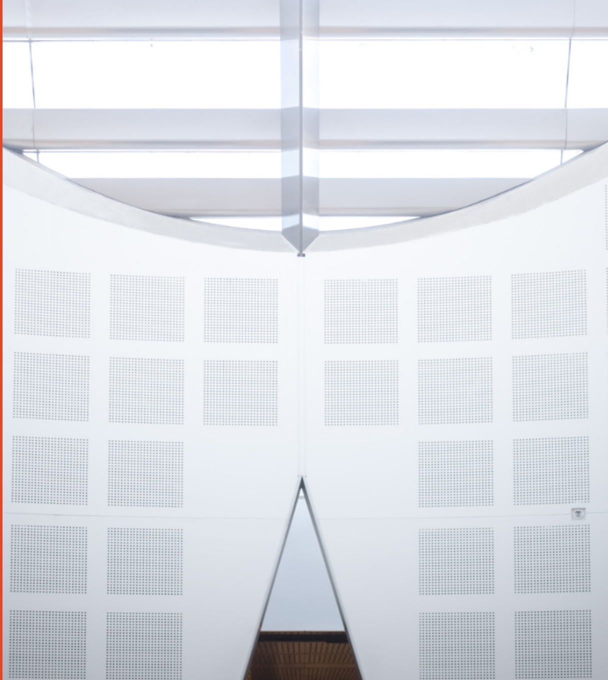
COMP5313/COMP4313 - Large Scale Networks

Week 10a: Power Laws

Lijun Chang

May 8, 2025

THE UNIVERSITY OF SYDNEY



Introduction

- We talked about evolution through decisions
  - Information cascade can depend on the outcome of few initial decisions
  - A technology can win simply because it reaches an audience before its competitors
- Let us talk now about a model of uncertain evolution
  - Various quantities, like popularity, have highly skewed distributions
  - How can this be explained?

Outline

Power Law Distribution of Popularity

The Preferential Attachment Model

Unpredictability of the Rich-get-richer Effect

Popularity

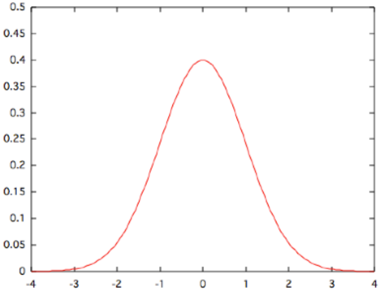
- Popularity is a phenomenon characterized by extreme imbalances
  - While most people are known only by people in their immediate social circles
  - Very few people achieve global name recognition
- How can we quantify these imbalances?
  - It is hard to answer these questions for people popularity
  - Let's try to answer these questions for Web page popularity

Popularity

- Let us define the popularity of a Web page as the number of its in-links
  - The links pointing to a given page are referred to as the in-links of the page
- To quantify the imbalance of popularity, let us consider the distribution of the number of in-links
  - As a function of k, what fraction of pages on the Web have k in-links?
  - k translates into popularity: the higher k is, the more popular the page

Distribution of Popularity

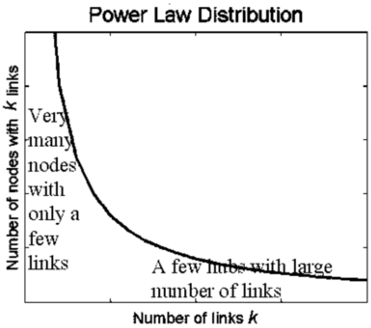
- What could be the distribution?



- The normal distribution is popular (cf. Central Limit Theorem)
  - The sum of independent random quantities follows the normal distribution
  - k would be normally distributed if pages would connect independently at random

Power Law Distribution of Popularity

- The distribution of links on the Web is different from a normal distribution. [1]
  - The fraction of Web pages that have k in-links is approximately $1/k^2$ (More precisely, the exponent on k is generally a number slightly larger than 2)



[1] A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins, J. Wiener. Graph structure in the web. Computer networks, vol. 33, no. 1-6, 2000.

Power Law Distribution of Popularity

- How does it differ from the normal distribution?
  - $1/k^2$ decreases much more slowly as k increases than in the normal distribution (there are more pages with a very high numbers of in-links)
  - $1/k^2$ is only $1/1000000$ for $k = 1000$, while $1/2^k$ is unimaginably low

- A power law is a function that decreases as k increases to some fixed power, such as $1/k^2$ in the present case

$\Longrightarrow$ There is an extreme imbalance in the distribution of in-links on Web pages

Power Law Distribution of Popularity

- Similar power laws exist:
  - The fraction of telephone numbers that receive k calls per day is $O(1/k^2)$
  - The fraction of books that are bought by k people is $O(1/k^3)$
  - The fraction of scientific papers that receive k citations is $O(1/k^3)$
  - ...
- Hence, if someone gives you a table showing the number of monthly downloads for each song at a large online music site, then it is worth testing whether it is approximately a power law $1/k^c$ for some c, and if so to estimate c.

Power Law Distribution of Popularity

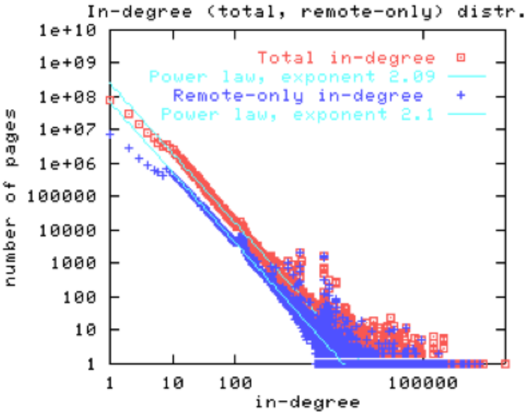
- How to test that some dataset is power law?
- Let $f(k)$ be the fraction of items that have value k
- Suppose you want to know whether the equation $f(k) = a/k^c$ approximately holds
- If we apply the log to both sides, we have:

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

$\Longrightarrow$ If we plot $\log f(k)$ as a function of $\log k$, then we should have a straight line with $-c$ the slope and $\log a$ the y-intercept

Power Law

- A power law distribution, like the Web page in-links, shows up as a straight line on a log-log plot



Outline

The Preferential Attachment Model

- Here is a simple model for the creation of links among Web pages
  1. Pages are created in order and named $1, 2, 3, \cdots, n$
  2. When page j is created, it produces a link to an earlier Web page by choosing between actions (a) and (b) below according to the following probabilistic rule (controlled by a single number p between 0 and 1):
     a) With probability p, page j chooses a page i uniformly at random from among all earlier pages and creates a link to 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 a page i points to

  (This described the creation of a single link from page j; one can repeat this process to obtain multiple, independently generated links from page j)

The Preferential Attachment Model

- If we repeat for many pages, the fraction of pages with k in-links follow a power law distribution $1/k^c$
- Step 2(b) is the key: j copies the behavior of node i instead of linking i
- We could have replaced Step 2(b) by:
  b) With probability $1-p$, page j chooses a page i with probability proportional to i's current number of in-links, and creates a link to i

$\Longrightarrow$ The probability that i's popularity increases is proportional to i's popularity

$\Longrightarrow$ Preferential attachment: links are formed "preferentially" to pages that already have high popularity. [2]

[2] A.-L. Barabasi, R. Albert. Emergence of scaling in random networks. Science, vol. 286, no. 5439, 1999.

Outline

Power Law Distribution of Popularity

The Preferential Attachment Model

Unpredictability of the Rich-get-richer Effect

Rich Get Richer

- Popularity grows at a rate proportional to its current value
- The populations of cities:
  - The fraction of cities with population k is roughly $1/k^c$ for some constant c. [3]
  - If we assume that cities are created at different times and grow at a rate proportional to its current size simply as a result of people having children, then we have roughly the same model
- Citations: new citations to a paper are proportional to the number it already has
- Sociology: Matthew effect
  - https://en.wikipedia.org/wiki/Matthew_effect

[3] H. A. Simon. On a class of skew distribution functions. Biometrika, vol. 42, no. 3/4, 1955

Unpredictability of the Rich-get-richer Effect

- For a Web page, a book, a song, or any other object of popular attention, the initial phase of its rise to popularity is relatively fragile.
- Once popularity is established, the rich-get-richer dynamics are likely to push it even higher.
- But getting this rich-get-richer process ignited is a precarious process, full of potential accidents and near misses.

Unpredictability of the Rich-get-richer Effect

- Salgankik, Dodds, Watts created a music download site
  - With 48 obscure songs of varying quality written by actual performing groups
  - Visitors were presented with a list of the songs and could listen to them
  - Each visitor would see a "download count" for each song
  - At the end, the visitor was proposed to download the song that she liked
- Upon arrival, visitors were redirected to one of 8 copies of the site
  - These copies were initially the same with download count set to 0
  - These copies evolved differently as visitors arrived
- The market share of the different songs varied considerably across the different parallel copies