CSC200: Lecture 33

Allan Borodin

Annoucements and Todays Agenda

- Announcements
  1. Quiz 7 takes place March 11, 2016
  2. Assignment 4 is due March 30, 2016. Initial questions will be posted by the end of the week (and hopefully by tonight).
  3. Usual tutorial this Friday
- Todays agenda:
  1. Finish discussion of impact of heavy tail in (for example) power law distributions.
  2. Briefly discuss recommendation systems
  3. Begin discussion of influence diffusion in social networks - Chapter 19 of text.

Why the Long Tail Matters?

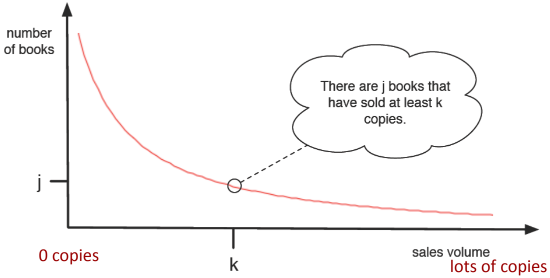
- Amazon was among the first to take advantage of the “long tail” so let’s discuss books (one might argue Sears-Roebuck did in 1880s)
- In Bricks-and-mortar bookstores, which types of books you find the most? “Hits” or “Niche books”.
- Bricks-and-mortar bookstores have limited capacity to hold inventory
  - If you can only stock 5,000-10,000 items, you are going to be sure they appeal to a wide audience.
  - Limits availability to readers, curtails production (authors have no outlets)
- What does an Amazon bring to the table?
  - Can store huge inventories, because of worldwide client base
  - Facilitated by internet, search tools, low-cost shipping, print-on-demand technologies, etc.
  - So does that change the nature of what we buy?

Impact of the Long Tail

- Turns out that consumer demand—when inventories are nearly "unlimited"—seem to satisfy power laws
- Let f(k) be the number of books that have been sold k copies, then

$$F(k) = \sum_{i \geq k} f(k)$$

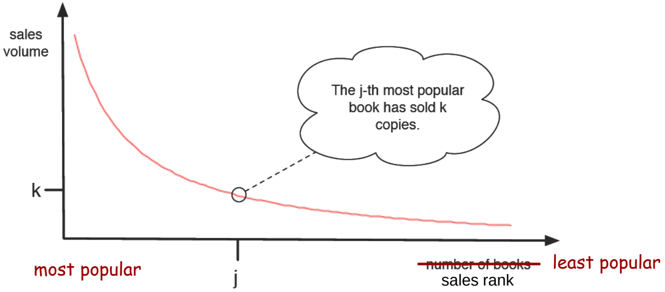
- F(k) shows the number of books that have been sold at least k copies.



Distribution of Popularity: How many books have sold at least k copies? E&K Ch.18, Fig.18.3.

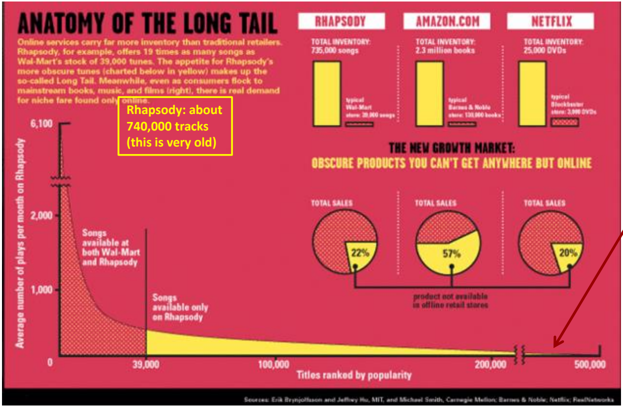
Impact of the Long Tail

- … or like this (just flip graph around): “sales rank” vs. sales volume



Distribution of Popularity: How many copies of the jth most popular book have been sold? E&K Ch.18, Fig.18.4.

Anatomy of the Long Tail



Even songs ranked 400,000th in Rhapsody receive a few dozen plays per month!

Chris Anderson, "The Long Tail"
http://www.wired.com/wired/archive/12.10/tail.html

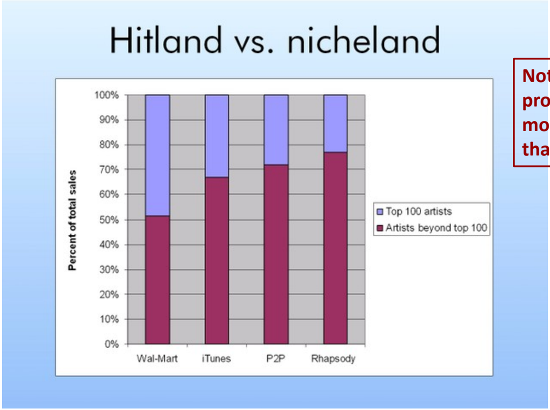
Online vs. Bricks and Mortar

- Let’s assume (!) you can carry millions of items
  - ``Currently’’ Rhapsody claims to carry 11 million tracks (2012)
  - Amazon has about 850,000 titles available digitally (e-books), and has sold approximately 7.5M distinct titles (not sure how many actually “in stock”, and since they linked to used sellers for out of print, it’s probably a flexible figure);
- Say B&M retailer can carry 50,000-150,000 items (the most popular)
- Let’s assume that with great search tools (directed and serendipitous) and recommender engines that people can find most anything that they might be interested in

Online vs. Bricks and Mortar

- OL sells roughly the same number of items as B&M in top 50,000
  - Sales of top sellers high, but tails quickly: avg. sales volume, say, 500
  - Total sales volume is *500*50,000 = 2.5M units*
- Now look at the extra stuff sold only by OL:
  - From 50,000 to 200,000, avg. sales volume 20: *total = 3M units*
  - From 200,000 to 500,000 avg. sales volume 5: *total = 1.5M units*
  - From 500,000 to 1M avg. sales volume 2: *total = 1M units*
  - From 1M to 2M, avg. sales volume 1: *total = 1M units*
- While the numbers are fictitious, the point is clear: if you can afford to sell down the long tail, you will be able to sell more number of units.
- But be cautious in any general conclusions:
  - Stocking, information management, shipping, etc. must be accounted for.
  - Search and accessibility must be accounted for.
  - Must address possibility of self-cannibalizing sales.

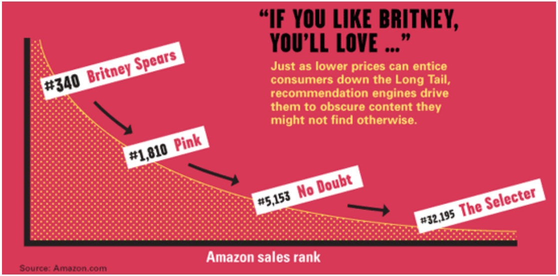
Percentage of Sales: Top 100 Artists vs. Rest



Note: Dated, probably even more skewed than this now

Chris Anderson, http://www.longtail.com/

Recommender Systems



Chris Anderson, "The Long Tail"
http://www.wired.com/wired/archive/12.10/tail.html

Need to be cautious in such conclusions!

Recommender Systems

- Ability to (cheaply) carry and deliver large inventories not enough
- Consumers must have the ability to navigate and explore!
- Search tools are part of it
  - Search engines (e.g. Google) can help find things: relevance/popularity
  - But with cultural products tastes are highly varied
- *Recommender systems*
  - attempt to predict what someone likes based on their “tastes”
  1. *Content-based recommendation*
     - based on content attributes of product (car, apartment, etc.)
  2. *Collaborative recommendation (collaborative filtering)*
     - predict interests/likes based on ratings or consumption of others whose “tastes” appear to be similar
     - increasingly common, especially in media (music, books, movies, …)
     - can you predict if you’ll like a song based on *genre, BPM, …*

Collaborative Filtering

- Suppose you want movie recommendations from Netflix
- Idea (very crudely)
  - Suppose you've rated/ranked/bought certain movies
  - There are thousands of movies you've never seen, even heard of
    - But millions of other people other have watched/rated/bought them
  - Define *overlap(A, you)* to be movies *both you* and *A* have rated
  - Define *unique(A, you)* to be movies *A* has rated *but you haven't seen*
  - If you rated movies in the set *overlap(A, you)* in a way that is very similar to *A*, maybe *A*'s tastes are similar to yours; so *A*'s ratings of movies in the set *unique(A, you)* can be used to predict whether or not *you* would like these movies
  - Of course, *unique(A, you)* might be a *small set* (*A* can only watch/rate a small fraction of all movies); and *A* may have similar tastes in comedies but different tastes in horror…
  - So you average results over millions of users' ratings to get very broad and very robust results!

Collaborative Filtering Example

Users/Movies	Titanic	Avatar	Terminator 1	Matrix
Alice	5	1	3	x
Bob	x	4	4	1
Carol	4	2	2	5
David	x	5	5	2
Edward	?	2	3	4

- x: missing ratings
- ?: a rating that we want to predict
- Alice, Carol, Edward have similar ratings for Avatar and Terminator.
- So, Edward's rating for Titanic should be close Carol's and Alice's ratings for Titanic. Let's simply average their ratings and predict Edward's rating as the average value of 4+5/2 = 4.5

Many Collaborative Filtering Models

- **Similarity-based methods**
  - Define distance between you and any other user to be a measure of divergence between your ratings and theirs on the overlap set
  - Then average everybody's ratings for a movie M, but weighted by their distance to you (closer users' ratings have greater weight in prediction)
- **Latent factor methods**
  - Use sophisticated machine learning and statistical techniques to identify latent (hidden) features of both users and movies that can be used to predict someone's rating of a movie
  - E.g., matrix factorization methods
- **Netflix Prize**
  - Awarded a $1M prize in 2009 to research group that was able to improve their prediction accuracy by 10%
  - 41,000 teams from 186 countries competed over a period of 2-3 years

Collaborative Filtering, Power Laws and the Long Tail

- Thought exercise:
  - *Content-based recommendation* vs. *collaborative filtering*: what impact will each of these have on the diversity of products consumed by users, and how will these impact the possibility of a long tail?
  - Consider the potential that each of them has to correlate people’s choices and increase/decrease product diversity
  - And for what types of product will content-based vs. collaborative filtering each be best suited?

Influence spread in a social network

- We begin a study of the spread/diffusion of products/influence in a social network (Chapter 19) in contrast to population wide spread phenomena as studied in Chapters 16, 17 and 18.

- The goal (as before) is to qualitatively understand the process in a highly stylized (but hopefully still interesting) setting.

- We will (as usual) be interested in what kind of general conclusions can be inferred from such an understanding?

Recalling population wide effects

- In Chapters 16 (herding or informational effects), 17 (direct benefit effects), and 18 (rich get richer models) we did not have a social network per se.
- These chapters dealt with population wide effects. Although :
  - One can construe Chapter 16 as taking place in a network where the ith individual is connected to all $i - 1$ previous individuals.
  - Chapter 17 can be construed as taking place in a network where everyone is directly connected (the network is a complete graph).
  - Chapter 18 studies random processes by which networks can grow and and one can think of situations where the resulting network is a social network.
- But still . . . these are basically population wide effects absent from an existing social network.

Social network effects

- Now we wish to consider an existing social network where edges (ties) between individuals represent some sort of friendship/relationship.
- This takes us back to concepts introduced in Chapters 3 and 4.
- There we saw the contrast between
  - homophily (we tend to be friends with people of similar backgrounds, geography, interests)
  - social influence (we join clubs, are influenced) by our friends/relations.

Models of influence spread/diffusion

- One of the most important themes of the text (and CSC 200) is that we construct models to gain insight.
  - Our models are often (maybe always) very simplified given the complexity of real social and economic networks.
  - There is always a tradeoff between the adherence to reality and our ability to analyze and gain insight.
- How we model diffusion in a social network will clearly depend on what product, idea, membership, etc. we are considering.
- There are many assumptions as to how products, ideas, influence are spread in a social network and what are the set of individual alternatives.
- The main emphasis in Chapter 19 is on a very simple process of diffusion where each person has 2 alternative decisions:
  1. stay with a current "product" B
  2. or switch to a (new) product A.

A simple model of diffusion in a social network

- Let's assume that we are making decisions based on the direct benefit of being coordinated with our friends beyond any intinsic value associated with the decision (e.g. when the decision is the purchase of an item).

- A standard example is what laptop or cell phone we decide to buy to the extent that we are mostly influenced by our friends rather than by general population wide usage. What influences you most? Friends or general population benefits?

A simple model of diffusion in a social network

- Let's assume that we are making decisions based on the direct benefit of being coordinated with our friends beyond any intinsic value associated with the decision (e.g. when the decision is the purchase of an item).
- A standard example is what laptop or cell phone we decide to buy to the extent that we are mostly influenced by our friends rather than by general population wide usage. What influences you most? Friends or general population benefits?
  - Maybe, choosing between two weekly television shows that occur at the same time or who to vote for is a better example.
- In fact, the model given in this chapter dictates that certain decisions (i.e. to change from B to A) are irreversible.
  - The text calls this a "progressive process" in the sense that it progresses in only one direction. Any good examples of truly (or essentially) irreversible decisions?
  - Amirali suggested the decision to get a tatoo. Joel suggested a parents decision to have a male child circumcised.

A threshold model for spread

- We assume that some number of individuals are enticed (at some time $t = 0$) to adopt a new product A.

- Outside of these "initial adopters", we assume all other individuals in the network are initially using a different product B (or equivalently this is the first product in a given market).

- This is not really a competitive influence model as B is not really competing. (More comments later.)

- The first model we consider for diffusion is that every node v has a threshold q (in absolute or relative terms) for how many of its neighbors must have adopted product A before v adopts A.

Threshold model (continued)

- For simplicity the text initially assumes that every node v (i.e. individual) in the network has the same threshold but then later explains how to deal with individual thresholds.
- If at some time t, the threshold for a node v has been achieved, then by time time $t + 1$, v will adopt product A.
- If the threshold has not been reached then v decides not to adopt A at this time.

Note

Although it is not explicitly stated, the initial adopters never reverse their adoption.

- Given these model assumptions, adopting A is irreversible for all nodes in the network.

Determining a (relative) threshold

- One way (some might say is usually the best way) to reason about a plausible threshold for a node is to view one's decision in economic terms.
- Specifically for every edge (v, w) in the network suppose
  - There is payoff a to v and w if both v and w have adopted product A.
  - There is payoff b to v and w if both v and w have adopted product B.
  - A zero payoff when v and w do not currently utilize the same product.
- This determines a simple coordination game.

		w: A	w: B
v	A	a, a	$0, 0$
	B	$0, 0$	b, b

Figure : $A - B$ coordination [Fig 19.1, E&K]

Coordination game induces threshold

- Suppose node v has not yet adopted A at time t, but a fraction p of the $d(v)$ neighbors of v have already adopted A, then:
  - By switching, the payoff to v is $p \times d(v) \times a$.
  - By staying with B, v has payoff $(1-p) \times d(v) \times b$.

- Thus node v will switch to A if

$$p \times d(v) \times a \geq (1-p) \times d(v) \times b$$

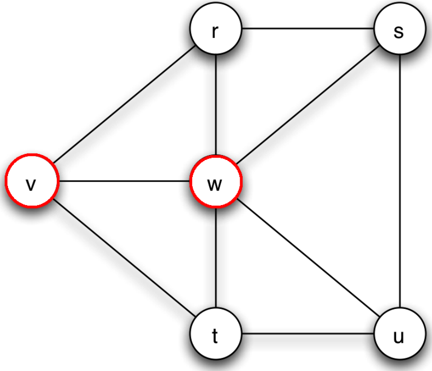
  (for simplicity say v switches when payoffs are equal).

- This is then equivalent to saying that v will switch whenever p is at least $\frac{b}{a+b} = q$ which is then the relative threshold.

- That is, whenever there is at least a (threshold) fraction q of the neighbours of node v that have adopted A, then v will also adopt A.

The process unfolds (example: $a = 3$ and $b = 2$)

[Fig 19.3, E&K]

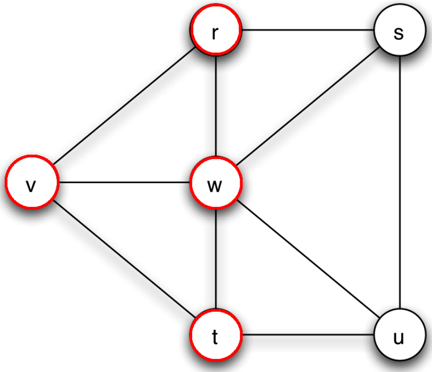


$t = 0$

- A node adopts A if and only if the threshold $q = \frac{b}{a+b} = 2/5$ is reached.
- Two nodes v and w are initial adopters.

The process unfolds (example: $a = 3$ and $b = 2$)

[Fig 19.3, E&K]

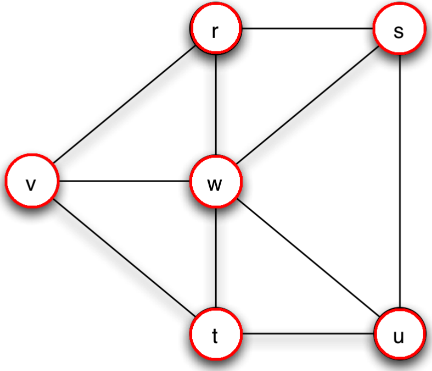


$t = 1$

- A node adopts A if and only if the threshold $q = \frac{b}{a+b} = 2/5$ is reached.
- Two nodes v and w are initial adopters.

The process unfolds (example: $a = 3$ and $b = 2$)

[Fig 19.3, E&K]

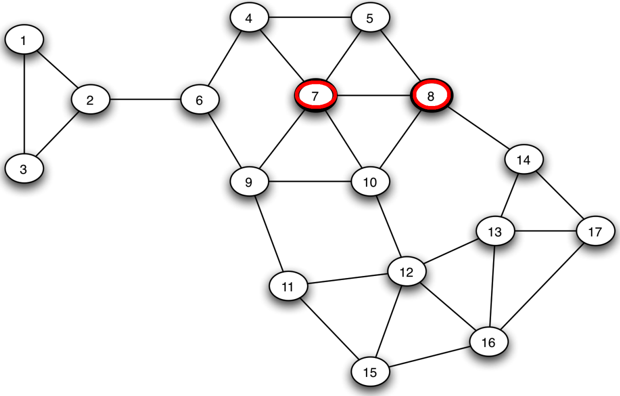


$t = 2$

- A node adopts A if and only if the threshold $q = \frac{b}{a+b} = 2/5$ is reached.
- Two nodes v and w are initial adopters.

Complete cascades vs tightly-knit communities (example: $a = 3$, $b = 2$, $q = 2/5$)

- The previous example showed a complete cascade where all nodes eventually adopt A.
- In the next example, "tightly-knit communities" block the spread.

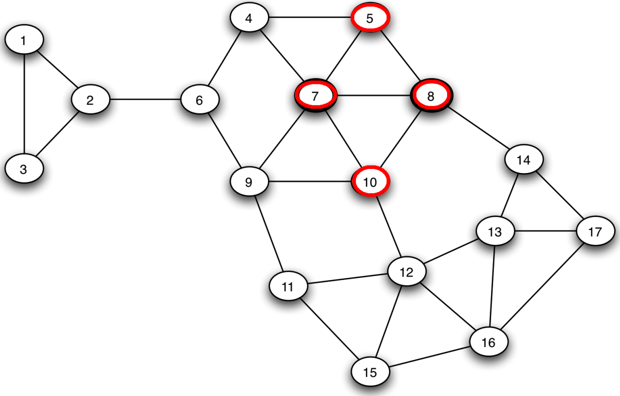


$t = 0$

[Fig 19.4, E&K]

Complete cascades vs tightly-knit communities (example: $a = 3$, $b = 2$, $q = 2/5$)

- The previous example showed a complete cascade where all nodes eventually adopt A.
- In the next example, "tightly-knit communities" block the spread.

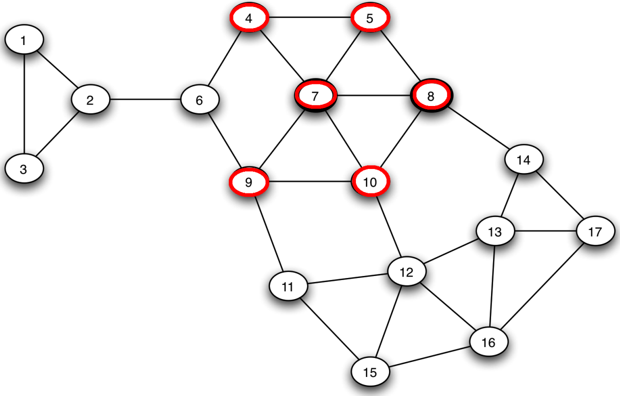


$t = 1$

[Fig 19.4, E&K]

Complete cascades vs tightly-knit communities (example: $a = 3$, $b = 2$, $q = 2/5$)

- The previous example showed a complete cascade where all nodes eventually adopt A.
- In the next example, "tightly-knit communities" block the spread.

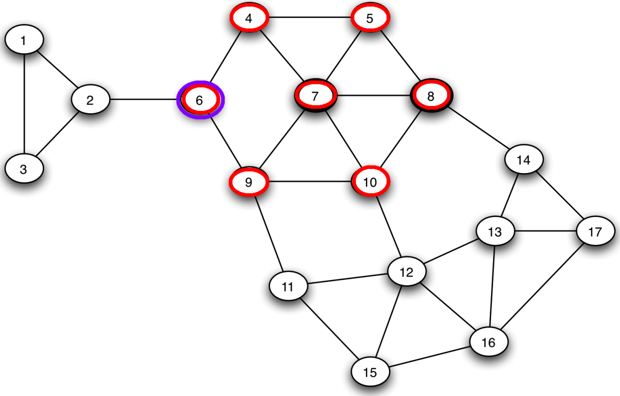


$t = 2$

[Fig 19.4, E&K]

Complete cascades vs tightly-knit communities (example: $a = 3$, $b = 2$, $q = 2/5$)

- The previous example showed a complete cascade where all nodes eventually adopt A.
- In the next example, "tightly-knit communities" block the spread.



$t = 3$

[Fig 19.4, E&K]