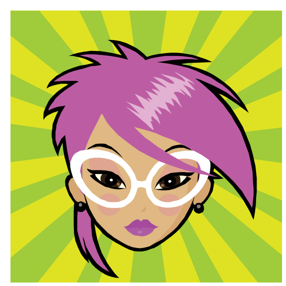


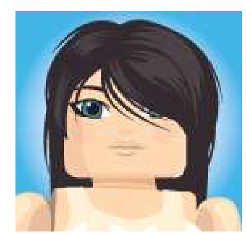
	Node 1	Node 2	Node 3	Node 4
Method (a) P	1/4	1/4	1/4	1/4
Method (b) 1-p	3/4	1/4	0/4	0/4
total	p/4 +3(1-p)/4	p/4+(1-p)/4	p/4	p/4

Rich-Get-Richer Models

- preferential attachment : "preferentially" to pages that already have high popularity.
  – people have a tendency to copy the decisions of people who act before them.







Analysis of Rich-Get-Richer Processes

- To prove that the fraction of nodes with k in-links is distributed approximately according to a power law $1/k^c$.
  - c depends on the basic features of the model (p).

1. Pages are created in order, and named 1, 2, 3, . . . , N.
2. When page j is created, it produces a link to an earlier Web page according to
   a) 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.
   b) 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. (new version)
   c) This describes the creation of a single link from page j; one can repeat this process to create multiple, independently generated links from page j. (We suppose that each page creates just one outbound link.)

A deterministic approximation of the rich-get-richer process.

- No. of in-links to a node j at a time t ≥ j is a random variable $X_j(t)$
- The initial condition : j is created at time j, $X_j(j) = 0$.
- The growth equation : at time t+1, t pages and t links
  - (a) Prob of linking to a randomly chosen page : $\frac{p}{t}$
  - (b) Prob of linking to node j with $X_j(t)$ in-link : $\frac{(1-p)X_j(t)}{t}$
  - the number of in-links to node j increases with probability $\frac{p}{t} + \frac{(1-p)X_j(t)}{t}$

Identifying a power law in the deterministic approximation

- No. of in-links increases with prob = $\frac{p}{t} + \frac{(1-p)X_j(t)}{t}$
- Use a deterministic model and replace $X_j(t)$ by a continuous function x_j
- Model the rate growth by the differential equation $\frac{dx_j}{dt} = \frac{p}{t} + \frac{(1-p)x_j}{t}$
- Put $q = 1 - p$

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

Identifying a power law in the deterministic approximation

- Solve the equation $\frac{dx_j}{dt} = \frac{p+qx_j}{t}$
- Re-arrange terms and integrate both sides

$$\int \frac{1}{p+qx_j} dx_j = \int \frac{1}{t} dt$$

- Solution

$$\ln\left(p+qx_j\right) = q\ln(\mathrm{t}) + \mathrm{c}$$

$$\ln\left(p+qx_j\right) = \ln(At^q) \qquad \text{for } A = e^c$$

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

- Initial condition at time t = j, $x_j = 0$, implies $\frac{1}{q}(Aj^q - p) = 0$

  We have $A = p/j^q$

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

Identifying a power law in the deterministic approximation

- No. of in-links of node j = $x_j = \frac{p}{q}\left[\left(\frac{t}{j}\right)^q - 1\right]$
- For a given value of *k*, and a time *t*, what fraction of nodes (among t nodes) have at least k in-links at time t?
  - what fraction of all functions x_j satisfy $x_j(t) \geq k$?
- $x_j = \frac{p}{q}\left[\left(\frac{t}{j}\right)^q - 1\right] \geq k$

- $\frac{p}{q}\left[\left(\frac{t}{j}\right)^{q} - 1\right] \geq k$ gives $\left(\frac{t}{j}\right)^{q} \geq \left[\frac{q}{p} \cdot k + 1\right]$
- So j $\leq t\left[\frac{q}{p} \cdot k + 1\right]^{-1/q}$
- At time t, we have nodes $x_1, x_2, \ldots x_t$. The fraction of nodes with at least k links is

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

- F(k) = the fraction of nodes with at least k links

$$= \left[\frac{q}{p}\cdot k + 1\right]^{-1/q}$$

- f(k) = the fraction of nodes with exactly k links

$$-\frac{dF}{dk} = \frac{1}{q}\frac{q}{p}\left[\frac{q}{p}\cdot k + 1\right]^{-1-1/q}$$

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

- f(k) = the fraction of nodes with exactly k links $= \frac{1}{p}\left[1 + \frac{1-p}{p} \cdot k\right]^{-(1+\frac{1}{1-p})}$
- When p is close to 1, link formation is mainly based on uniform random choices

$$(1 + \frac{1}{1-p}) \rightarrow \infty$$

nodes with very large numbers of in-links become increasingly rare

- When p is close to 0, network is strongly governed by rich-get-richer behaviour

$$(1 + \frac{1}{1-p}) \rightarrow 2$$

power-law exponent decreases toward 2.

1. Pages are created in order, and named 1, 2, 3, . . . , N.
2. When page j is created, it produces a link to an earlier Web page according to
   a) 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.
   b) With probability 1 – p, page j instead chooses a page i uniformly at random from among all earlier pages, and creates a link to the page that i points to.
   c) This describes the creation of a <u>single link</u> from page j; one can repeat this process to create multiple, independently generated links from page j. (We suppose that each page creates just one outbound link.)

The Unpredictability of Rich-Get-Richer Effects

- If we roll time back and run history again, would the most popular items be the same ?
- Salgankik et al's experiment
  - a music download site, with 48 obscure songs of varying quality
  - visitors were given
    - the opportunity to listen to them and
    - the current "download count"
    - the opportunity to download copies of the songs

The Unpredictability of Rich-Get-Richer Effects

- Eight "parallel" copies of the site with initial download counts of zero.
- The "market share" of the different songs varied considerably across the different parallel copies
  - the best songs never ended up at the bottom and the worst songs never ended up at the top.
- The 9th version : no download counts was provided
  - there was significantly less variation in the market share of different songs.

Closer Relationships between Power Laws and Information Cascades?

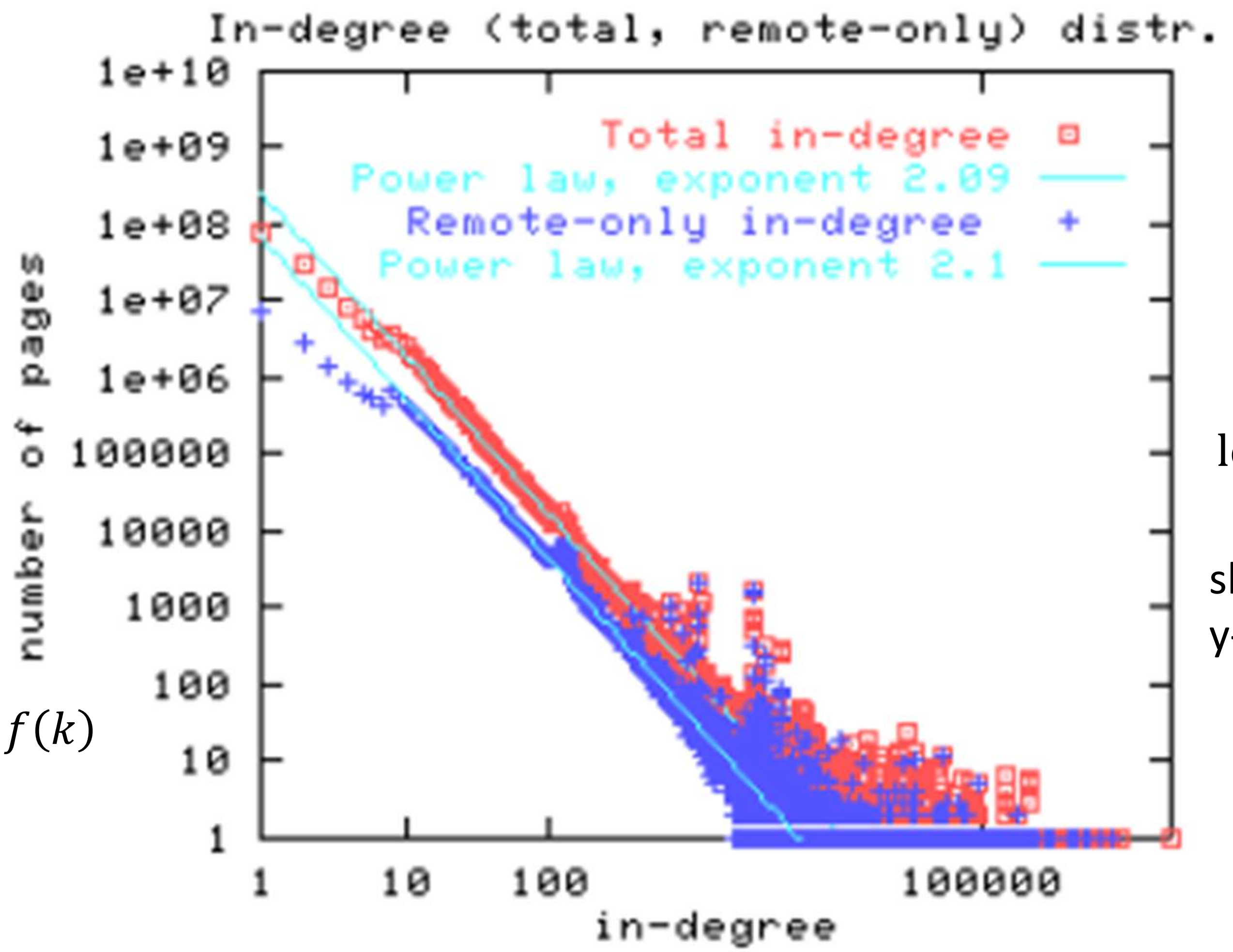
- Information cascade
  - accepting or rejecting
  - know the decision of a population
  - rational decision making
- Power law
  - many choices of pages
  - consult the decision of just one other page
  - assume later people imitate the decisions of earlier people

The long tail

- Are most sales being generated by
  - a small set of items that are enormously popular
    - Hits !!
  - a much larger population of items that are each individually less popular?
    - niche products (products that are made and marketed for use in a small and specialized but profitable market.)

Current Hit vs oldies

Number of in-links (from Broder)



$\log f(k)$

$\log k$

$$f(k) = a/k^c$$

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

slope = -c
y-intercept = log a

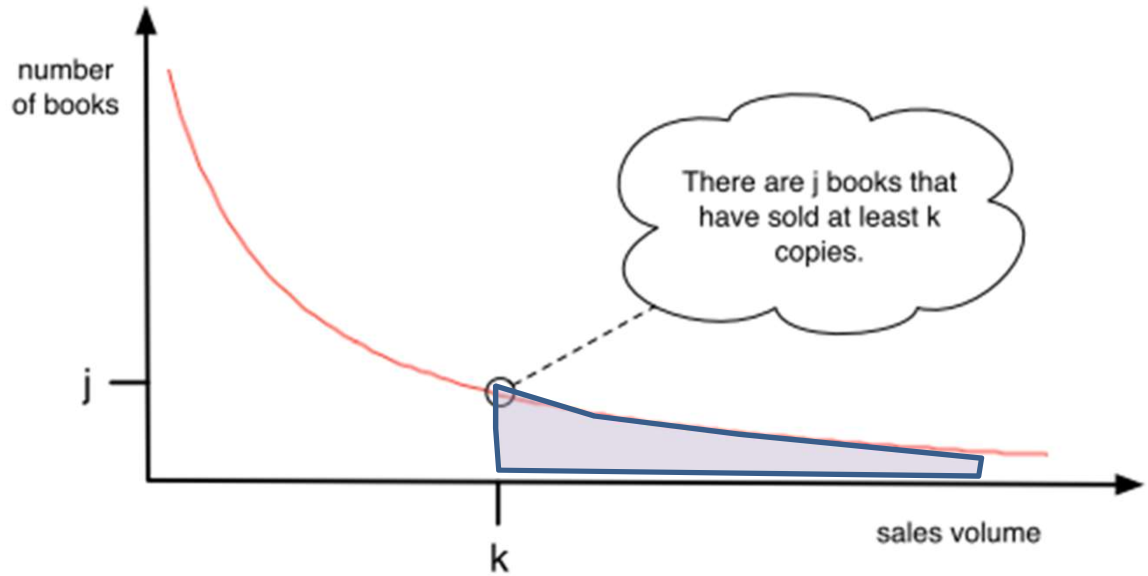
Visualizing the Long Tail

- As a function of k, what fraction of items have popularity exactly k?

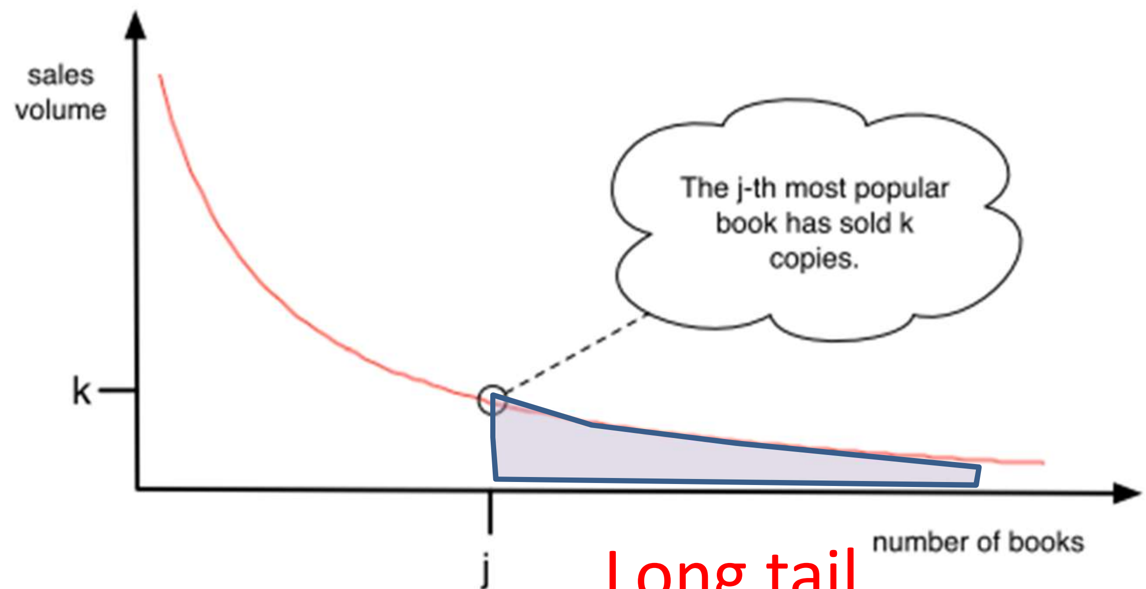
↓

- As a function of k, what number of items have popularity at least k?

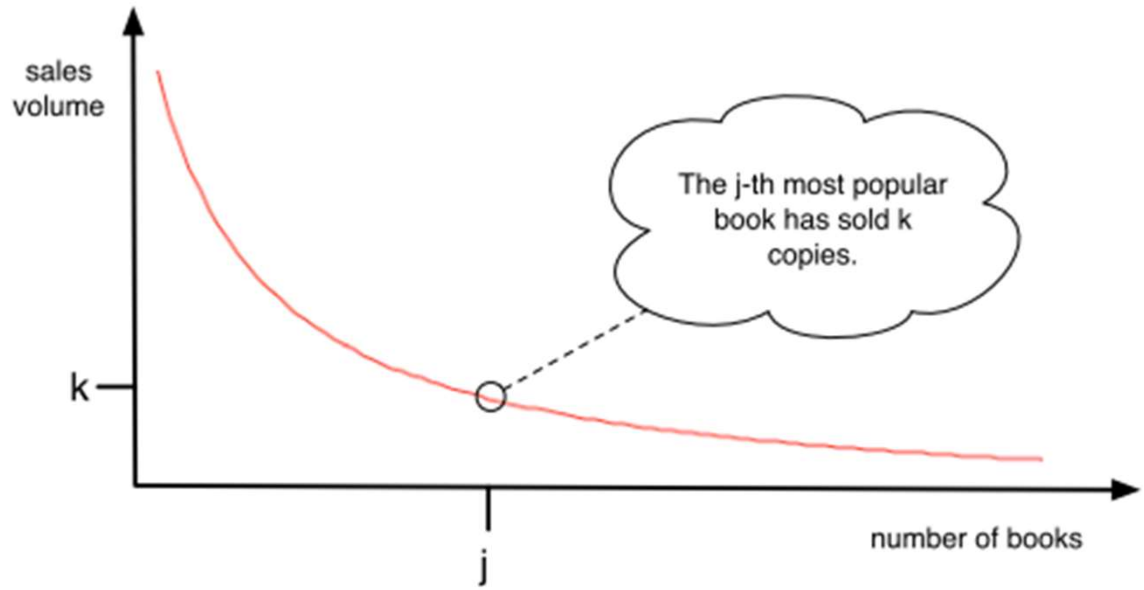
- As a function of k, what number of items have popularity at least k?



- interchange the x and y axis



Long tail



- Zipf plots
- Zipf’s Law : the frequency of the j th most common word in English (or most other widespread human languages) is proportional to 1/j.

Top hundred words

Zipf's law is a neat, general fact about word frequency distribution. G K Zipf discovered that the frequency of the *k*th most frequent word is proportional to 1/*k* (*Human Behavior and the Principle of Least Effort, an Introduction to Human Ecology* (Reading, MA, Addison-Wesley, 1949), cited in Knuth, *The Art of Computer Programming: vol 3, Sorting and Searching* (Reading, MA: Addison-Wesley, 1973), 397). The top hundred words in this database adhere to the law quite well.

	frequency (per milln)	cumulative frequency	frequency rank	alphabet rank
the	68351.63	68351.63	1	318525
of	33008.66	101360.29	2	212425
and	28651.11	130011.40	3	11331
to	27599.22	157610.62	4	322312
a	23160.48	180771.10	5	1
in	20670.81	201441.91	6	149032
is	10571.15	212013.06	7	156934
that	10549.02	222562.08	8	318470
was	9939.26	232501.34	9	356587
it	9882.90	242384.23	10	157771
for	9309.44	251693.67	11	114281
on	7636.66	259330.33	12	213645
with	7171.07	266501.39	13	361235
he	7167.84	273669.23	14	134413
be	7153.17	280822.40	15	27945
I	7036.88	287859.28	16	146205
by	5866.89	293726.17	17	44040
as	5793.35	299519.52	18	19178
at	5154.12	304673.64	19	20631
you	5043.27	309716.91	20	364651
are	5000.14	314717.05	21	17618
his	4963.47	319680.52	22	139433
had	4922.27	324602.79	23	131212
not	4899.77	329502.56	24	209444
this	4789.41	334291.97	25	319827
have	4685.82	338977.79	26	134106
from	4625.21	343603.01	27	117354
but	4616.26	348219.26	28	43732
which	4131.11	352350.37	29	358956
she	3991.77	356342.14	30	285912
they	3982.95	360325.09	31	319435
or	3975.58	364300.67	32	214838
an	3836.07	368136.73	33	10593
her	3692.13	371828.86	34	137067
were	3482.45	375311.31	35	358233
there	3025.87	378337.18	36	319027
we	2953.92	381291.10	37	357241
their	2929.78	384220.88	38	318680
been	2924.28	387145.16	39	28958
has	2873.74	390018.90	40	133676
will	2775.94	392794.84	41	360225
one	2764.69	395559.53	42	213720
all	2630.80	398190.33	43	7706
would	2617.11	400807.44	44	362548
can	2355.35	403162.80	45	46162

https://www.cs.cmu.edu/~cburch/words/top.html

The Effect of Search Tools and Recommendation Systems

- Search engines use popularity measures to rank Web pages
  - accentuate rich-get-richer dynamics
- Search tools target more closely to users' specific interests and enable people to find unpopular items more easily
  - counteract the rich-get-richer dynamics.