

## CS 103: Lecture 17 Power Laws and the Rich-Get-Richer Phenomenon

Dan Sheldon

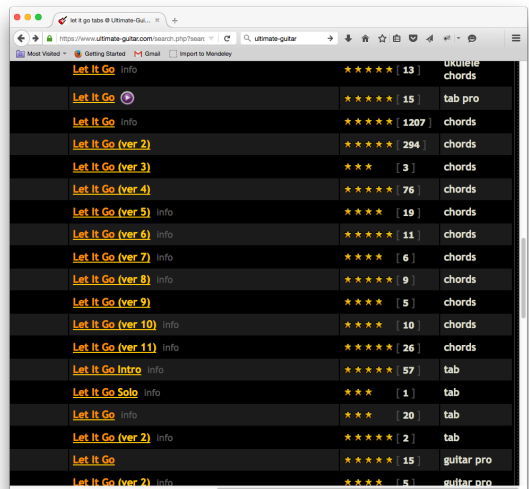
December 10, 2015

## Announcements

- ▶ No HW 7
- ▶ Tuesday: final review
- ▶ Exam will be open book

## Popularity

How is *popularity* distributed? How does it arise?



Version	Stars	Count	Category
Let It Go	★★★★★	13	ukulele chords
Let It Go	★★★★★	15	tab pro
Let It Go	★★★★★	1207	chords
Let It Go (ver 2)	★★★★★	294	chords
Let It Go (ver 3)	★★★	3	chords
Let It Go (ver 4)	★★★★★	76	chords
Let It Go (ver 5)	★★★★	19	chords
Let It Go (ver 6)	★★★★★	11	chords
Let It Go (ver 7)	★★★★	6	chords
Let It Go (ver 8)	★★★★★	9	chords
Let It Go (ver 9)	★★★★	5	chords
Let It Go (ver 10)	★★★★	10	chords
Let It Go (ver 11)	★★★★★	26	chords
Let It Go Intro	★★★★★	57	tab
Let It Go Solo	★★★	1	tab
Let It Go	★★★	20	tab
Let It Go (ver 2)	★★★★★	2	tab
Let It Go	★★★★★	15	guitar pro
Let It Go (ver 2)	★★★★	5	guitar pro

## Popularity

How is *popularity* distributed?

Extreme imbalances

- ▶ Harry Potter and the Philosopher's Stone: 107 million sales
- ▶ Katy Perry: 78M twitter followers
- ▶ Adele's '25': 1.11M sales in 2 weeks

Small fraction of people/things gain immense popularity. What does distribution of "popularity" look like? How does it arise?

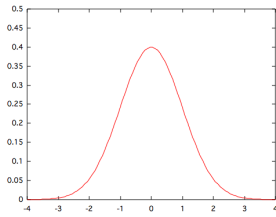
## A Model System: Web Pages

Links to a page = votes for its popularity

What fraction of pages have  $k$  links?

Let  $f(k) = \#$  of pages with  $k$  links. What is the form of this function?

## First Guess: Normal Distribution



**Central Limit Theorem:** if a quantity is the result of adding many small and *independent* effects, it will follow a normal distribution

**Examples:** test scores, plant heights

## Popularity: Power Law Distribution

Do we see a normal distribution? No.

**Sketch distribution on board**

Power law: number of items with  $k$  "likes" proportional to  $\frac{1}{k^r}$

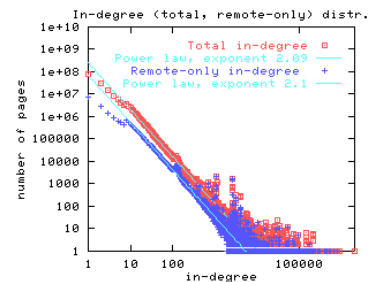
## Power Laws

Power laws are a robust natural phenomenon:

- ▶ # web pages with  $k$  links  $\approx \frac{c}{k^2}$
- ▶ # books read by  $k$  people  $\approx \frac{c}{k^3}$
- ▶ # of scientific papers that receive  $k$  citations  $\approx \frac{1}{k^3}$

What explains this?

## Finding Power Laws: The Log-Log Plot



**Board work:** a power law is a line with slope  $-r$  on a log-log plot

## Copying and Power Laws

How do power laws arise?

Central limit theorem = many small independent effects

Popularity is *not* independent. Follow someone on Twitter, read a book, watch a movie: based on what others do/recommend.

"Rich-get-richer". How can we model this?

## Rich-Get-Richer: a Simple Model

Return to web pages. **Example on board**

- ▶ Pages created in order  $1, 2, \dots, n$
- ▶ Each page creates one link
- ▶ When page  $j$  is created:
  - ▶ With probability  $p$ , link to a previous page chosen uniformly at random
  - ▶ With probability  $1 - p$ , select a previous page uniformly at random, then **copy it's link**

**Fact:** this leads to a power law. The fraction of pages with  $k$  links is roughly  $1/k^c$ , where  $c$  depends on  $p$  **Discuss.**

## Power Laws and The "Long Tail"

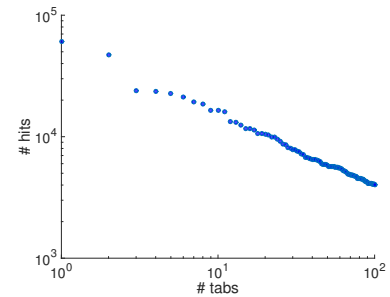
The Long Tail: popular manifestation of the same phenomenon w/ relevance to retail and marketing

### Board work

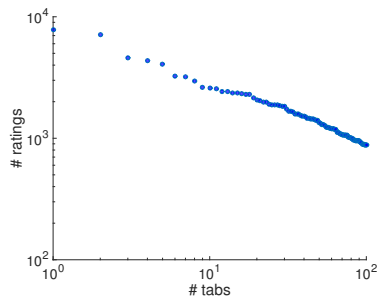
- ▶ "Head": small number of immensely popular items
- ▶ "Tail": huge number of less popular items ("niche")

Discuss: \* Retail strategies. "Brick and mortar" vs. online. \* Recommendation systems

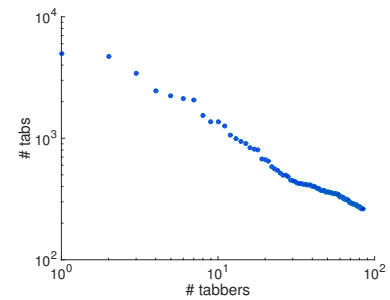
## You Can Find Long Tails Everywhere



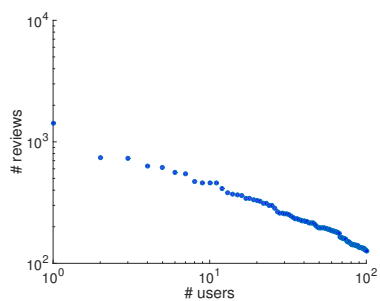
## You Can Find Long Tails Everywhere



## You Can Find Long Tails Everywhere



## You Can Find Long Tails Everywhere



## Popularity and Luck

Salgankik, Dodds, Watts 2006, *Science*

- ▶ Experiment setup **screenshot**
- ▶ Users assigned to one of 8 parallel sites w/ download counts starting at zero
- ▶ 9th version of site: no download counts
- ▶ Final popularity varied dramatically within 8 trials
- ▶ Much more equitable in version with no download counts