Lecture 10 Network Centralities

which nodes are important?

1. Degree centrality
2. Eigenvector centrality

Set $x_i = 1$ for all nodes, at $t = 0$

$x_i' = \sum A_{ij} x_j$ — improved measure of importance

$\Rightarrow \quad \underset{\sim}{x}' = A \cdot \underset{\sim}{x}$

After "time" t: $\quad \underset{\sim}{x}(t) = A^t \cdot \underset{\sim}{x}(0)$

Eigenvalue & eigenvectors of A: $\quad A \cdot \underset{\sim}{u}_i = \kappa_i \underset{\sim}{u}_i$

$\Rightarrow \quad \underset{\sim}{x}(0) = \sum_i c_i \underset{\sim}{u}_i \qquad \kappa_1 \geq \kappa_2 \cdots \geq \kappa_n$

$\Rightarrow \quad \underset{\sim}{x}(t) = A^t \cdot \sum c_i \underset{\sim}{u}_i = \sum c_i \kappa_i^t \underset{\sim}{u}_i$

$= \kappa_1^t \sum c_i \underbrace{\left(\frac{\kappa_i}{\kappa_1}\right)^t}_{\leq 1} \cdot \underset{\sim}{u}_i$

$t \to \infty$: $\quad \underset{\sim}{x}(t) \to c_1 \kappa_1^t \underset{\sim}{u}_1$

Leading eigenvector of A $\to$ <u>centrality</u> $\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$

$A \cdot \underset{\sim}{x} = \kappa_1 \underset{\sim}{x} \qquad A_{ii} = 0$

$\Rightarrow \quad x_i = \kappa_1^{-1} \sum_j A_{ij} x_j \qquad \Rightarrow \sum A_{ij} x_j \equiv$ sum of centralities of i's neighbors

How to make x_i large?

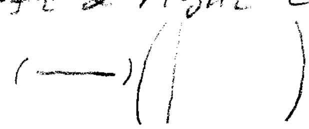
(i) Many neighbors (ii) some neighbors are important (iii) both

Note: – $x_i \geq 0 \quad (\because \underset{\sim}{x}(0) > 0)$

– works for undirected networks

Directed networks? $\Rightarrow$ Left & right eigenvectors

Left: $\quad \underset{\sim}{u}^T A = \kappa^L \underset{\sim}{u}^T$



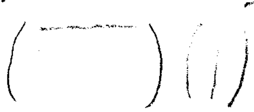
$\Rightarrow$ outgoing links

$A_{i1} \qquad i \leftarrow 1$

Right: $\quad A \cdot \underset{\sim}{v} = \kappa^R \underset{\sim}{v}$

A_{1j}

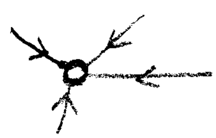
$j \to 1$



$\Rightarrow$ ingoing links

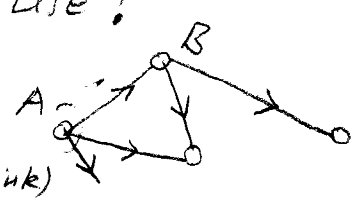
which one to use?

WWW



$\Rightarrow$ Right eigenvector

Difficulty:



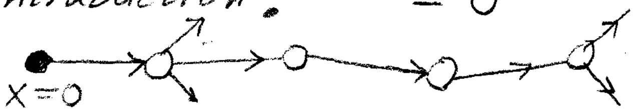
$x_A = 0$ (no ingoing link)

but $x_B > 0$ (one ingoing link)

$x_B = \kappa_1^{-1} \sum A_{Bj} x_j \quad j \in$ B's ingoing neighbors

contradiction! $\quad = 0$

Extreme case: (Acyclic net.)



$x = 0$

$x_i = 0$ for all i

3. $\Rightarrow$ Katz centrality

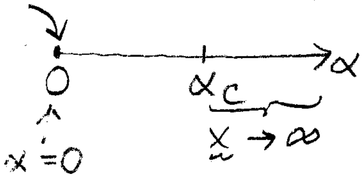
$$x_i = \alpha \sum_j A_{ij} x_j + \beta$$ ← "free" amount of centrality regardless of its position in the net.

α – free parameter

$$\underset{\sim}{x} = \alpha A \cdot \underset{\sim}{x} + \beta \underset{\sim}{1}$$

$\underset{\sim}{x} = \underset{\sim}{1}$ $\Rightarrow$ $$\underset{\sim}{x} = (\mathbb{I} - \alpha A)^{-1} \cdot \underset{\sim}{1} \quad \left(\begin{array}{l}\text{setting } \beta = 1\\ \Rightarrow \text{Katz}\end{array}\right)$$

$$(\mathbb{I} - \alpha A)^{-1} = \mathbb{I} + \alpha A + \alpha^2 A^2 + \cdots$$



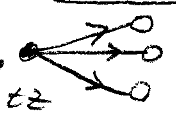
$\alpha_c = ?$

$$\det(\mathbb{I} - \alpha A) = 0$$

$$\Rightarrow \det(A - \underbrace{\alpha^{-1}}_{\text{eigenvalue}} \mathbb{I}) = 0$$

$$\boxed{\alpha_c = \frac{1}{\kappa_1}}$$

Problem



} All high Katz due to propagation

$\underset{\sim}{x}' = \alpha A \cdot \underset{\sim}{x} + \underset{\sim}{1}$ ↗

opposite to eigenvector centrality

4. PageRank

$$x_i = \alpha \sum A_{ij} \frac{x_j}{k_j^{out}} + \beta$$

If $k_j^{out} = 0 \Rightarrow A_{ij} = 0 \Rightarrow \frac{0}{0}$ — convention: set $k_j^{out} = 1$

$$\underset{\sim}{x} = \alpha A \cdot D^{-1} \cdot \underset{\sim}{x} + \beta \underset{\sim}{1}, \quad D = \begin{pmatrix} \diagdown & 0 \\ 0 & \diagdown \end{pmatrix}$$

$\beta = 1$

$$\underset{\sim}{x} = (\mathbb{I} - \alpha A \cdot D^{-1})^{-1} \cdot \underset{\sim}{1} \qquad D_{ii} = \max(k_j^{out}, 1)$$

$$\boxed{\underset{\sim}{x} = D (D - \alpha A)^{-1} \cdot \underset{\sim}{1}} \text{ — PageRank}$$

$\boxed{\alpha < (\text{largest eigenvalue of } A \cdot D^{-1})^{-1}}$

Choice of α ?

Say A – symmetric

$$A \cdot D^{-1} \begin{pmatrix} k_1 \\ \vdots \\ k_n \end{pmatrix} = A \cdot \underset{\sim}{1} = 1 \cdot \begin{pmatrix} k_1 \\ k_n \end{pmatrix}$$

↑ unit eigenvalue of $A \cdot D^{-1}$

Let $\mathbb{B} \equiv A \cdot D^{-1}$, $B_{ij} > 0$

Say $\mathbb{B} \cdot \underset{\sim}{v} = \gamma_1 \underset{\sim}{v}$, $\gamma_1 > 0$

$$\gamma_1 \underset{\sim}{v}^T \cdot \underset{\sim}{v} = |\gamma_1 \underset{\sim}{v}^T \cdot \underset{\sim}{v}| = |\underset{\sim}{v}^T \cdot \mathbb{B} \cdot \underset{\sim}{v}|$$

$\underset{\sim}{u} \equiv \begin{pmatrix} |v_1| \\ \vdots \\ |v_n| \end{pmatrix}$

$$= \left|\sum_{ij} B_{ij} v_i v_j\right| \leq \sum_{ij} |B_{ij} v_i v_j|$$

$$= \sum B_{ij} |v_i| |v_j| \equiv \underset{\sim}{u}^T \cdot \mathbb{B} \cdot \underset{\sim}{u}$$

$$\Rightarrow \gamma_1 \leq \underset{\sim}{u}^T \cdot \mathbb{B} \cdot \underset{\sim}{u} / \underset{\sim}{v}^T \cdot \underset{\sim}{v} = \underset{\sim}{u}^T \cdot \mathbb{B} \cdot \underset{\sim}{u} / \underset{\sim}{u}^T \cdot \underset{\sim}{u}$$

But for an arbitrary vector $\underset{\sim}{w} = \sum c_i \underset{\sim}{v}_i$

$$\frac{\underset{\sim}{w}^T \cdot \mathbb{B} \cdot \underset{\sim}{w}}{\underset{\sim}{w}^T \cdot \underset{\sim}{w}} = \frac{(\sum c_j \underset{\sim}{v}_j^T) \cdot \mathbb{B} \cdot (\sum c_j \underset{\sim}{v}_j)}{\sum_i c_i^2} = \frac{\sum c_i^2 \gamma_i}{\sum c_i^2}$$

$$\leq \sum c_i^2 \gamma_1 / \sum c_i^2 = \gamma_1$$

$$\Rightarrow \gamma_1 = \underset{\sim}{u}^T \cdot \mathbb{B} \cdot \underset{\sim}{u} / \underset{\sim}{u}^T \cdot \underset{\sim}{u}, \quad \Rightarrow \underset{\sim}{v} = \underset{\sim}{u}$$

Perron-Frobenius theorem:
1. Largest eigenvalue unique
2. The associated eigenvector $\underset{\sim}{u}$ has no negative elements
3. $\underset{\sim}{u}$ is unique

Maximum eigenvalue of $\mathbb{B} = A \cdot D^{-1}$ is one!

$\boxed{\alpha < 1}$ Google: $\alpha = 0.85$

PageRank

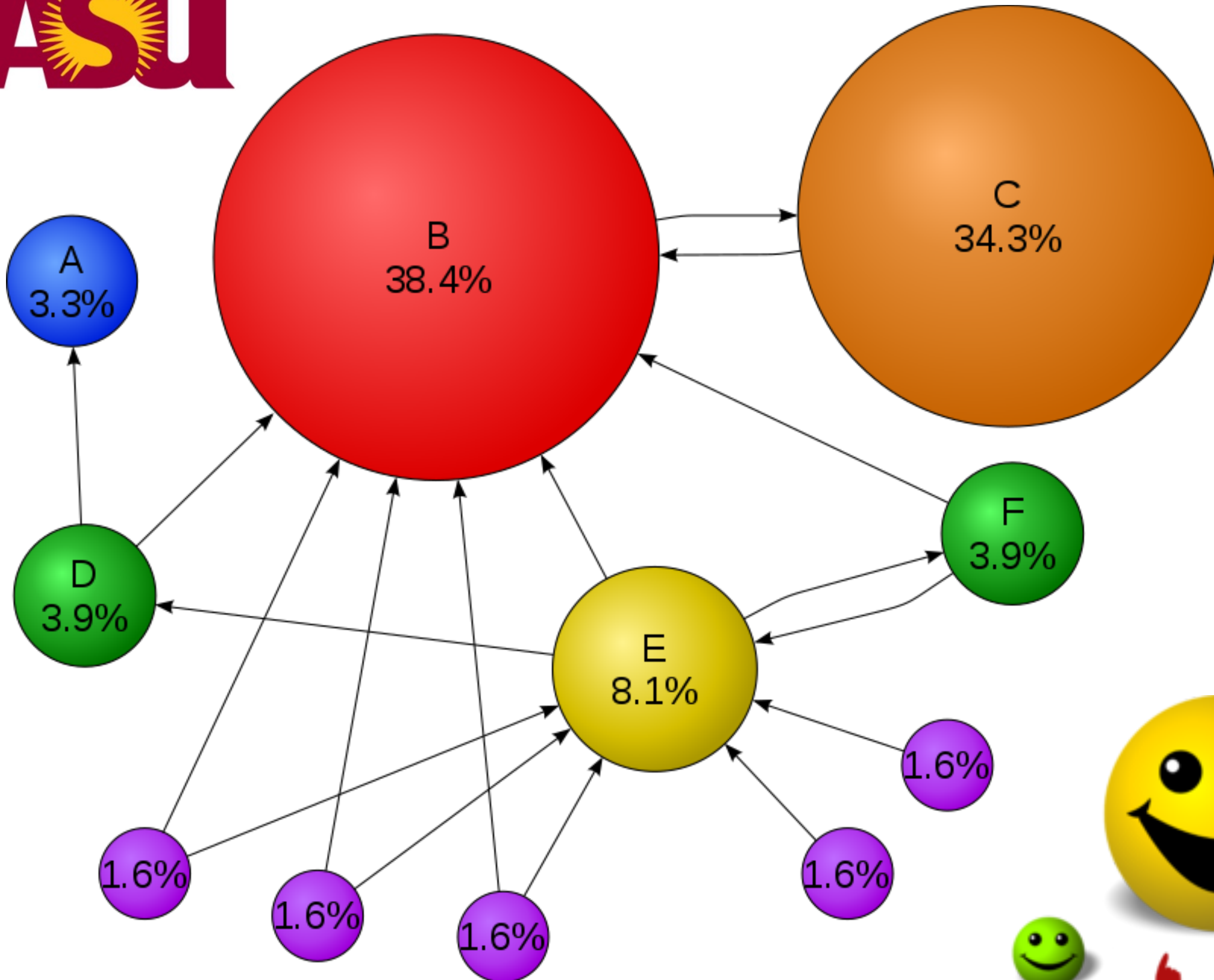
The Anatomy of a Large-Scale Hypertextual Web Search Engine

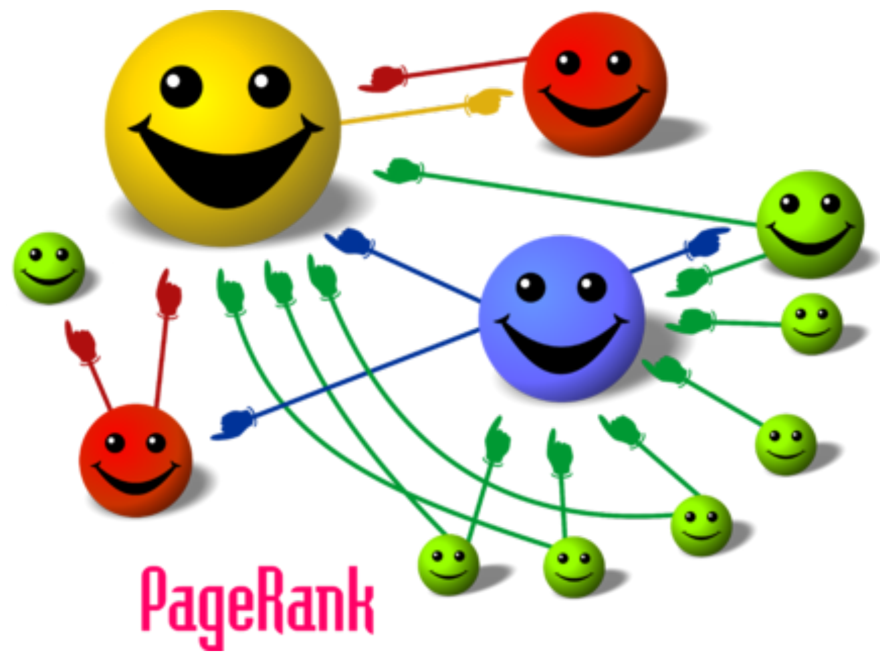
Sergey Brin and Lawrence Page

Computer Science Department,
Stanford University, Stanford, CA 94305, USA
sergey@cs.stanford.edu and page@cs.stanford.edu

Abstract

In this paper, we present Google, a prototype of a large-scale search engine which makes heavy use of the structure present in hypertext. Google is designed to crawl and index the Web efficiently and produce much more satisfying search results than existing systems. The prototype with a full text and hyperlink database of at least 24 million pages is available at http://google.stanford.edu/
To engineer a search engine is a challenging task. Search engines index tens to hundreds of millions of web pages involving a comparable number of distinct terms. They answer tens of millions of queries every day. Despite the importance of large-scale search engines on the web, very little academic research has been done on them. Furthermore, due to rapid advance in technology and web proliferation, creating a web search engine today is very different from three years ago. This paper provides an in-depth description of our large-scale web search engine -- the first such detailed public description we know of to date. Apart from the problems of scaling traditional search techniques to data of this magnitude, there are new technical challenges involved with using the additional information present in hypertext to produce better search results. This paper addresses this question of how to build a practical large-scale system which can exploit the additional information present in hypertext. Also we look at the problem of how to effectively deal with uncontrolled hypertext collections where anyone can publish anything they want.


B
38.4%
C
34.3%
A
3.3%
D
3.9%
F
3.9%
E
8.1%
1.6%
1.6%
1.6%
1.6%
1.6%


PageRank

Going Beyond the Web

SIAM Review
Vol. 57, No. 3, pp. 321–363

PageRank Beyond the Web*

David F. Gleich†

Abstract. Google's PageRank method was developed to evaluate the importance of web-pages via their link structure. The mathematics of PageRank, however, are entirely general and apply to any graph or network in any domain. Thus, PageRank is now regularly used in bibliometrics, social and information network analysis, and for link prediction and recommendation. It's even used for systems analysis of road networks, as well as biology, chemistry, neuroscience, and physics. We'll see the mathematics and ideas that unite these diverse applications.