NBA 600: Day 15
Online Search
116 March 2004

Daniel Huttenlocher
Today’s Class

- Finish up network effects topic from last week
- Searching, browsing, navigating
- Reading “Beyond Google”
  No longer available on Tech Review Site
  http://www.technewsworld.com/perl/story/33020.html
Network Effects

- Online communities often exhibit network effects
  - Value proportional to number of other people

- Metcalfe’s law
  - Value of network proportional to square of its number of users, \( n^2 \)
    - Value to each user is (proportional to) \( n \)
    - Times \( n \) users = \( n^2 \)

- Network effects were often claimed in the Internet bubble
  - But did not materialize
Classical Network Effects

- Telephone network
  - Value higher if more subscribers
  - But inter-operability across providers means individual firm does not benefit from effect

- Standard interchange formats
  - Fax, MS Word, Adobe Acrobat, VHS video tape
  - Again often standardized with little return to a given firm unless license fees for standard
    - License fees may inhibit adoption however

- Value because “everyone uses it”
Switching Costs

- Network effects lead to high collective switching costs
  - Worse than individual switching cost because of incompatibility with others
- Online communities can exhibit this
  - Auctions?
  - Gaming?
  - Reviews?
- What other Internet services exhibit this?
- Amazon vs Ebay?
Initiating Network Effects

- “Chicken and egg” problem
  - Value comes from critical mass
  - First mover can be important
    - But still need to get boot-strapped
- Financial markets have similar issue
  - Enough participants to get liquidity
- If enough improvement an old network can be displaced
  - E.g., standards-based networks VHS to DVD
- Alliance of those who stand to benefit
Role of Search Online

- People use the Internet for many things
  - News, entertainment, research, discussion, communication, commentary, opinion, shopping
- Often looking for things online
  - Highly variable degree of specificity
- Many times easier to find things online than offline
  - But issues of authority, trust, brand, convenience
  - Increasingly too much clutter online
Search and Top Web Sites

1. Yahoo, 30% reach (portal)
2. MSN 23% reach (portal)
3. Google 15% reach (pure search)
4. Microsoft
5. Ebay 4% reach (commerce)
6. Passport.net
7. Amazon 3% reach (commerce)
8. Go (Disney)
9. Doubleclick
10. CNN 2% reach source: alexa.com
How Alexa Computes Rankings

- Have a free toolbar that helps people navigate the Web
  - Access to Google search
  - Related sites based on other Alexa users
  - Notes every page visited by each user

- Owned and promoted by Amazon
  - Millions of active users

- Aggregate information about sites these users visit
  - For rankings use two things, visit to site compared to other sites and page views on site
Importance of Search

- Google has increased from fifth to third place site over past year
  - While portals Yahoo and MSN have stayed steady in first and second
  - Only site with reach approaching that of major portals

- Search alone increasingly appears able to compete with directories, editorials and recommendations
  - Yahoo and Microsoft have both recently undertaken large initiatives in search arena
Search in the Value Chain

- Search based on keywords or descriptions
  - Part of many activities
    - Can view many transactions as search, payment, fulfillment

- Within site, across sites within category, across sites and across categories
  - Auction sites like ebay
  - Ecommerce sites like Amazon
  - Travel sites like Expedia
  - File sharing networks like Kazaa
  - News sites like CNN
Different Views on Search

- Ebay has long guarded their listings
  - Only way to search listings is on their site
    - Have defended their right to restrict
- Amazon has open API’s that enable search of their site (common in online retail)
  - Google uses for search engine (~7M pages)
  - Various independent vendors use for finding items on their site
- Advantages and disadvantages of each?
  - Audience size
  - Where your customers turn to first
Ebay and Search

- Ebay has had a substantial impact on classified advertising
  - A form of search (for buyers and sellers)
- Larger audience, easier to find listings
- Auction format differs from classified ad listings
- More broadly, newspaper online classified also generally controlled access
  - Not made accessible to search engines
  - No payment processing in classified ads
Some Recent Events

- Yahoo stopped using Google for search and replaced with own technology
  - Based on Inktomi and Overture acquisitions
- Amazon established visible R&D group focused on search
- Gates acknowledged Microsoft mistakenly did not appreciate value of search
  - Now focused on Google as challenging Web portals such as MSN
- Patent disputes between Yahoo, Google, others
Anatomy of a Search Engine

- Three main components
  - Crawler or spider
    - Finds sites and pages on the Web
    - Usually by following links from already known pages
  - Index
    - Fast access to pages based on keywords
  - Query processing
    - Displaying pages most relevant to a user request
    - Some method for ranking by relevance
      - E.g., Google’s PageRank
Web Crawlers

- Web pages are “unstructured data”
  - Text, layout, images, movies, audio
  - One key aspect of layout is hyper-links
    - Identified by particular format

- Crawlers trace chains of hyper-links in search of new pages
  - Pages that are not already in the index

- Crawlers also check for known pages that have changed
  - Rate of checking for changes can be variable based on type of content or site
Page Indexing

- Efficiently map from query terms to web pages
  - Sometimes only URL, sometimes also content
    - “Cached” copy of page
- Which terms are important for the page
  - Initial indexing methods used just the text of the page
    - Relatively easy to manipulate
  - Current methods also rely heavily on text of referring pages
    - Near link to this page
Page Relevance

- Perhaps most important aspect to user is ordering of most relevant pages
  - Initial search engines used only text information
    - On page and perhaps referring pages
  - Methods such as Google’s PageRank and Kleinberg’s hubs & authorities use link structure
    - The importance of a page depends not only on its content but in how it is linked to
  - PageRank uses popularity – pages with more links to them ranked higher and in turn their votes for other pages count more
    - Question about such popularity based measures
Constant Tuning

- Search engines constantly tune their ranking methods
  - Balancing link structure, text content, rate of change, degree of linking
    - Partly to counteract “search engine optimizers”
- Often has substantial effect on rankings
  - Google’s changes now named in style of storms
  - Recent examples
    - Rise of weblogs, which have high degree of linking, led to re-balancing
    - Incorporation of Amazon’s pages into rankings
Revenue Models for Search

- For many years was a challenge
  - Paid inclusion allows sponsors to have their pages listed in response to certain queries
    - Can lead to lowering of search quality
    - FTC investigations about un-credited paid inclusion being misleading
    - Hard for advertisers to assess value
- Google has always argued against paid inclusion advertising
  - Sought ways of having “relevant ads” appear adjacent to search results
Bidding for Terms

- Overture launched a market for search terms
  - Advertisers bid a per-click fee for being listed for certain words or phrases
    - Placement ordered by bid
    - Gave advertisers “context” for their fees
- Google still sought mechanism for assessing relevance of ads
  - Focused on providing user with useful information even in ads
  - Highest bidder not necessarily relevant
AdWords

- Google’s mechanism of bidding for listings
  - Placement based on product of CTR and payment
    - Items that are more relevant tend to be placed higher due to more click-throughs
      - To address quality concern
  - Advertisers bid a maximum per-click fee they are willing to pay
    - Actual fee is lowest amount that is $.01 larger than next lowest ranked bidder
    - Can set daily expenditure budget as hard to predict actual costs