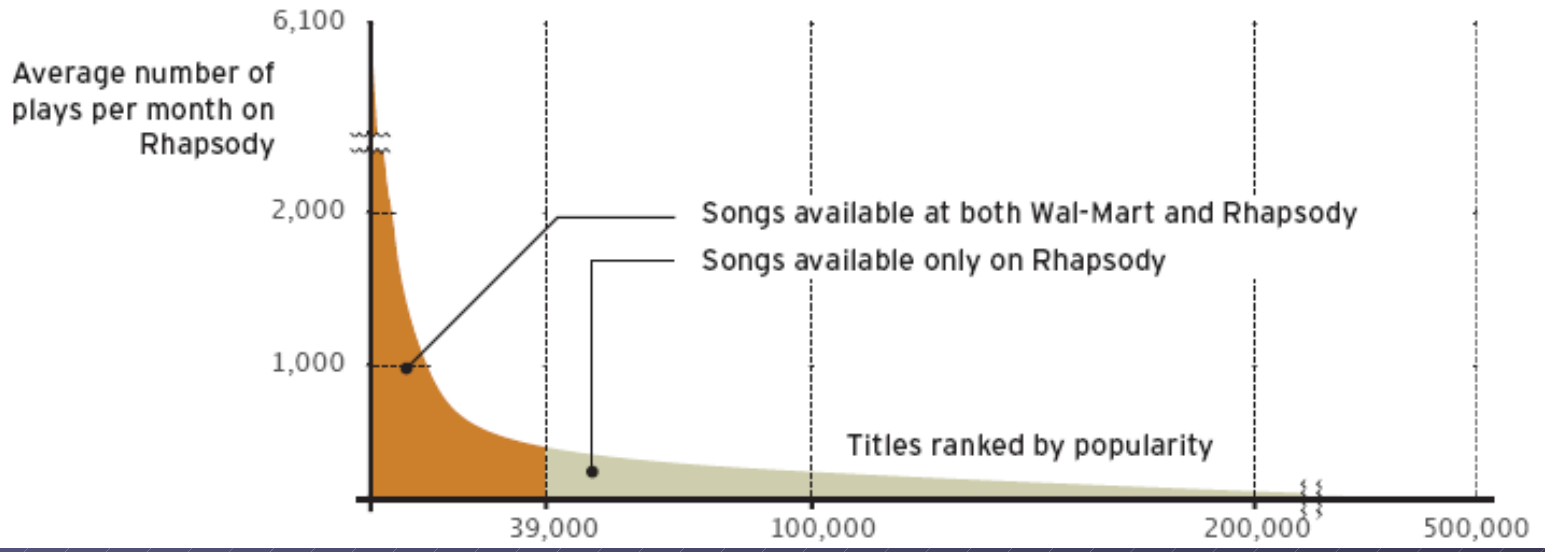
The Long Tail

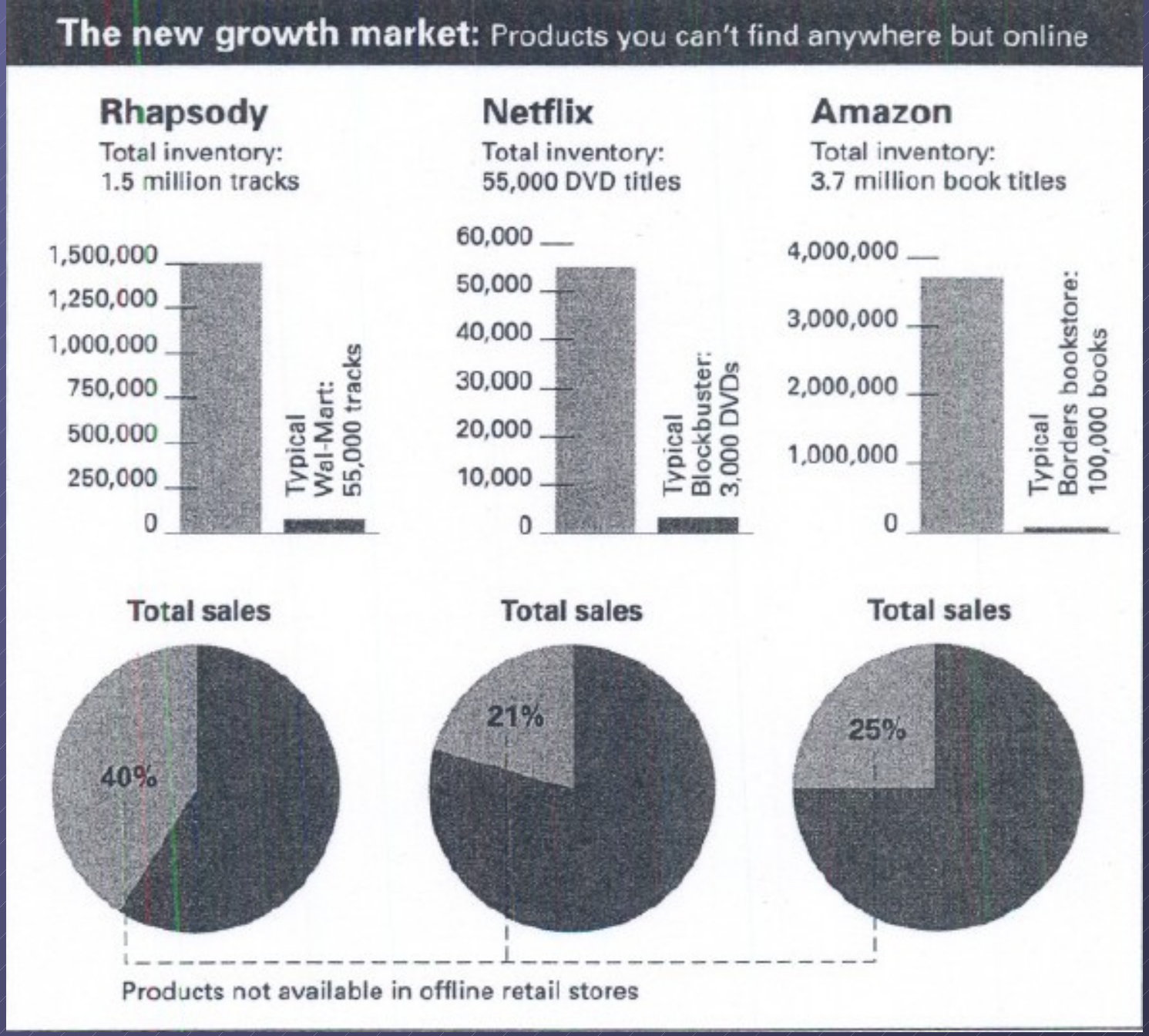
Jeff MacKie-Mason

SI 646



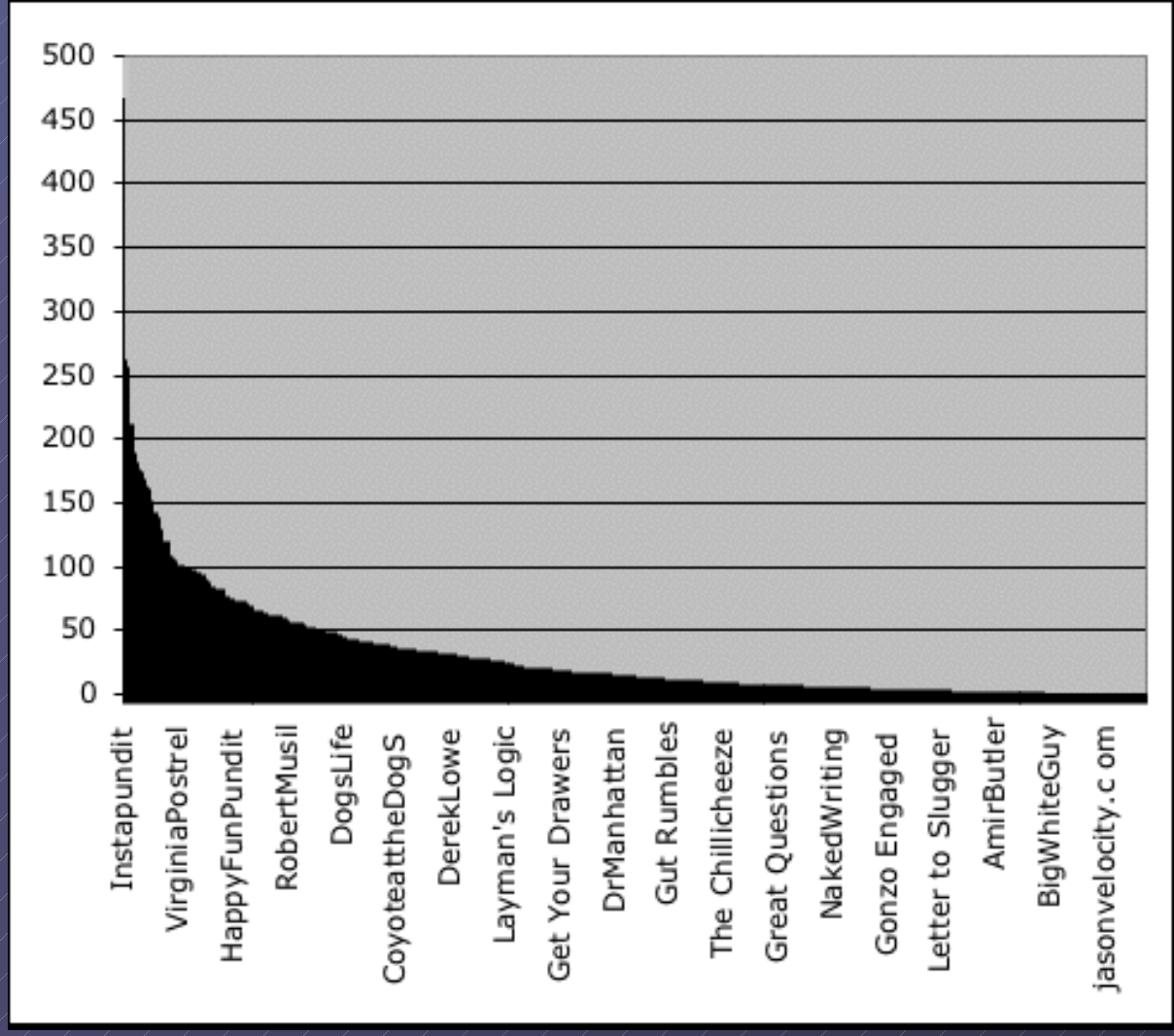
Rhapsody: More than 2 million tracks
40% of revenue from tracks not at Wal-Mart

Source: Chris Anderson (2006), The Long Tail (Hyperion). 25.



Source: Chris Anderson (2006), The Long Tail (Hyperion). 23.

433 blogs arranged in rank order by number of inbound links



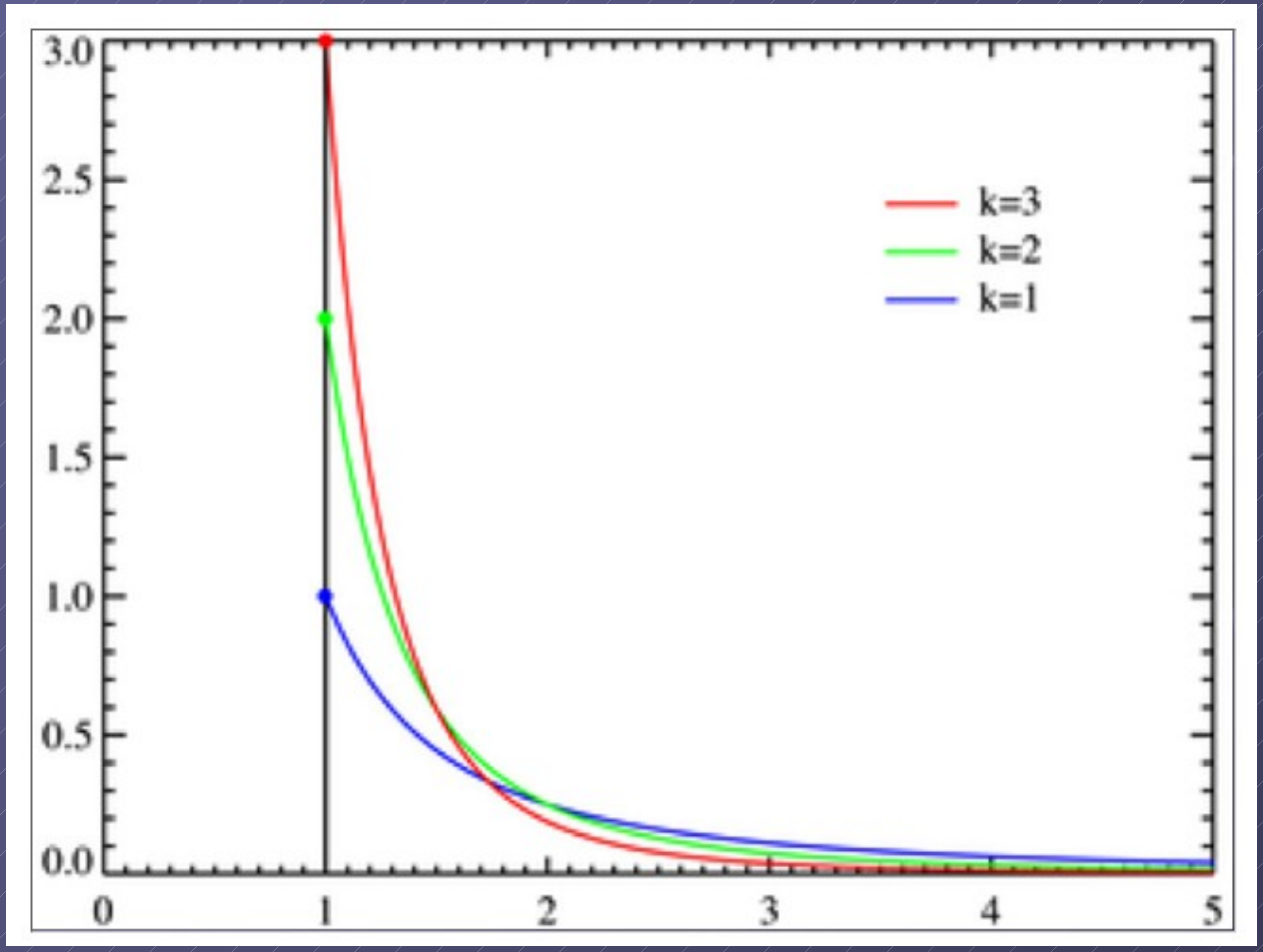
The data are drawn from N.Z Bear's 2002 work on the blogosphere ecosystem.
The current version of this project can now be found at http://www.myelin.co.nz/ecosystem/.

Source: http://www.truthlaidbear.com/ecosystem.shtml

Pareto distribution:
What is the probability that a person has wealth x?

$$f(x; k, x_m) = k \frac{x_m^k}{x^{k+1}} \text{ for } x \geq x_m.$$

f(x)



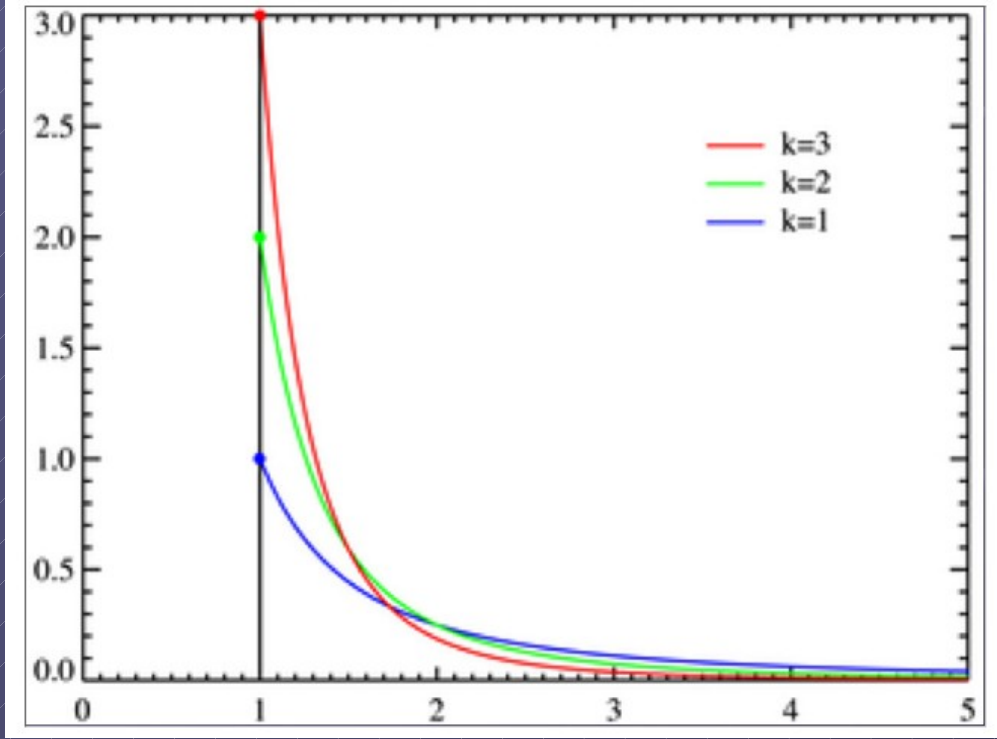
x

k constant shape parameter > 0

x_m constant scale parameter

Source : Undetermined

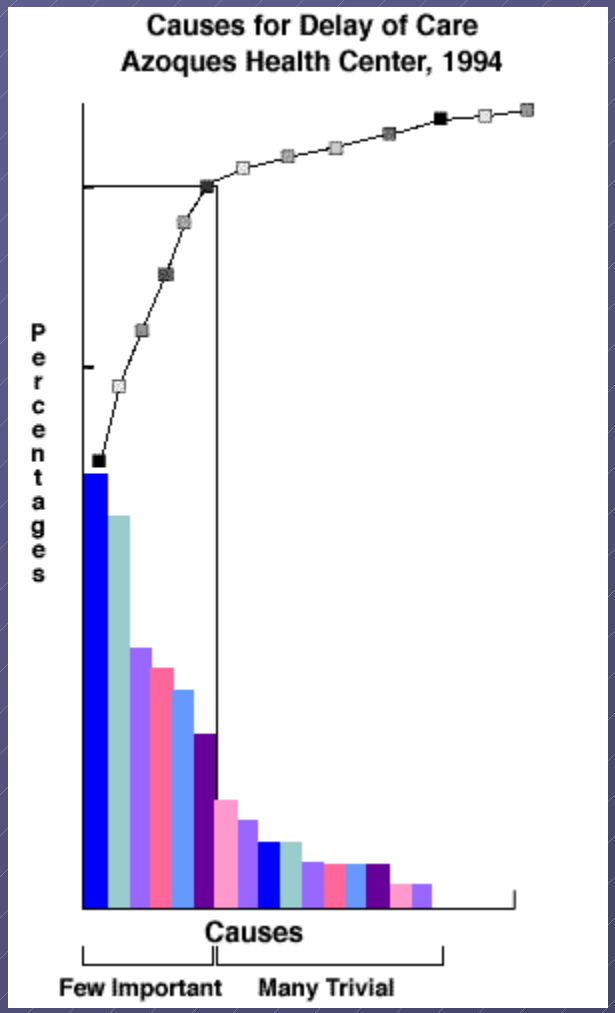
Pareto distribution: Most people have low wealth, few have very high

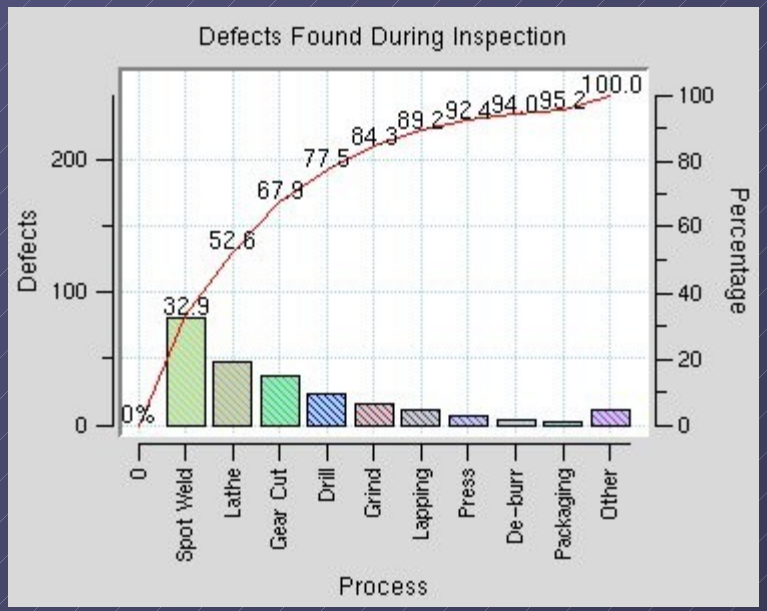


Source: Undetermined

Examples:

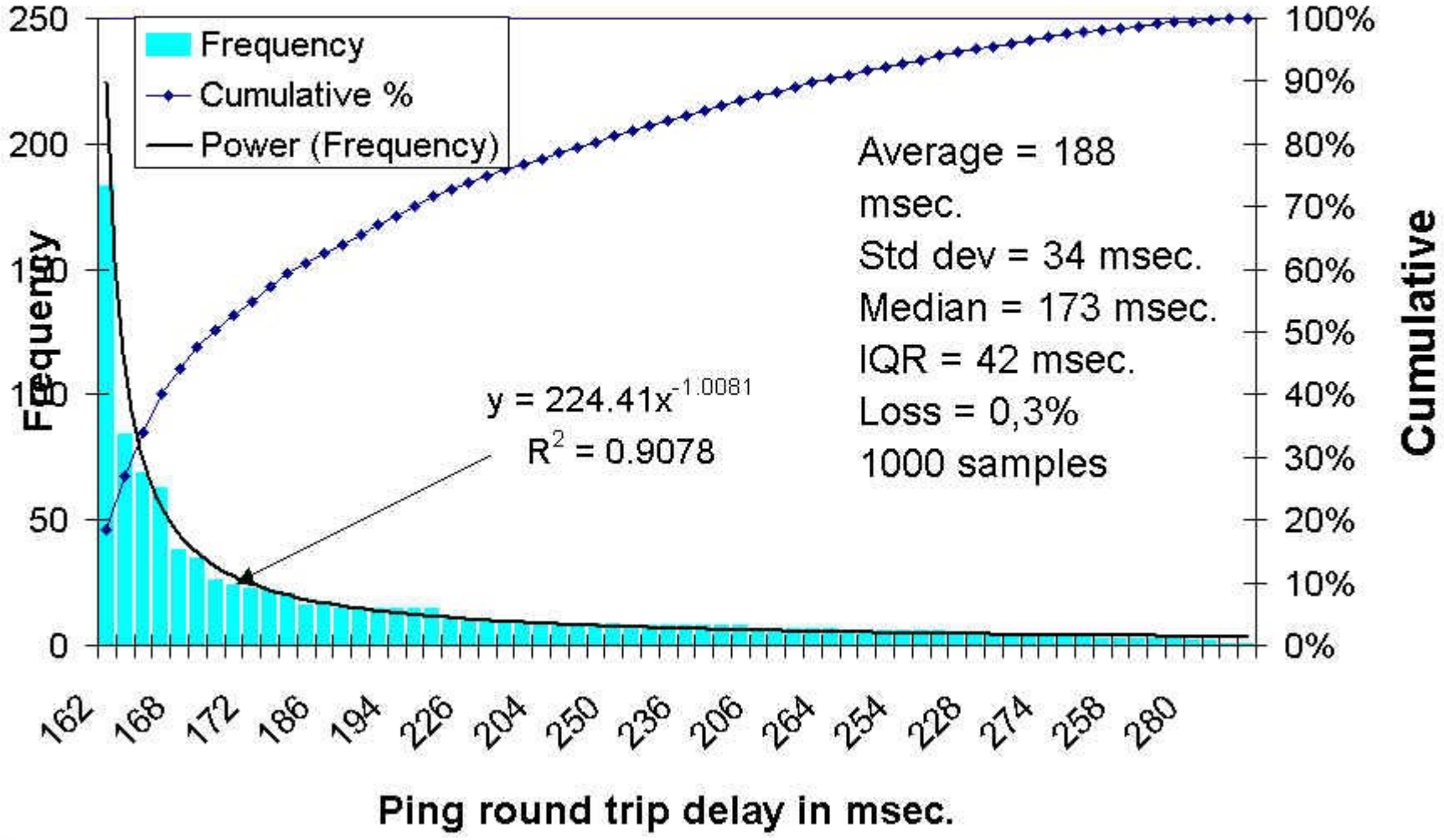
- Human settlement sizes
- File sizes transferred over Internet
- Sizes of oil fields
- Rates of return on corporate equities (stocks)
- Areas burnt in forest





Sources:
Graph on left: http://erc.msh.org/quality/pstools/pspareto.cfm
Graph on right: http://tcltk.free.fr/blt/

SLAC<=>CERN round trip delay Pareto histogram



Source: http://www.slac.stanford.edu/comp/net/wan-mon/resp-jitter-pareto.jpg

Zipf's law: "In a corpus of natural language utterances, the frequency of any word is roughly inversely proportional to its rank in the frequency table"

Source: ?

More generally, the size of the r'th largest occurrence of the event is inversely proportional to it's rank:

$$y = a\, r^{b}$$

with b close to unity

In the "Brown" corpus,

- *the* accounts for 7% = .07/1
- *of* for 3.5% = .07/2
- *and* for 2.8% = .07/2.5

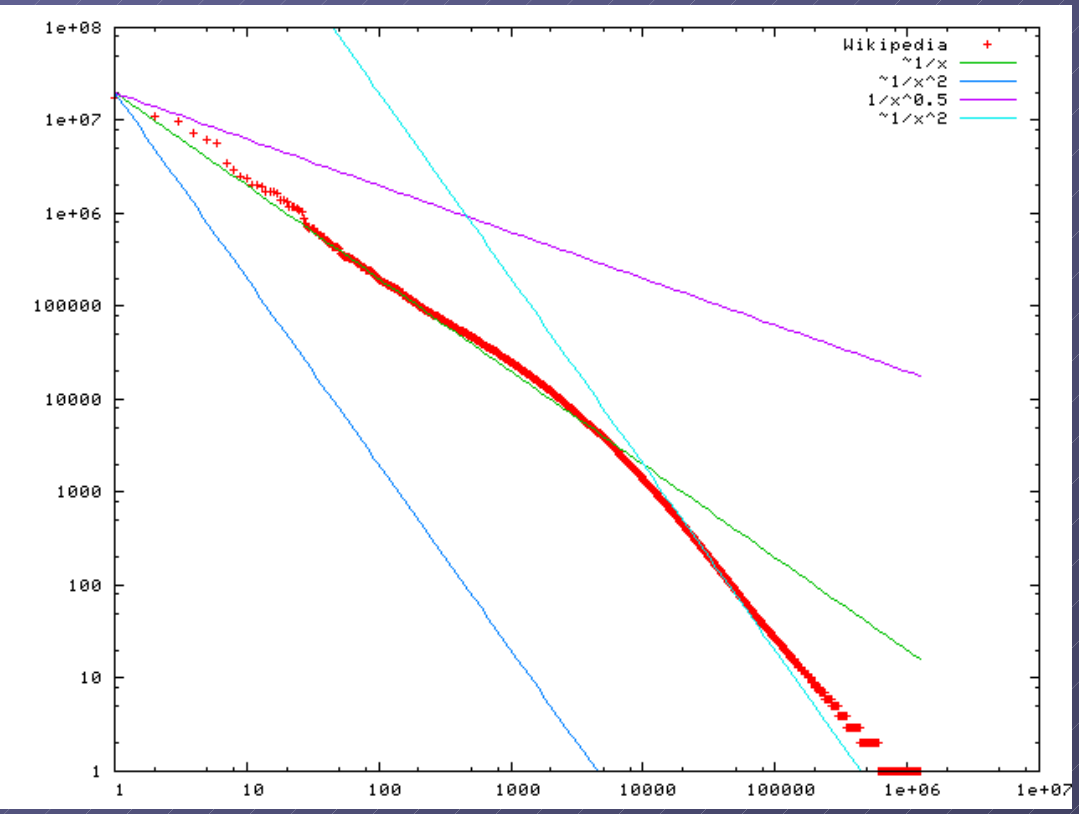
The first 135 words account for 50%

Pareto & Zipf are both examples of a power law:

$$y = a\, x^k$$

Take logs of both sides: $\log y = \log a + k \log x$

which is linear: $z = a + b\, v$



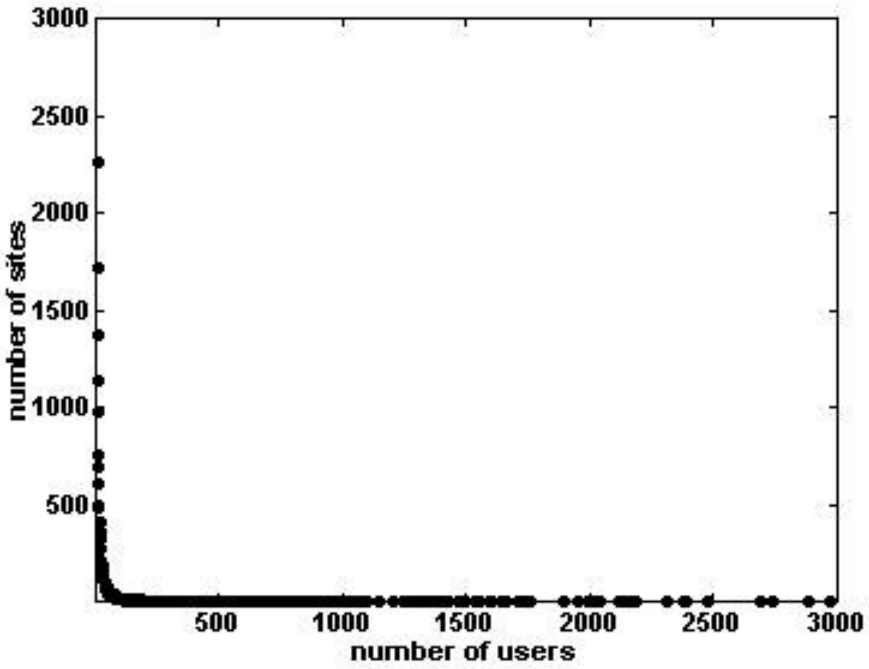
A plot of word frequency in Wikipedia (November 27, 2006).

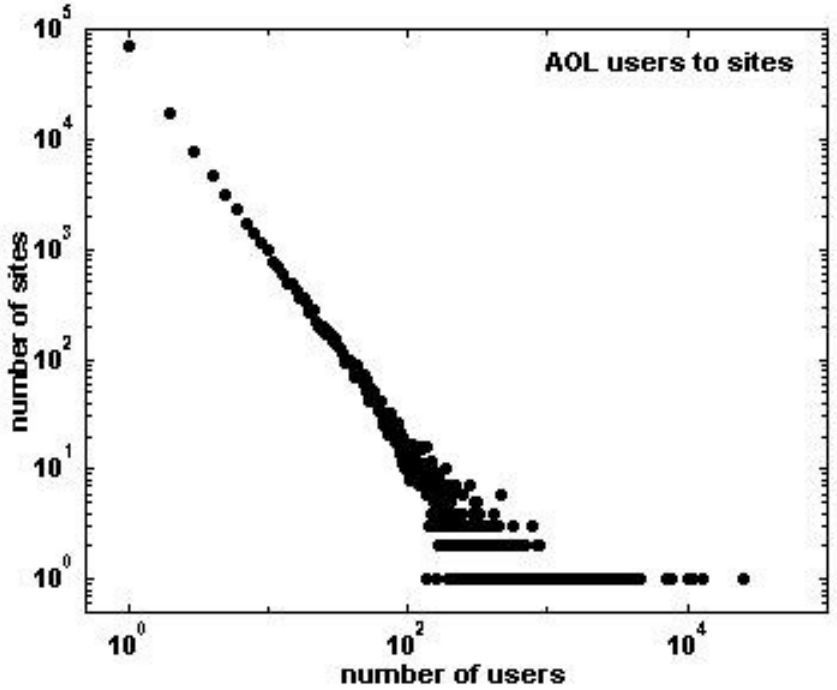
The plot is in log-log coordinates.

x is rank of a word in the frequency table; y is the total number of the word's occurrences.
Most popular words are "the", "of" and "and", as expected.

Zipf's law corresponds to the upper linear portion of the curve, roughly following the green $\log y = 1.3 \times 10^7 - \log x$ line.

Source: http://en.wikipedia.org/wiki/Zipf's_law

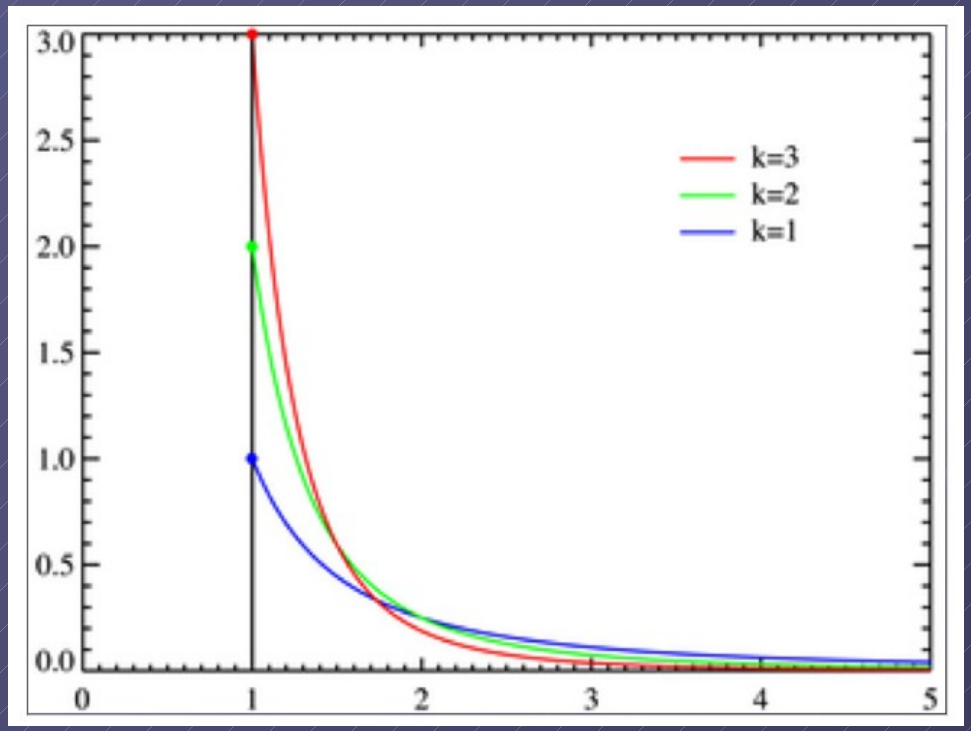




Source: Sites visited by AOL users, December day 1997.
L. Adamic, “Zipf, Power-laws, and Pareto - a ranking tutorial”,
http://www.hpl.hp.com/research/idl/papers/ranking/ranking.html

Pareto = Zipf = Power law

probability

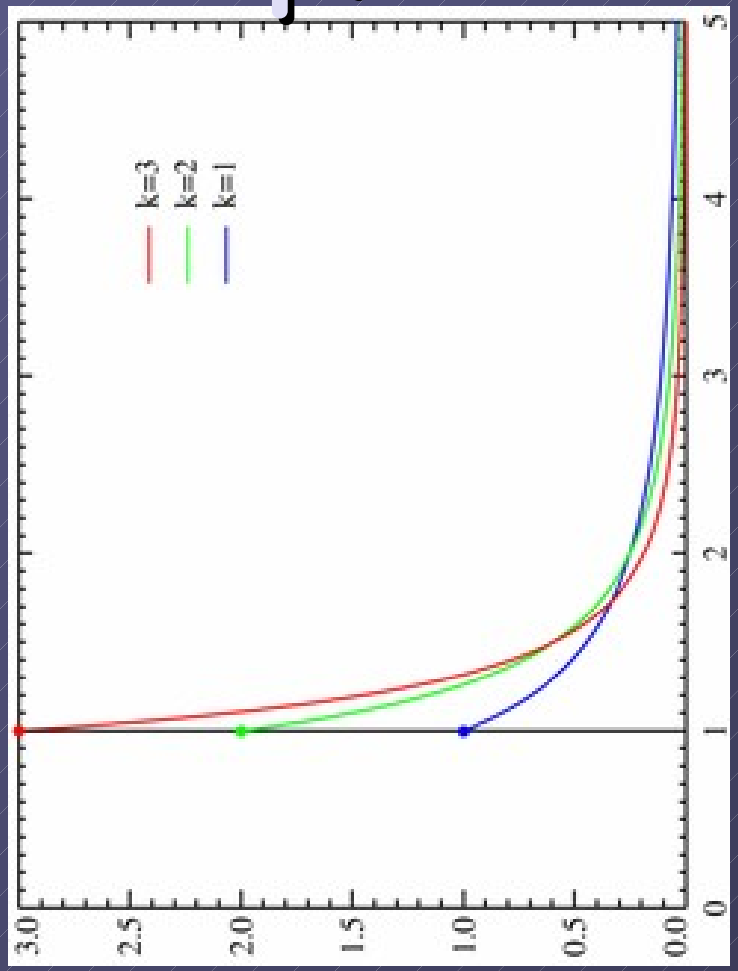


size of x

Source: Undetermined

Rotate…

Pareto = Zipf = Power law

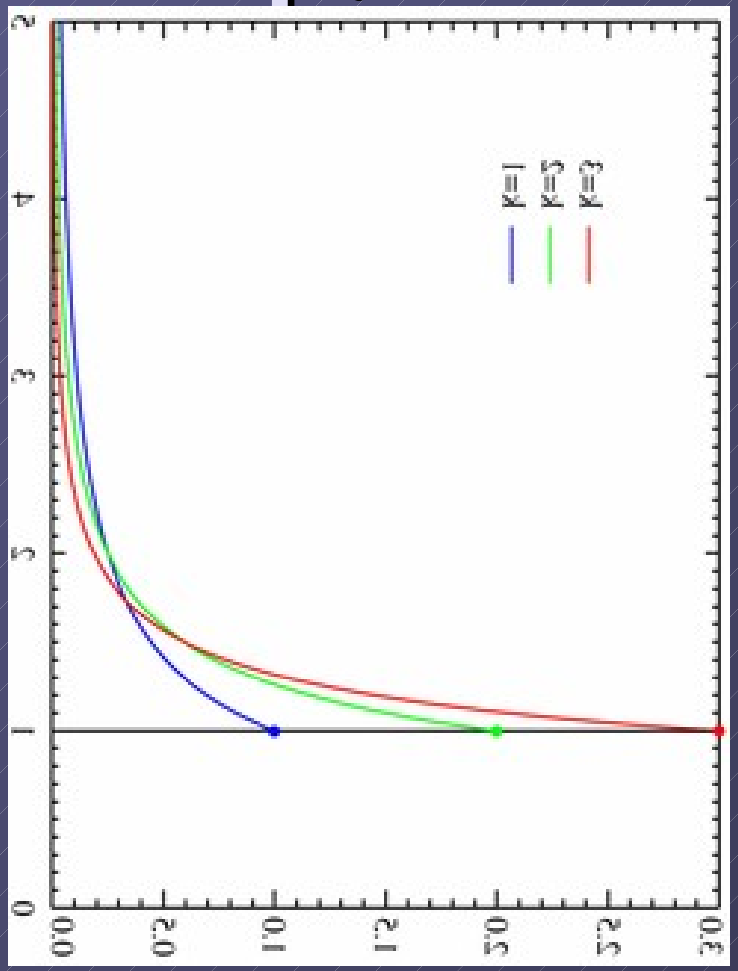


Source: Undetermined

Then flip…

Pareto = Zipf = Power law

size of x

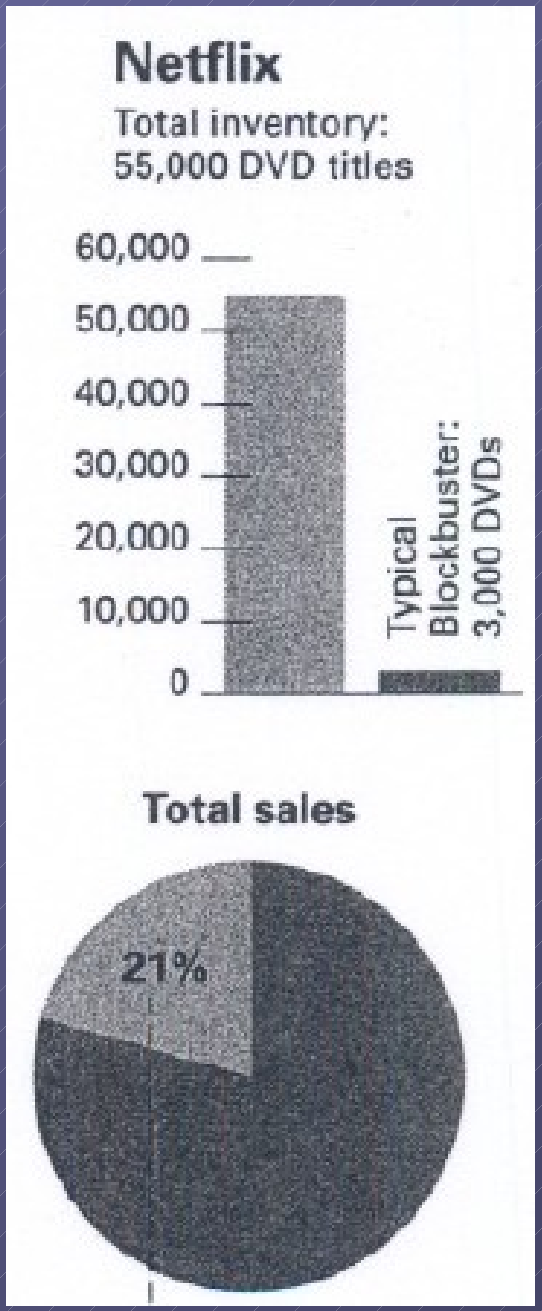


rank of x

Source: Undetermined

What's this about an 80-20 rule?

For many phenomena, 80% of the consequences stem from 20% of the causes (*Pareto principle*)



Source: Chris Anderson (2006), The Long Tail (Hyperion). 25.

3000 out of 55,000 titles: 5.5%

79% of sales

For Netflix, “80 – 6 rule”

What causes power laws for goods variety?

variety

quality variation

network effects

scarcity

What are we measuring? Rank orders or significance of *what*?

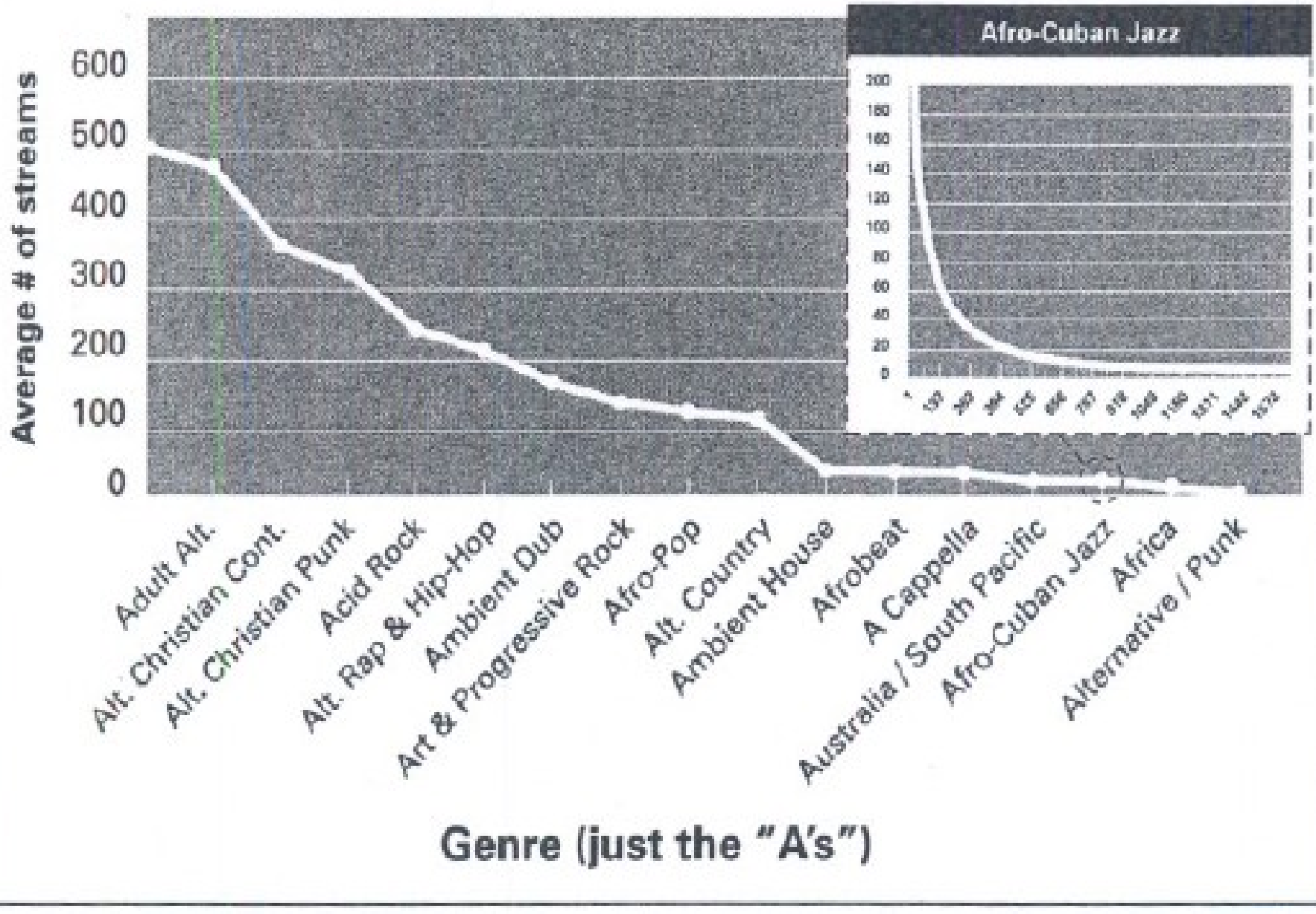
Maybe, if we're Amazon.

E.g., do we want to know about books?

What if we're O'Reilly (technical publisher)?

Generally, goods power laws apply to subcategories (e.g., genres) as well…

Tails within Tails



Source: Chris Anderson (2006), The Long Tail (Hyperion). 140.

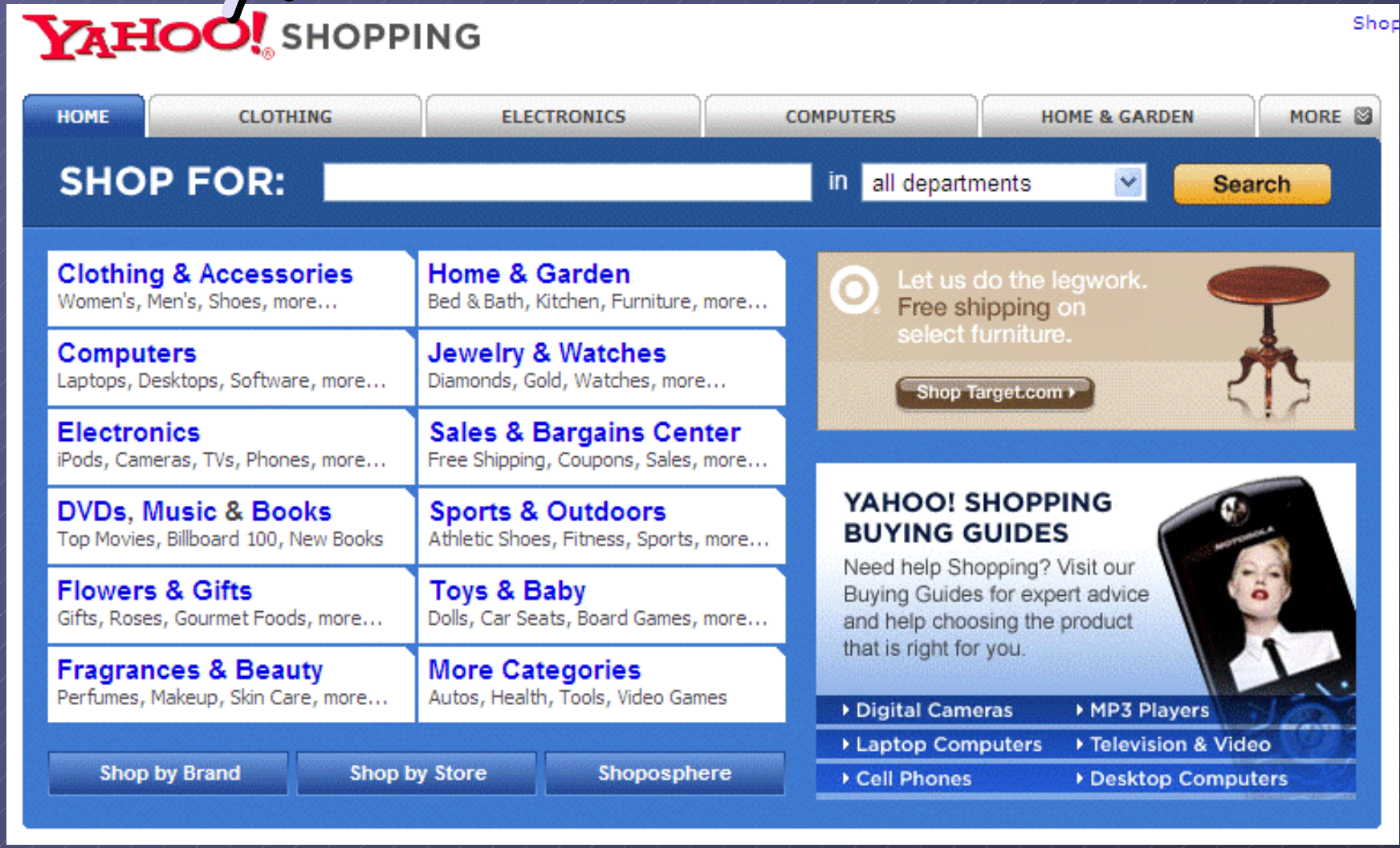
Network effects and recommender services most effective at genre or subcategory level

Let's do the basic economics of variety: demand for, supply of

Why the long tail now?

Didn't consumers demand variety before?

Need to know about and find variety.



Source: http://shopping.yahoo.com/

Other big changes are in cost of *providing* variety, which are…?

Suppose there are distribution fixed costs.

$\pi = (p\text{-}mc)Q - F$

Offer if expect $\pi > 0$ ➔ $Q > F/(p\text{-}mc)$

If F gets smaller, more goods offered

So, fixed distribution costs favor mass market hits. Lower fixed costs favor niche goods.

See MacKie-Mason, Shenker and Varian, "Service Architecture and Content Provision: The Network Provider as Editor," in Telecommunications Policy, vol. 20, no. 3, April 1996: 203-17.

Image Source: Telecommunications Policy, vol. 20, no. 3, April 1996

Does the marketing the Long Tail imply a shorter head?

Does marketing to the Long tail increase demand or just shift it?

Anderson claims:

"Some forms of entertainment, such as music, are 'non-rivalrous' for attention, which is to say you can consume them while you're doing something else."

Agree?

Source: Chris Anderson (2006), The Long Tail (Hyperion).

Should prices be higher or lower for products down the tail?

1. Make everything available
2. Help me find it

Anderson: Long Tail spawns two imperatives.

Do we need an economics of abundance?

Anderson's nine rules

Source: Chris Anderson (2006), The Long Tail (Hyperion).

1. Move inventory way in…or way out

2. Let customers do the work

3. One distribution method doesn't fit all

4. One product doesn't fit all

5. One price doesn't fit all

6. Share information (lose control)

7. Think “and”, not “or”

8. Trust the market to do your job

9. Understand the power of free