BA Network Analytics
Lecture 4

Kalyan Talluri

Imperial College Business School, London

kalyan.talluri@imperial.ac.uk

Jan-Feb 2021

Eigenvector based centralities for undirected graphs[26]

[26] Main reading activity for next week (for class discussion)

- Read article in this week's folder *Romantic Partnerships and the Dispersion of Social Ties: A Network Analysis of Relationship Status on Facebook*, Backstrom and Kleinberg. Authors have unique access to the ground-truth!

Eigenvector Centrality

Attributed to Bonacich: If A is the adjacency matrix of an undirected graph, define centrality vector $\boldsymbol{x}$ (each element of corresponds to a node) if it satisfies

$$A\boldsymbol{x} = \lambda\boldsymbol{x}$$

Questions: what is the role of λ ? Why should x exist?

- As A is symmetric, the matrix has all *real* eigenvalues and eigenvectors (that's a relief!)
- But . . . problems:
  - Can have n eigenvalues $\lambda_{\max} \geq \lambda_2 \ldots \geq \lambda_{\min}$ for a graph with n nodes, and all satisfy $A\boldsymbol{x} = \lambda\boldsymbol{x}$. Which one?
  - $\boldsymbol{x}$ may be real, but why should it be non-negative? If they have negative elements, how to interpret a negative value as a centrality measure?
- Partly for this reason, we settle on $\lambda_{\max}$ as the default choice. Perron-Froebenius guarantees that the corresponding eigenvector has all non-negative values (note that all elements of A are non-negative)

Still not entirely satisfactory—we would like to get $A\boldsymbol{x} = \boldsymbol{x}$ rather than a scaling factor $\lambda_{\max}$. Variations: Normalize the weights by the degree to make A a stochastic matrix (sum of each row is 1) when you get $\lambda_{\max} = 1$

Power, prestige, centrality[27]

Sociologists and political scientists discuss the subtleties of these three concepts (none of them are easy to nail down with any precision):

- Centrality: A pure network concept based on the topology of the network
- Prestige: Asymmetric relationship. Others follow you (say on twitter) but you do not follow them.
- Power: A bilateral concept and usually interpreted as *bargaining* power between two players—the two parties are sharing a pie between themselves. "Power comes from being connected to those who are powerless".
  - Trump is important because he *knows* Putin as an equal (importance); Trump has little power *over* Putin (bargaining power)
  - Boss has power *over* employees (hierarchy)—however, this depends on outside options, how critical the employee is!

Sociologists have come up with various measures to "attempt" to capture these effects. These measures work on some networks, miss in others . . . i.e., no dominant measure in existence

[27] We won't delve too much into these topics, but if you are interested, EK, §12 is a good easy read into notions of "power" from econ/sociology point of view

Eigenvector $\rightarrow$ Bonacich, Katz, etc.

Similar ideas, differences in detail to account for notions of power.

- Katz Prestige-1: Almost identical to eigenvector centrality, but normalize each column by its degree (scale the rows of the adjacency matrix A by the row's degree, then take the transpose of the matrix)—a node's importance gets equally distributed to its neighbors. For directed graphs, normalize by either in-degree or out-degree.
- Katz Prestige-2: Prestige is sum of (discounted distances) to other nodes[28]

$$\boldsymbol{x} = \sum_{i=1}^{\infty} \beta^i A^i 1 = \beta[I - \beta A]^{-1} A1$$

  - $|\beta|(< \frac{1}{\lambda_{\max}})$ is an attenuation factor; we use β as a radius tuning-parameter to measure centrality: local influence? take small β; more global? take larger β.
  - Elements of A^i gives the number of length-i paths between pairs of nodes[29].

[28] Recall algebra fact $\frac{1}{1-a} = (1-a)^{-1} = 1 + a + a^2 + \ldots$ for $a < 1$; Holds for matrices too under some conditions

[29] Check out this fact about adjacency matrices by experimenting on a small graph: elements of the matrix A^i gives the number of length-i walks in the graph. See the examples from Jackson's book `http://press.princeton.edu/chapters/s2_8767.pdf`. Can prove by induction if you are so inclined.

PageRank

The internet is a directed graph: Original idea of Google's ranking algorithm (importance of results) for the directed graph representing the web

Ranking items by relevance/importance comes up in many many contexts

Results[30] must be presented in some order

- What order? Relevance, recency, popularity, reliability?
- Some ranking methods tried (most discarded)
  - Presence of keywords in title of document
  - Closeness of keywords to start of document
  - Frequency of keywords in document
  - Link popularity (how many pages point to this one)
- Critical issue: Can the page owner influence the ranking? Manipulation!
  - PageRank turned out (at that time) to be the most difficult to manipulate
  - Still, secrecy on how the ranking is done is the best defense

[30] Google indexes all pages and extracts out the subset that contain your keyword

PageRank idea

A link in page 1 to page 2 is a recommendation (a "vote") of page 2 by the author of page 1 (we say 2 is successor of 1)

- The "quality" of a page is related to the number of links that point to it (its in-degree)
- A "vote" from an important page is worth more
- A page is important if it is pointed to by other important pages
- Define the pagerank r_j of node j as

$$r_j = \sum_{(i,j) \in G} \frac{r_i}{d_o(i)}$$

i.e.: the importance of each node i is proportionally split based on the out-degree of i (the $d_o(i)$). PageRank of j is the sum of these fractional PageRanks of the nodes that point to j.

PageRank idea

PageRank is calculated by solving a set of linear equations (software; many Python packages have a PageRank routine)

- In matrix notation[31]

$$\boldsymbol{r} = W^T \boldsymbol{r}$$

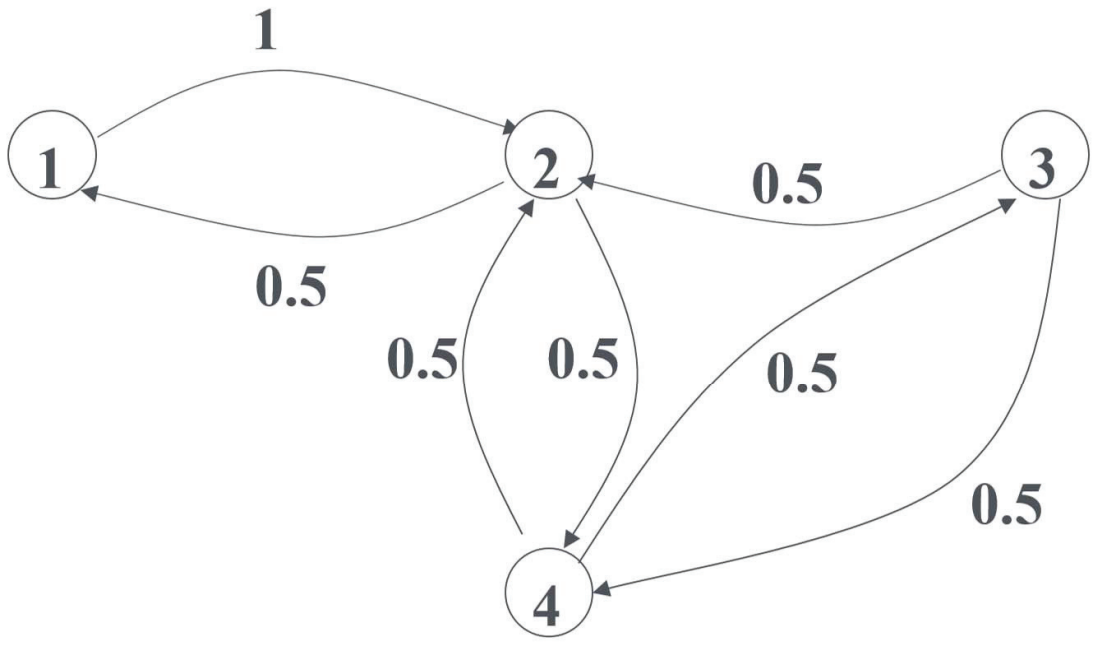
  where W is the square adjacency matrix ((i, j) of the *directed* graph with each row normalized by its out-degree, i.e., has 1 if there is an arc going from i to j) divided by out-degree of i

Note a small technical problem: W, need not be symmetric! Means no guarantee that it has all real eigenvalues, let alone it has an eigenvalue of 1.

[31] This is actually the (right) eigenvector of W^T; Why do we take the transpose W^T, instead of W?

PageRank calculation example

Consider the following graph, where edge-weights are normalized by the out-degree of that node



$$A = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{pmatrix} \quad W = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0.5 & 0 & 0 & 0.5 \\ 0 & 0.5 & 0 & 0.5 \\ 0 & 0.5 & 0.5 & 0 \end{pmatrix} \quad W^T = \begin{pmatrix} 0 & 0.5 & 0 & 0 \\ 1 & 0 & 0.5 & 0.5 \\ 0 & 0 & 0 & 0.5 \\ 0 & 0.5 & 0.5 & 0 \end{pmatrix}$$

Ask yourselves: What is special about this matrix W? (Hint: Sum the elements of each row—they are called stochastic matrices).

PageRank calculation example

Need to solve $\boldsymbol{r} = W^T \boldsymbol{r}$; that is the eigenvector corresponding to eigenvalue of 1

$$\boldsymbol{r} = \begin{pmatrix} 0 & 0.5 & 0 & 0 \\ 1 & 0 & 0.5 & 0.5 \\ 0 & 0 & 0 & 0.5 \\ 0 & 0.5 & 0.5 & 0 \end{pmatrix} \boldsymbol{r}$$

In algebraic form:

$$r_1 = 0.5r_2; r_2 = r_1 + 0.5r_3 + 0.5r_4 \ldots$$

Why should a solution exist? Why should there be an eigenvalue of 1? The matrix is not even symmetric $\rightarrow$ why should eigenvalues be even real?

PageRank calculation theory: existence

Recall linear algebra theory

- W is a stochastic matrix, i.e., rows are non-negative and sum to 1
- The largest of these eigenvalues $\lambda_{\max}$ is always 1
- The vector $\boldsymbol{r}$ is an eigenvector of W (or W^T) corresponding to eigenvalue 1. Perron-Frobenius says it has all non-negative real elements

So existence of a solution is assured and r has non-negative entries Solving the system is relatively easy: for our example, turns out to be (interpret it)

$$\begin{pmatrix} 0.2 & 0.4 & 0.133 & 0.2667 \end{pmatrix}$$

Question: How can you check this is the correct answer?

Google's version of PageRank

Google-specific issues

- First crawls the web to get all the pages and downloads them (crawler)
- Indexes them by words (each word has a list of all pages where it occurs)
- When a user types a keyword, looks up the pages where it occurs
- Uses PageRank (type algorithm) to rank them

Major computational problem

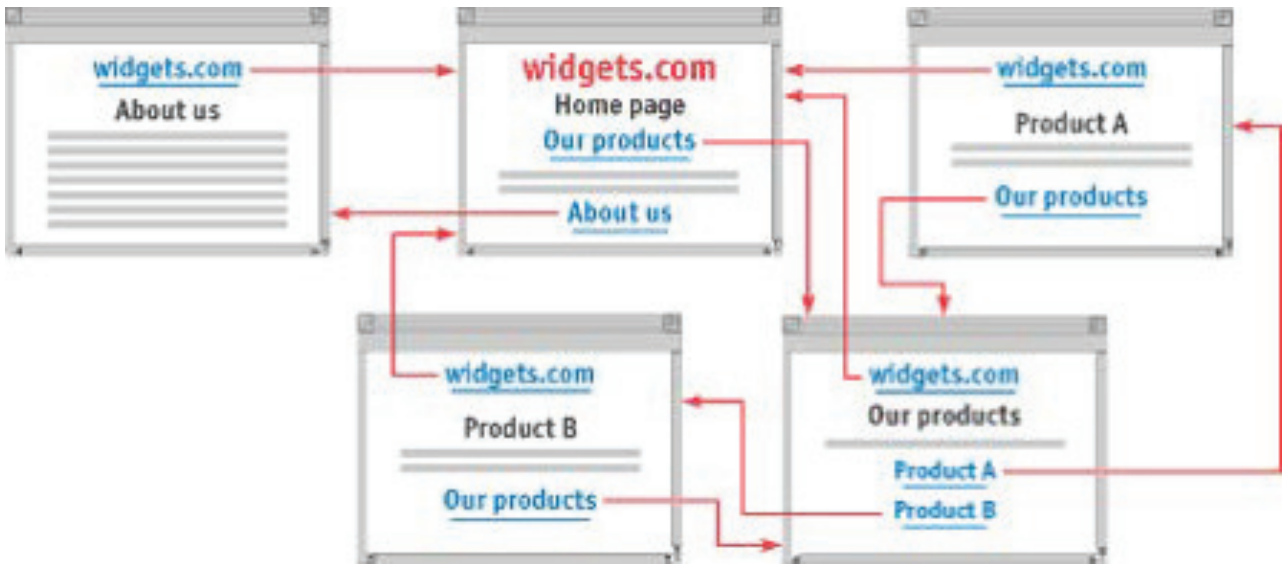
- Solving an eigenvector system scales as $O(n^3)$
- For the entire Web graph $n \sim 1$ trillion (!!)
- So direct solution is not feasible

Google uses the power method of solving the system

$$\boldsymbol{r}^{(k+1)} = W^T \boldsymbol{r}^{(k)}, \quad k = 1, 2, \ldots$$

found to converge for k around 20 (but no theoretical guarantee—we need the network to be irreducible in a Markov-chain sense; aperiodic and strongly connected)

Power iteration example[32]

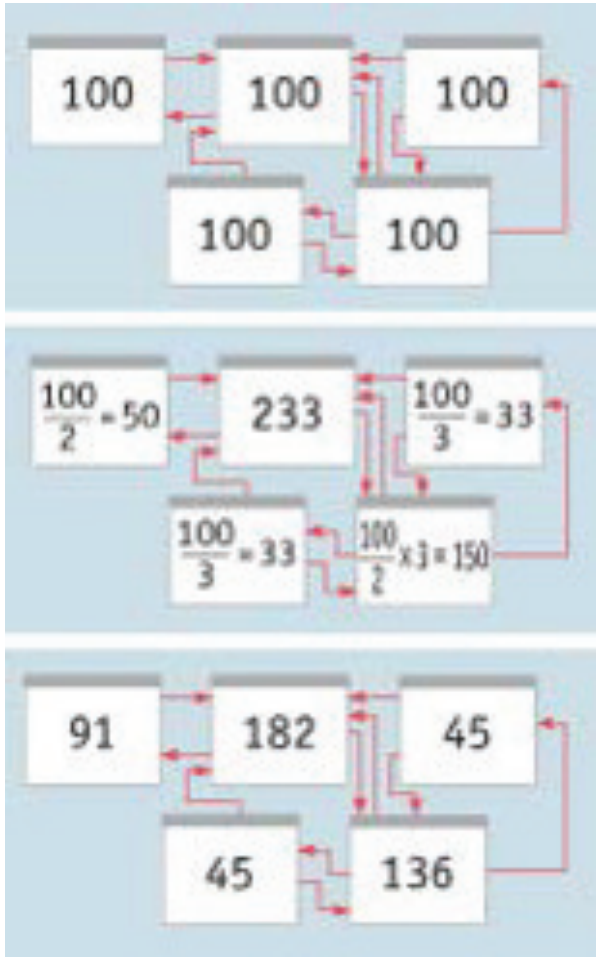


Write down the graph, and the matrix W first

[32] From http://www.economist.com/node/3172188

Power iteration example

Initial vector $\begin{pmatrix} 100 & 100 & 100 & 100 & 100 \end{pmatrix}$

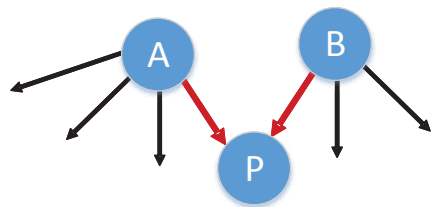


Google's model for PageRank

From original paper of Brin & Page

- Consider the following infinite random walk (surfing): Initially the surfer is at a random page. At each step, the surfer proceeds
  - stops surfing or goes to a random disconnected page with probability $1 - d$
  - with remaining probability d clicks on a randomly chosen successor of the current page
- The PageRank of a page p is the fraction of steps the surfer spends at p as the number of steps approaches infinity

d is referred to as a damping factor, and 0.85 is a commonly quoted number[33]



PR(P)= 0.15/N + 0.85 × [PR(A)/4 + PR(B)/3]

$$PR(p) = \frac{1-d}{n} + (d) \sum_{(q,p)\in E} \frac{PR(q)}{D_o(q)}$$

where $PR(\cdot)$ is the PageRank and $D_o(\cdot)$ is the out-degree

[33] The technical purpose of this is the following: By adding random jumps, the Markov-chain becomes aperiodic. In addition if it is strongly connected, it guarantees that a stationary distribution exists, and therefore an eigenvalue of 1 exists.

Google's version of PageRank

Current algorithm?

- No one knows for sure (kept secret to reduce manipulation)–certainly has evolved a lot from the original paper
- A huge number of business rules, experimentation, improvements, quality feedback on validity . . . constantly tweaked
- PageRank removed from public view; Rumored to be: Links, Content and AI based

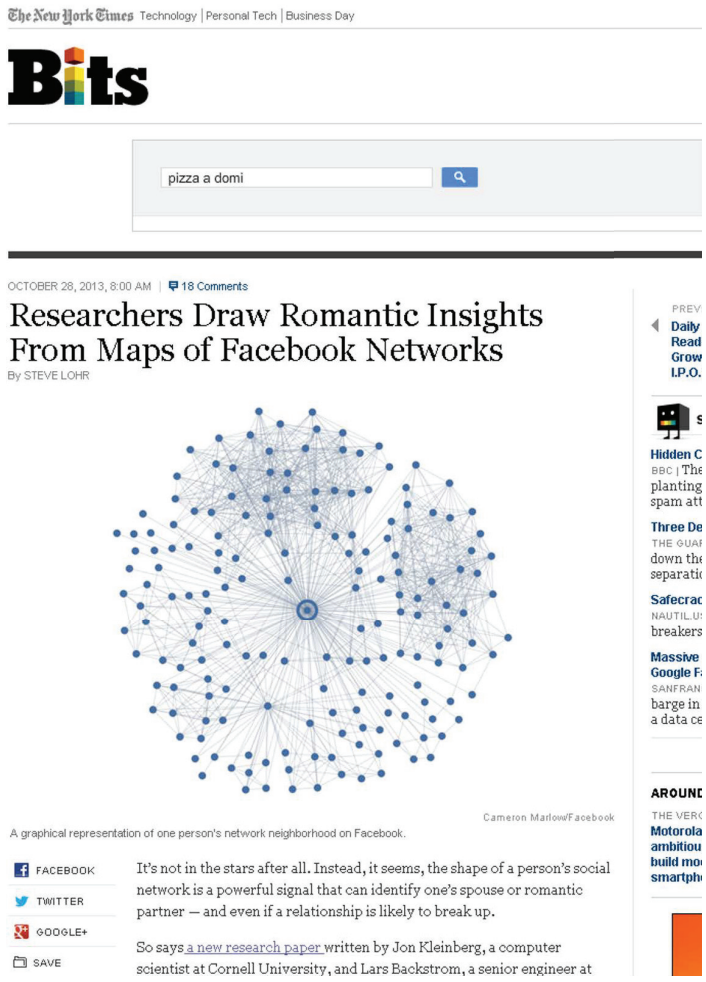
Originality of idea? The eigenvector measures were well-established in social-network theory. Great achievement is operationalizing it on scale–network has billions of pages, and the network constantly needs to be updated.

Casestudy: Relationship prediction based on a centrality measure
prediction based on some network measures (knowing only the network topology)[34]

Discussion question: How do they use centrality measure to understand an unobservable phenomenon?

[34] Takeaway: we can come up with some task-specific measures also that might work better

Casestudy: Relationship prediction based on a centrality measure[35]



The New York Times Technology | Personal Tech | Business Day

Bits

pizza a domi

OCTOBER 28, 2013, 8:00 AM | 18 Comments

Researchers Draw Romantic Insights From Maps of Facebook Networks

By STEVE LOHR

Cameron Marlow/Facebook

A graphical representation of one person's network neighborhood on Facebook.

FACEBOOK
TWITTER
GOOGLE+
SAVE

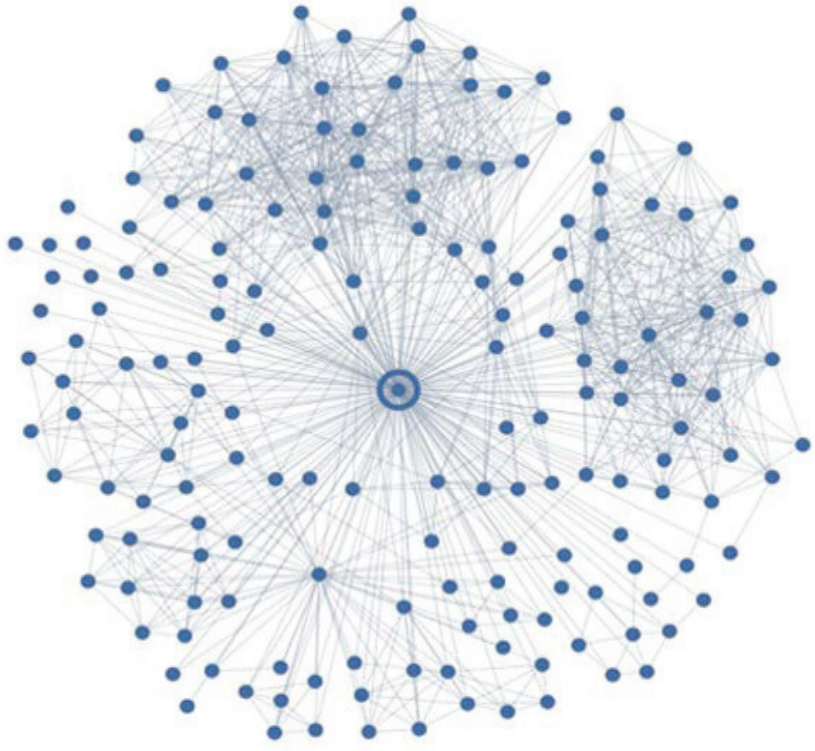
It's not in the stars after all. Instead, it seems, the shape of a person's social network is a powerful signal that can identify one's spouse or romantic partner — and even if a relationship is likely to break up.

So says a new research paper written by Jon Kleinberg, a computer scientist at Cornell University, and Lars Backstrom, a senior engineer at

[1] *Romantic Partnerships and the Dispersion of Social Ties: A Network Analysis of Relationship Status on Facebook*, Backstrom and Kleinberg. Authors have unique access to the ground-truth!

Example: Relationship prediction

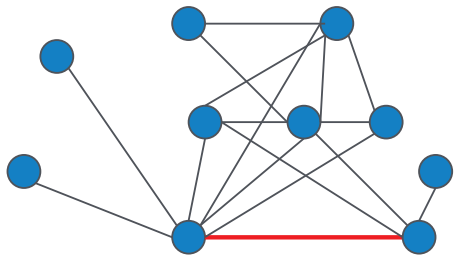
The person declares on Facebook he/she is in a relationship. Prediction task: With whom?



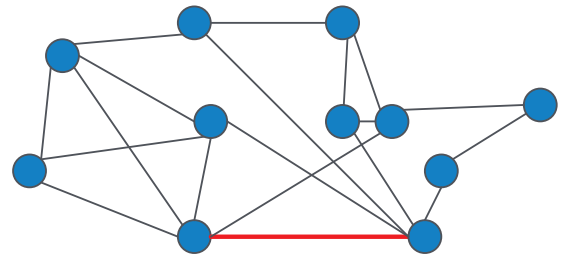
Data: The person's ego network (the induced subgraph of their neighbors)
Performance measure: How often can you predict correctly over 1.3 million such users
Information: Network structure only, or [+ Features (messages, photos, photos with both appearing, . . .)]

Design of an algorithm for prediction

Define two new local neighborhood measures



Embeddedness: How many mutual friends do they have? Example of a measure: mutual friends of the pair/all friends of the pair

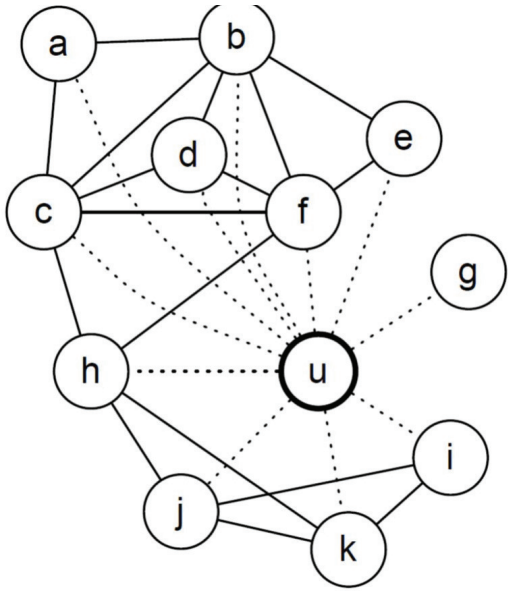


Dispersion: They *do* have a few mutual friends but the friend groups are dispersed

Typical prediction problem; essentially define a good measure that captures what we are interested in. (*Note measures based only on network structure.*)

Intuition/hunch $\rightarrow$ measures

Define a suitable (reasonable, easy to calculate) measure, Rank over the measure.



Embeddedness of an edge: Number of mutual friends shared by the two nodes of that edge
Dispersion of an edge (u, v):

$$disp(u, v) = \sum_{s,t \in C_{uv}} d_v(s, t)$$

where C_{uv} is the set of common neighbors, $d_v(s, t)$ is a distance function, $= 1$ if s, t are not connected by an edge and have no common neighbors in G_u (u and its neighbors) other than u and v, $= 0$ otherwise.

Example

Calculate measures for u: The links from u to b, c, and f all have embeddedness 5 (highest) u to h has an embeddedness of 4.
However, $disp(u, h) = 4$ (the pairs $(c, j)(c, k)(f, j)(f, k)$) , $disp(u, b) = 1$ (pair (a, e))

Ranking based on various measures

Amongst all neighbor nodes of u, pick the one with the maximum $\frac{disp(u,v)}{emb(u,v)}$, the normalized dispersion: 48% success rate

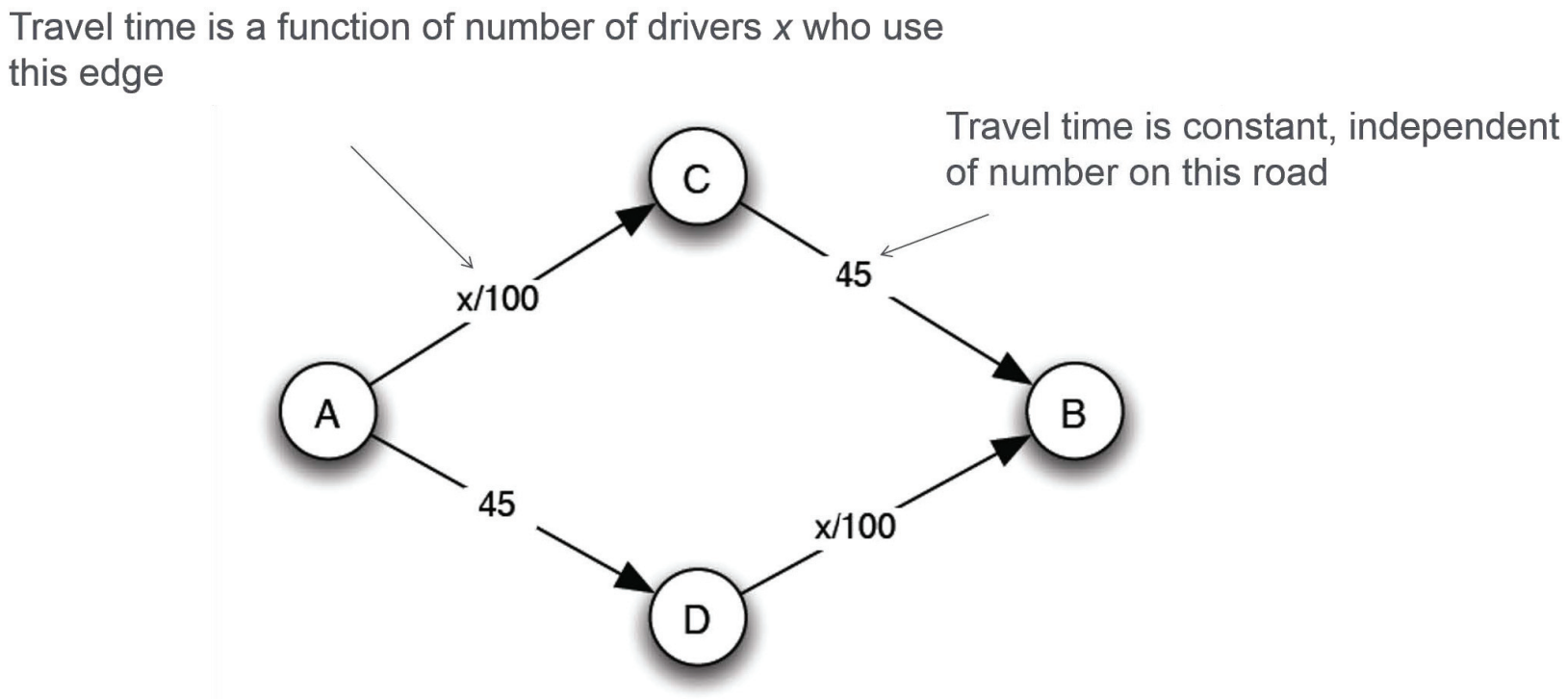
- experiment with variations of this idea
- explore size and time-effects
- compare with M/L methods

Keep in mind: they know the ground-truth. This is often missing in social network research based only on observed data (i.e. outside laboratory settings)

Final footnote on “Network effects” as used in Economics

"Network Effects" in Economics usually refers to network externalities—effect of your action or the others via the network

Negative externalities of traffic: the Braess Paradox[36]



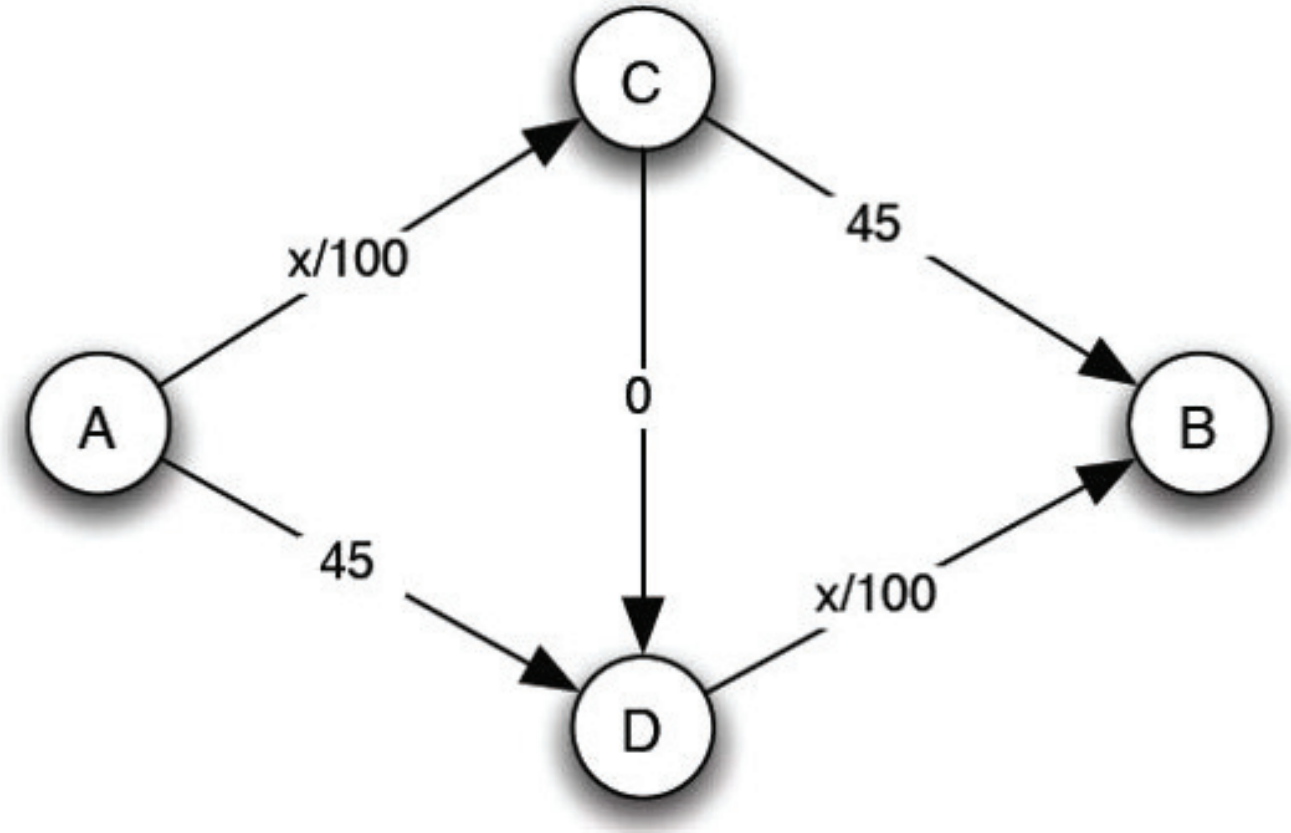
4000 drivers who want to go from A to B (no other drivers). Equilibrium: half take A-C-B, half take A-D-B. Travel time of 65

Nash equilibrium concept: Everyone does what is best for them; given other's actions, no reason to change my action

[36] See EK §8.2

Braess Paradox

Government builds a new fast road C $\rightarrow$D



New fast road would reduce drive time, right? What would you do if you are a driver?

The End