CSC200: Lecture 25

Allan Borodin

Announcements and Todays Agenda

- Announcement
  - There are now three questions on Assignment 3.
- Todays agenda
  - Quick review of Hubs and Authorities
  - Page Rank
  - The advanced material in chapter 14
  - Introducing Sponsored search (Chapter 15) as special case of Matching Markets (Chapter 10)

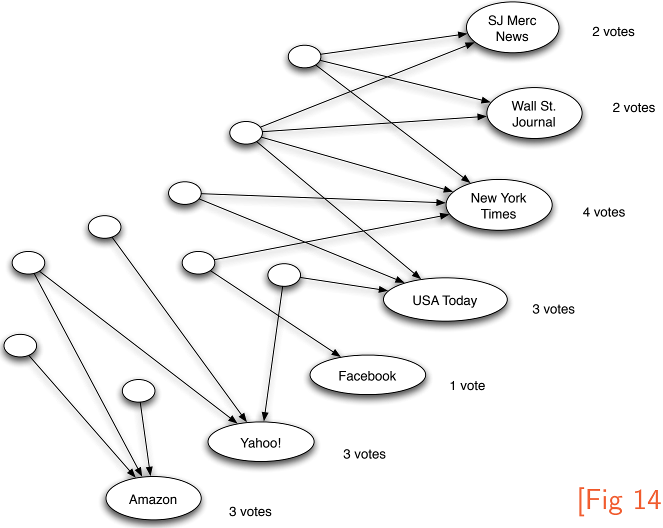
Hubs and Authorities

- The simplest way to utilize links to rank web pages would be to think of each link from A to B as an endorsement or vote by A for B.
- And then use the number (or weight) of endorsements as a key feature determining the rank. Of course, one would have to adjust such scores coming from say the same domain name.
- Even after adjusting for such "vote fixing", if Leo Komarov or Dion Phaneuf has a web site and a link suggesting where he buys his hockey equipment you might think that is more meaningful than if say where I buy hockey equipment.

Reinforcement of Hubs and Authorities.

- This then becomes the motivation (and seemingly circular reasoning) behind hubs and authorities.
- The best "authorities" on a subject (places to buy equipment) are being endorsed by the best hubs (people who know where to buy equipment).
- Similarly, the best hubs are those sites that recommend the best authorities. Conceptually the link structure induces a bipartite graph. The same web page can be both a hub and an authority.
- Comment: The word "authority" is not generally an accurate way to describe high ranking documents. These might better be referred to (barring other information) as the most relied upon sites. This is also different from "the most popular" sites which might better be measured in terms of the number of clicks being received. Hubs then are the most reliable endorsers.

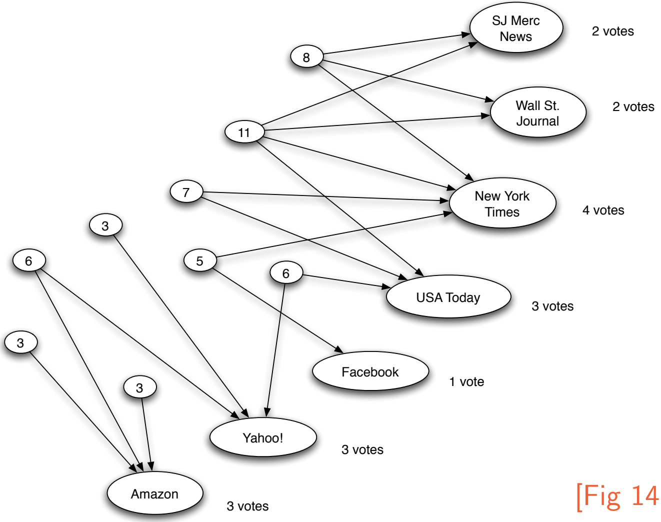
- The result of applying the authority update rule: for each page p, auth(p) is the sum of hub values (initially just the number) of hubs pointing to p.



[Fig 14.1, E&K]

Figure : Counting in-links to pages for the query "newspapers."

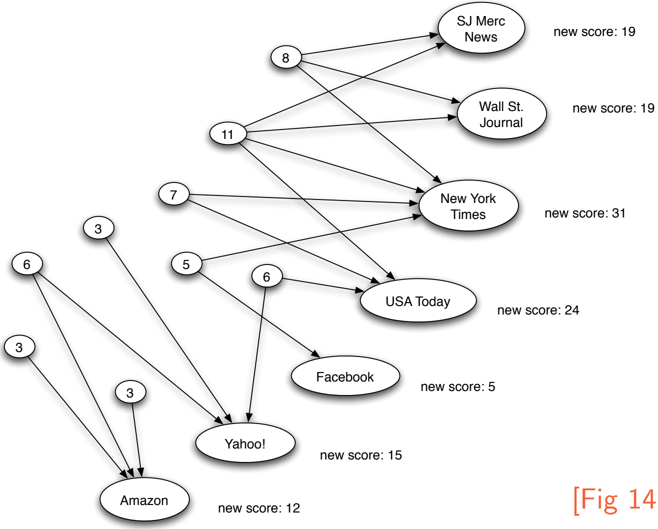
- Then to recalibrate hub values, we use the hub update rule: for each page p, hub(p) is the sum of values of all authorities that p points to.



[Fig 14.2, E&K]

Figure : Finding good lists for the query "newspapers": each page's value as a list is written as a number inside it.

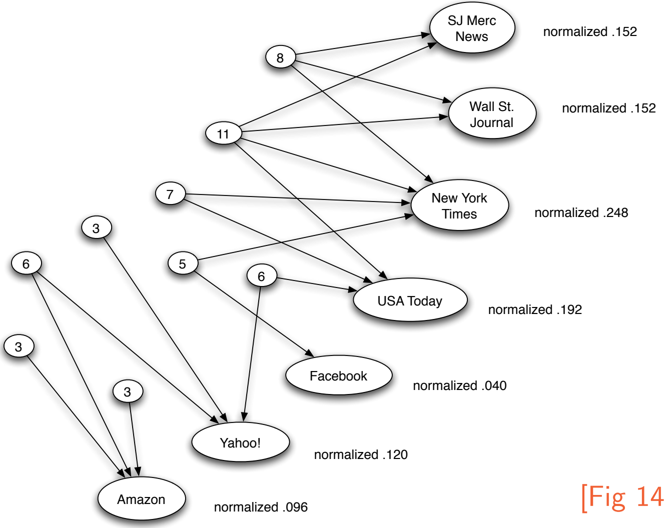
- Applying the authority update rule again we get figure 14.3.



[Fig 14.3, E&K]

Figure : Re-weighting votes for the query "newspapers": each of the labeled pages new score is equal to the sum of the values of all lists that point to it.

- Since we only care about the relative values of these numbers, both authority and hub scores can be normalized to sum to 1 (to allow convergence and avoid dealing with large numbers).

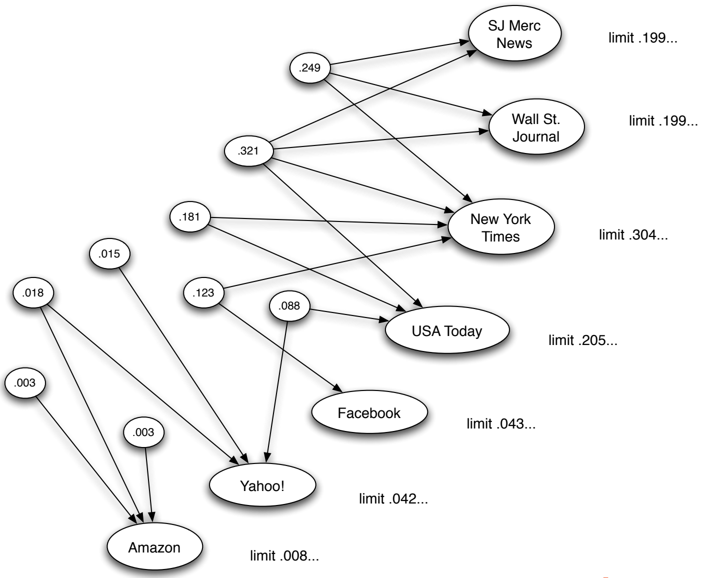


[Fig 14.4, E&K]

Figure : Re-weighting votes after normalizing for the query "newspapers".

Keep repeating a good idea

- Now having recalibrated and normalized both the authority and hub scores, we can continue this process to continue to refine these scores.
- That is, the hubs and authorities procedure is as follows:
  - Initialize all hub values (say to some positive vector perhaps depending on usage or content)
  - For sufficiently large k, perform the following k times
    - Apply authority update rule to each page
    - Apply hub update rule to each page
    - Normalize so that sum of A and H weights $= 1$.
- Using linear algebra, it can be shown (in Section 4.6) that these A and H normalized values will converge to a limit as $k \to \infty$ (which can be approximated by some sufficiently large k)!



[Fig 14.5, E&K]

Figure : Limiting hub and authority values for the query "newspapers".

Page Rank

- The motivation behind page rank is a somewhat different view of how authority is conferred.
  - Endorsement of authority is conveyed by other authorities
  - That is, no hub concept
  - This is how peer review works in the academic and scholarly world.

- Authorities themselves convey authority on those they link to. This naturally leads to a forumation in terms of Two equivalent views of page rank:
  1. Authorities directly conveying authority (without hubs)
  2. Authority values resulting from long term behaviour of a random walk on a graph.

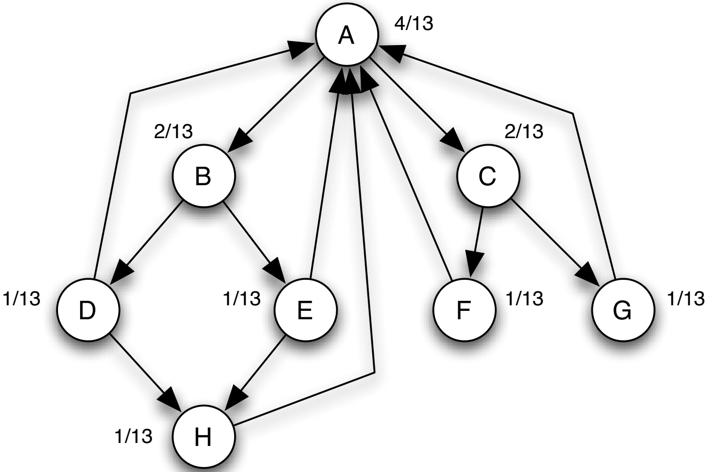
How does Page rank spread authority?

- Suppose at any point of time we have relevant authority scores.
  - A page spreads its authority equally amongst all of its out links.
  - If a page has no outlinks then all authority stays there.
- This redistributes the authority scores. (We are not creating or losing any authority, we are just redistributing it.)
- We can initially start with every relevant page having authority $1/n$ where there are n pages. Then we repeat this process k times for some sufficiently large k .
- With the exception of some "degenerate cases" (e.g. the process is periodic) it can be proven (again using linear algebra) that this process has a limiting behavior as $k \rightarrow \infty$.
- The resulting limit values will form an equilibrium.
- If the network is strongly connected then there is a unique equilibrium,

Remark

In many cases this won't reflect the desired authority. Namely, if the network has any sinks (or SCC that are sinks) which it will surely have, then all of the authority will pass to such sinks.

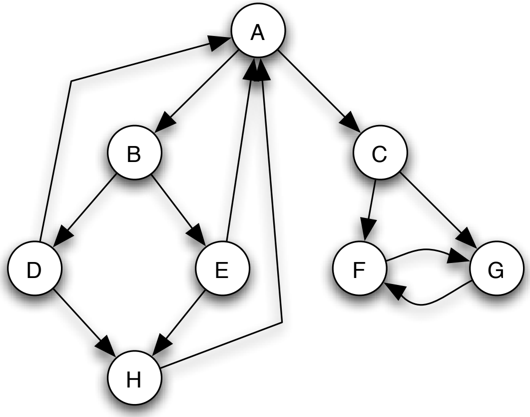
Page rank equilibrium for a network



[Fig 14.7, E&K]

Figure : Equilibrium PageRank values for the network of eight Web page.

Where has all the authority gone when we redirect (F, A) and (G, A) edges?



[Fig 14.8, E&K]

Figure : The same collection of eight pages, but F and G have changed their links to point to each other instead of to A. Without "scaling", all the PageRank would go to F and G.

Scaled page rank

- The way around this sink hole of authority is to have a scaled version of page rank where
  - only a fraction s of the authority of a page is distributed to its out links
  - the remaining $(1 - s)$ fraction is distributed equally amongst all relevant pages.

- For any value of $s < 1$ (which effectively makes the graph strongly connected), we get convergence to a unique set of scores for each page and that is its page rank (for that particular value of s). It is reported that Google uses $0.8 \leq s \leq 0.9$.

- (See the footnote on page 410 of E&K as to why in the previous example, nodes F and G will still get most of the authority but that for realistically large networks, the process works well.)

Some additional remarks

- The limiting scores for both the authority and hubs approach and the page rank approach are equilibrium points for an appropriate algebraic process.
- That is, if we actually were in the limiting state, we would be in the equilibrium state. In practice, we can stop the process when the change in each iteration is sufficiently small.
- We can weight the network edges (say according to some concept of link importance) and apply the same authority and hubs or page rank approach distributing authority in proportion to these weights.

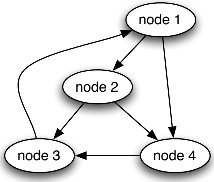
Advanced material (section 14.6): Handwaving argument why these processes converge

We have already suggested that both the page rank and hubs and authorities processes can be understood in terms of an algebraic process, namely, a linear transformation.

- Suppose we are considering a web network of n pages. We can represent the hub, authority or page rank values at any time k of the process by an n-dimensional (column) vector, denoted (respectively) by $\mathbf{h}^{(k)}, \mathbf{a}^{(k)}, \mathbf{r}^{(k)}$.
- Here we are using boldface $\mathbf{v} =< v_1, \ldots, v_n >$ to represent a vector whose components are the v_j so that (for example), $r_j^{(k)}$ repesents the page rank of the j^{th} web page after k steps of the page rank process.
- Let $\mathbf{v}$ be any of the hub, authority or page rank vectors. In each case it is not difficult to see that the process can be viewed as a linear transformation $\mathbf{v}^{(k+1)} = M\mathbf{v}^{(k)}$ for some appropriate $n \times n$ matrix M whose entries are non negative real numbers.

Advanced material continued: page rank convergence

- Section 14.6 tells us how to define the appropriate matrices and gives the conditions that will guarantee the convergence of the process; that is, when there exists $\mathbf{v}^{(*)} = \lim_{k \to \infty} \mathbf{v}^{(k)}$ and when this limit vector $\mathbf{v}^{(*)}$ is unique and independent of the starting vector $\mathbf{v}^{(0)}$.
- Figure 14.3 of the text illustrates a simple directed graph and the matrix N that defines the unscaled page rank update process. That is, $< r_1^{k+1}, \ldots, r_n^{k+1} >= N^{tr} < r_1^{k}, \ldots, r_n^{k} >$ where N^{tr} is the transpose of matrix N.



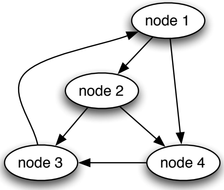
$$\begin{bmatrix} 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 1/2 & 1/2 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

[Fig 14.13, E&K]

Figure : A toy web graph and the associated matrix N describing the unscaled update process.

Page rank analysis for the scaled update

Similarly Figure 14.4 illustrates the same graph and the matrix $\tilde{N}$ that defines the scaled page rank update process with scaling factor $s = .8$.



$$\begin{bmatrix} .05 & .45 & .05 & .45 \\ .05 & .05 & .45 & .45 \\ .85 & .05 & .05 & .05 \\ .05 & .05 & .85 & .05 \end{bmatrix}$$

[Fig 14.14, E&K]

Figure : The same toy web graph and the associated matrix $\tilde{N}$ describing the scaled update process with $s = 0.8$.

- It follows that $\mathbf{r}^k = (\tilde{N}^{tr})^k \mathbf{r}^0$
- If the process is converging then it would be converging to some $\mathbf{r}^*$ satisfying $\mathbf{r}^* = N^{tr}\mathbf{r}^*$

Now comes the necessary linear algebra

So far we have mainly used matrices as a convenient way to represent the process. But to understand convergence we need to mention some more essential aspects of linear algebra. (Note to CS students: This is one of the many reasons linear algebra is a required course)

- Let $M_{n \times n}$ be a full rank matrix. Recall that the matrix-vector multiplication $M\mathbf{v}$ can rotate and expand/shrink the vector $\mathbf{v}$.
- Since it is hard to "visualize" an n-dimensional vectore space, we can simply think about the meaning of such a linear transformation in 2-space or 3-space.
- A vector $\mathbf{v}$ is an eigenvector of M with associated eigenvalue λ if $M\mathbf{v} = \lambda\mathbf{v}$. It follows that $\mathbf{v}$ is also an eigenvector of M^k with eigenvalue λ^k.
- When $\lambda = 1$, the eigenvector then becomes an equilibrium of the process!

More linear algebra

- For each full rank matrix there is a set of n eigenvectors with (not necessarily distinct) associated eigenvalues $\lambda_1, \ldots, \lambda_n$; these eigenvectors span the n-dimensional Euclidean space so that any vector can be expressed as a linear combination of the eigenvectors.
- An important result from linear algebra (Perron's Theorem) states that any matrix which has all positive entries has a unique eigenvector $\mathbf{y}$ corresponding to the largest positive eigenvalue λ_1 and $\lambda_1 > |\lambda_i|$ for $i > 1$.
- Since $\lambda_1 > |\lambda_i|$ for $i > 1$, and since every vector is a linear combination of the eigenvectors, it follows that as $k \to \infty$, the transformation M^k is being dominated by the largest eigenvalue acting on its associated eigenvector.
- For the scaled matrix $\tilde{N}^{tr}$, all entries are positive and the largest eigenvalue is 1. It follows that as $k \to \infty$, $(\tilde{N}^{tr})^k\mathbf{v}$ will converge to the eignevector $\mathbf{y}$ associated with the largest eignevalue 1.

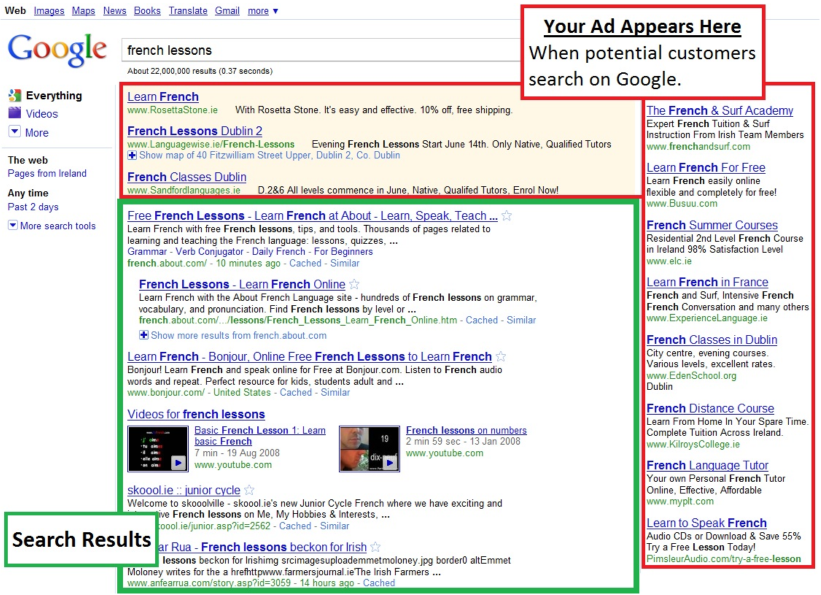
Similar analysis for hubs and authorites

- If M is the adjacency matrix of the web graph, then the process can be described by $\mathbf{h} = M\mathbf{a}$ and $\mathbf{a} = M^{tr}\mathbf{h}$.
- Then
  1. $\mathbf{a}^{(1)} = M^{tr}\mathbf{h}^{(0)}$
  2. $\mathbf{h}^{(1)} = M\mathbf{a}^{(1)} = MM^{tr}\mathbf{h}^{(0)}$
- It follows that
  1. $\mathbf{a}^{(k)} = (MM^{tr})^{k-1}M^{tr}\mathbf{h}^{(0)}$
  2. $\mathbf{h}^{(k)} = (MM^{tr})^{k}\mathbf{h}^{(0)}$

Hubs and authorities analysis continued

- The matrices (MM^{tr}) and $(M^{tr}M)$ are symmetric and have non-negative entries.
- Any $n \times n$ symmetric matrix S with non negative entries has an orthonormal set of n eigenvectors all of whose associated eigenvalues are real. By normalizing the scores, we can assume that the largest eigenvalue $\lambda_1 = 1$.
- If the largest eigenvalue is unique (which is what would happen in a real web graph), then the same analysis for page rank applies (assuming that the starting hub scores are all positive).

New Topic: Sponsored Search



Sponsored search: search becomes profit

> “it soon became clear that a lucrative market existed. . . combining search with advertising targeted to the queries that users were issuing.”

VS

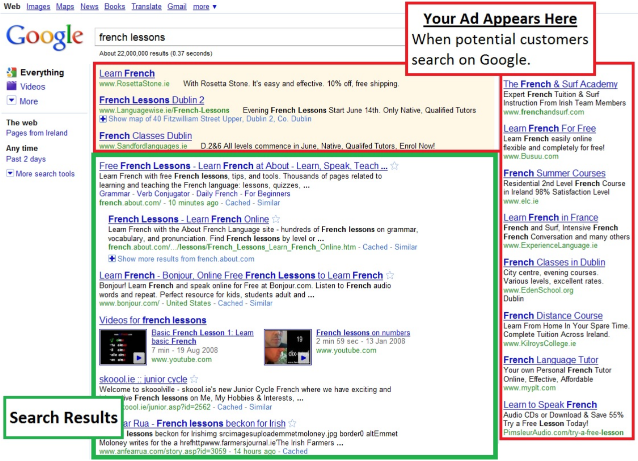
> “We all knew that search was going to be big but we didn’t know if or how it could be profitable.”

- It wasn’t clear how pervasive online search would become and it wasn’t clear (and maybe things will change) if you could be profitable just selling desired information (vs selling advertising).
- Online vs more traditional advertising:
  - To some (relatively small) extent all advertising tries to market to an intended audience.
  - Television advertisers choose programs based on what they know of the audience.
  - Advertisers in print media ask for ads to be placed in certain sections.

The advantage of online search ads

- "An ad that is based on the query is catching a user at a very receptive moment"; that is, when a user is expressing intent for information $\Rightarrow$ a much more refined way to target advertising.
- Query search is so widespread, ads can be read anywhere in the world or targeted to geographic regions
- Users are often currently searching for particular items.
- Keyword based search has become the main (perhaps 98%) of search engine company (eg Google) revenues.
- Google has extended query based targeted search to its gmail service.
- The idea of combining such targeted search in combination with a social network seems like the next step. (Is Facebook exploiting the network?)

Selling slots in sponsored search



- Google and other search engines display several paid advertisements along with the unpaid algorithmic ranking of documents in response to any query.

Highest cost per click bids - 2006

- For top 50 most expensive adwords (as of 2006) see
  http://wpmarketing.org/2009/08/the-top-50-most-expensive-adwords/

- The top 5 in Adword Average CPC (cost per click):
  1. school loan consolidation $69.16
  2. college loan consolidation $68.35
  3. car insurance quotes $66.88
  4. school consolidation $66.29
  5. auto insurance quotes $65.90

Google's top 5 cost per click CPC - January 2014

For some (confusing) January-2014 data see http://www.poweredbysearch.com/expensive-ppc-keywords-adwords/ for January 2014 costs per keyword

1. Insurance $54.91 24%
2. Loans $ 44.28 12.8%
3. Mortgage $ 47.12 9%
4. Attorney $ 47.07 3.6%
5. Credit $ 36.06 3.2%

Topical and very expensive CPC data

- More recent most expensive CPC (and volume per month) data can be found at http://http://www.adluge.com/blog/marketing-intelligence/paid-search/top-20-expensive-pay-per-click-terms/

Keyword	Search volume		Average CPC		Category
mesothelioma settlement	1,900	~	$142.67	↑↑↑	asbestos
mesothelioma asbestos attorney	5,400	~	$121.68	↑	asbestos
asbestos attorney	9,900	~	$117.23	↓	asbestos
asbestos law firms	2,900	~	$106.77	↑	asbestos
sell annuity payment	1,900	~	$95.16	↑	annuity
annuity settlements	5,400	~	$92.36	↑	annuity
structured annuity settlement	6,600	↑↑↑	$92.36	↑	annuity
auto donation	33,100	~	$84.51	↑	donate
virtual data rooms	4,400	~	$79.59	↑	data
donating a used car	2,900	~	$79.35	↑	donate
auto accident attorney	135,000	~	$73.12	↑	auto

- And for a video on how to generate low cost (e.g. 2 cent) adwords, see http://www.colinklinkert.com/cheap-ppc/

Selling slots as an example of matching markets

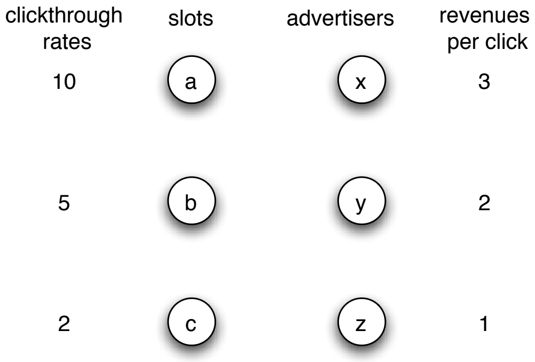
- Ignoring game theoretic aspects (of advertiser and search engine self-interest), what is the algorithmic problem?
- The search engine has some number of slots it wishes to sell to some number of advertisers/agents.
- As the text points out, this is analogous to a University wanting to fill a certain number of dorm rooms or courses.
- Here we might assume that the administration (of the University) is altruistic and only wants to optimize some sense of social welfare, matching students to the placements that they want the most.
- In many such cases, the provider just gives out placements on a first come first serve basis or perhaps based on some merit but some Business schools let students auction for courses.

Matching markets

- The adword problem is a special case of matching markets. This was the subject matter of Chapter 10 and our lectures 17,18,19.
- In a matching market, we have sellers for each item (or one common seller for all the items) and buyers desiring those items where say buyer or agent i has some intrinsic value $v(i, j)$ for item j.
- This is a generalization of the single item auction where only one item is being sold, but it is a special case of the general combinatorial auction where agents are interested in one or more subsets of items.
- We assume now that each buyer or agent only wants a single item; he/she may value many different items but only wants to receive at most one item. (This is called *unit demand* in auctions.)

Bipartite graph of advertisers and slots

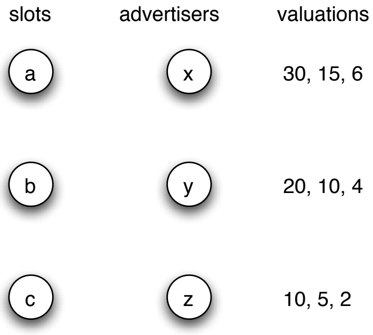
Without loss of generality we can assume that all edges exist and there an equal number of advertisers and slots. Why can we assume this?



[Fig 15.2, E&K]

Setting $v(i,j) = r(i)v(j)$

- Let $r(i)$ be the click-through rate for slot i (= number of clicks per say hour)
- Let $v(j)$ be the (expected) revenue per click for the jth advertiser.
- Usually one assumes that the click-through rates favour the top slots; more specifically we assume $r(i) > r(i+1)$.



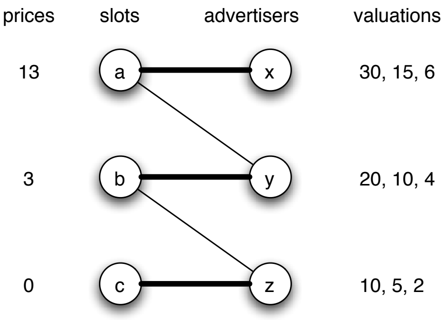
(a) *Advertisers' valuations for the slots*

[Fig 15.3 (a), E&K]

Maximum value matching

Fact

The problem of computing a maximum value matching in a bipartite graph (or any graph) can be computed optimally in time $O(|V| \cdot |E|)$.



(b) *Market-clearing prices for slots*

[Fig 15.3 (b), E&K]

The stated assumptions

Before preceding to the game theoretic aspect where advertisers are making not necessarily truthful bids as to their revenue per click values, we make the following assumptions:

- Advertisers and the search engine know the click-through rates. (This is not a worrisome assumption as statistics are kept.)
- The click-through rate is independent of the quality and relevance of the ad. Is this a valid assumption?

The stated assumptions

Before preceding to the game theoretic aspect where advertisers are making not necessarily truthful bids as to their revenue per click values, we make the following assumptions:

- Advertisers and the search engine know the click-through rates. (This is not a worrisome assumption as statistics are kept.)
- The click-through rate is independent of the quality and relevance of the ad. Is this a valid assumption?
  - This is not a valid assumption and the text goes on to show how to correct this.
  - Namely, we need to incorporate a quality factor $q(j)$ depending on the appropriateness of the advertiser j's ad to the query.
  - This changes click-through to be $q(j)r(i)$ if advertiser j gets slot i.
  - The value $v(i,j)$ becomes $q(j)r(i)v(j) = r(i)q(j)v(j) = r(i)v'_j$.
- The click-through rate does not depend on what is in the other slots.
  - This is also not a valid assumption and is a complex issue.

Some unstated assumptions/issues

- Advertisers usually state a bid per click $v(j)$ and also a maximum budget $\mathrm{Budget}(j)$ which cannot be exceeded.

- The conversion rate (when click becomes purchase) is higher for slots not quite at the top.

- Advertisements can vary slightly in the amount of space they occupy.

- The search engine can have a reserved price for slots.

- This is a repeated game where advertisers are constantly monitoring and changing their bids.

- It is the search engine that determines the quality factor $q(j)$ and this is essentially a search ranking problem using the proposed ads as the documents to be ranked.