michele.tizzanio@unito.it

di.unito.it

Michele Tizzani - Università degli Studi di Torino (Italy)

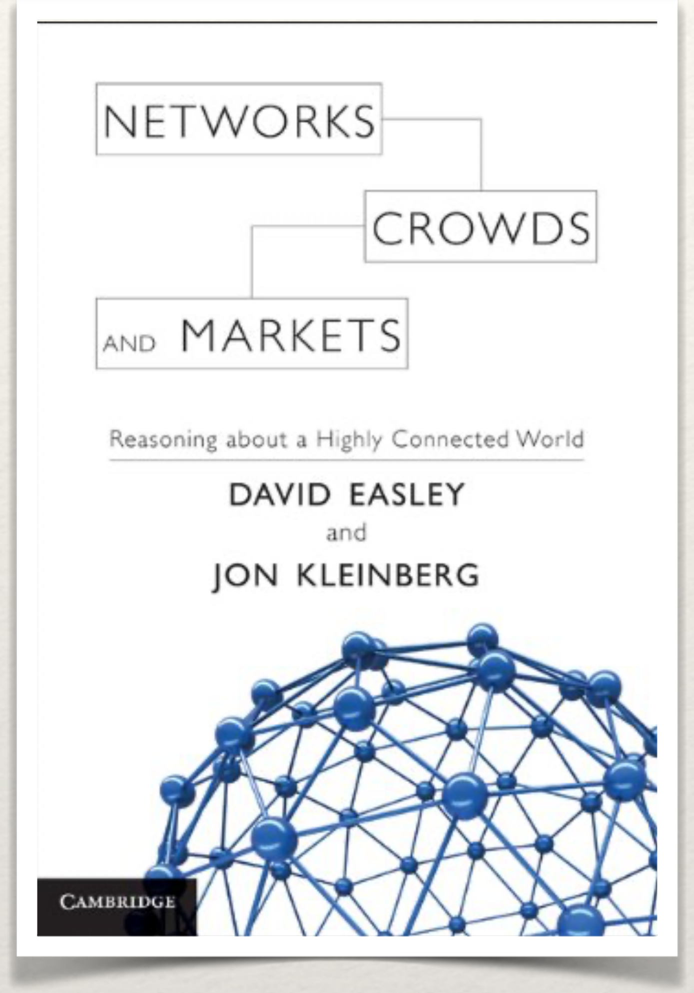
Lecture 24ns16

Link Analysis and Web Search

Course: **Complex Networks Analysis and Visualization**
Sub-Module: **NetSci**

-

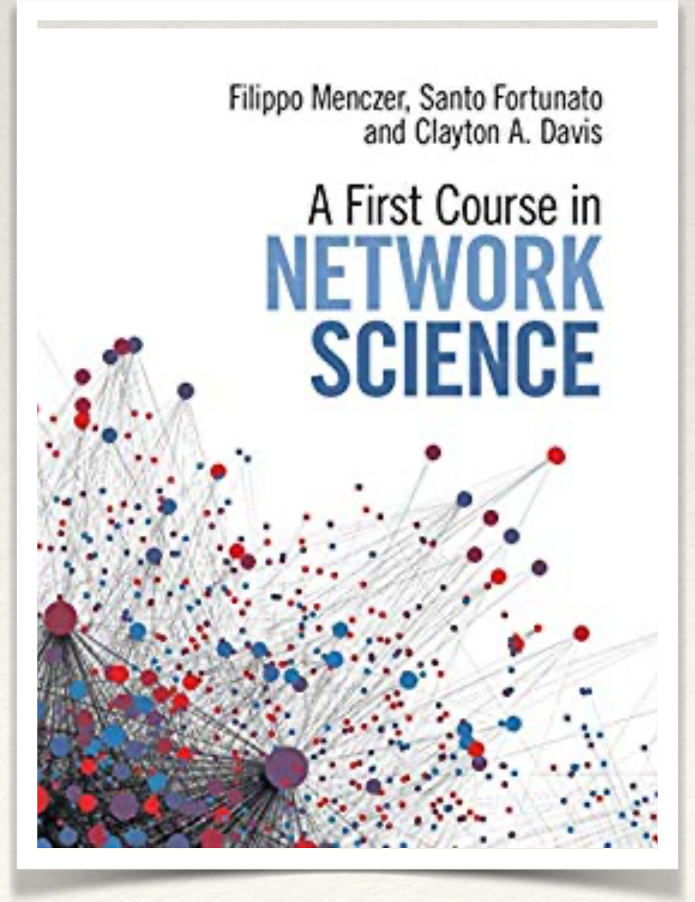
References



[ns2] Chapter 14 (14.1 - 14.6)

"Link Analysis and Web Search" —->

https://www.cs.cornell.edu/home/kleinber/networks-book/networks-book-ch14.pdf



[ns1] Chapter 4 (4.3)

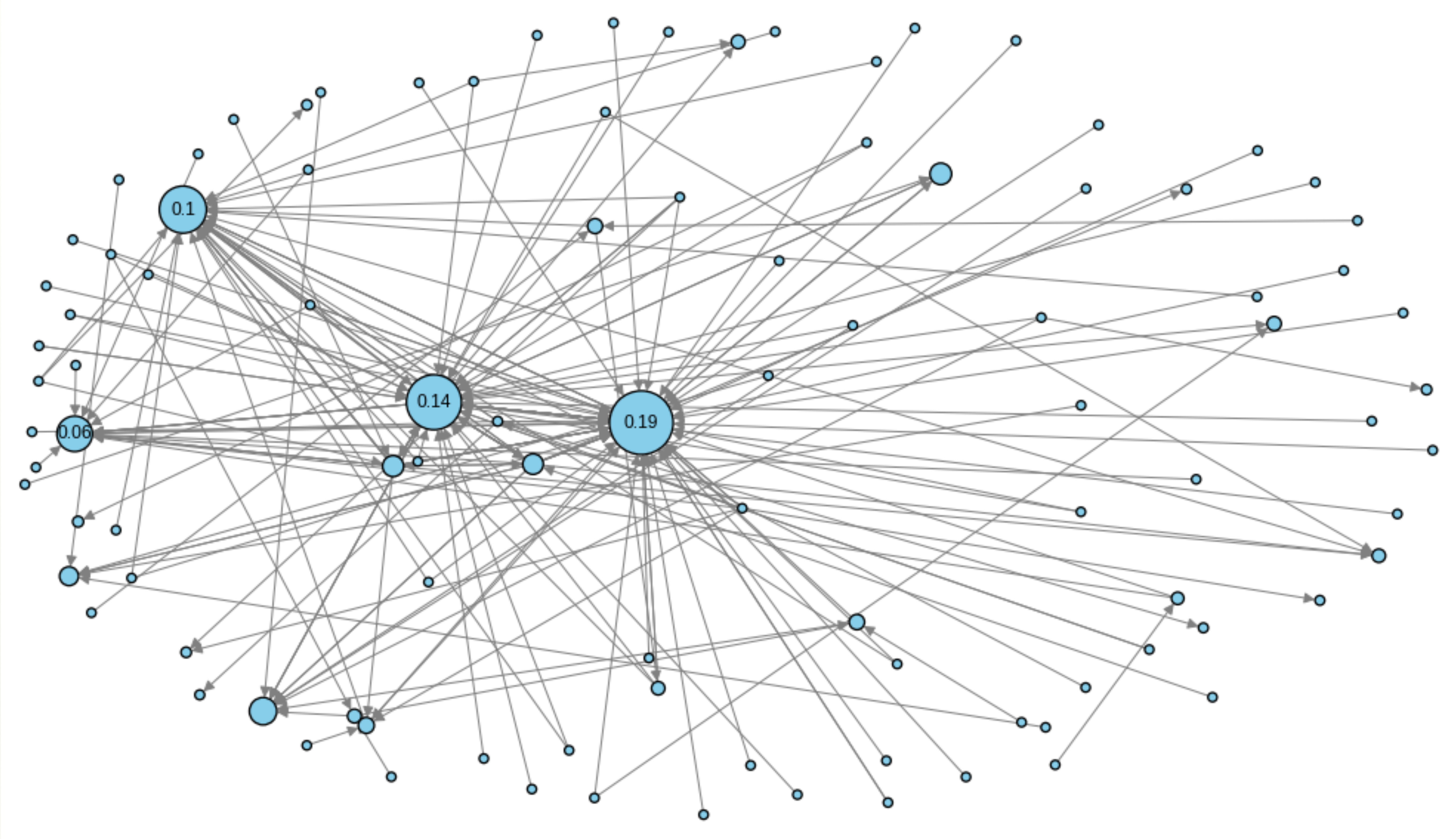
——> very simplified introduction to PageRank

Agenda

- Searching the Web
- Link Analysis
  - HITS: Hubs and Authorities (+ Spectral Analysis)
  - Page Rank (+ Spectral Analysis)
  - Random Walks and PR
- Practical implications
  - Modern Web search
  - Link Analysis beyond the Web

PageRank and HITS

- **Centrality** measures for nodes in directed networks
- Sergey Brin and Larry Page introduced PageRank in 1998 as a key ingredient of Google
- Jon Kleinberg introduced HITS in 1999
- Both are based on eigenvector centrality and designed for web information retrieval




```
PR_dict = nx.pagerank(D)      # D must be a DiGraph
```

```
H_dict, A_dict = nx.hits(G)    # G should be a DiGraph
```

Sources

- Page, Lawrence; Brin, Sergey; Motwani, Rajeev and Winograd, Terry, The PageRank citation ranking: Bringing order to the Web. 1999 http://dbpubs.stanford.edu:8090/pub/showDoc.Fulltext?lang=en&doc=1999-66&format=pdf
- Jon Kleinberg, Authoritative sources in a hyperlinked environment Journal of the ACM 46 (5): 604-32, 1999. doi:10.1145/324133.324140. http://www.cs.cornell.edu/home/kleinber/auth.pdf
- A. Langville and C. Meyer, "A survey of eigenvector methods of web information retrieval.", SIAM Review, vol. 47, No. 1https://epubs.siam.org/doi/pdf/10.1137/S0036144503424786

Searching the Web

- Search Engine
  - problem: how to rank (web) pages related to a given topic
- Information Retrieval
  - automated strategies to search in libraries, scientific papers, repositories, …
  - in response to keywords based queries
- List of keywords is "inexpressive" (e.g., polysemy, synonymy)
- "diversity": given a topic we find pages written by many kind of authors

Searching the Web (cont.'d)

- Pages are dynamic and always changing
- Filters: what is "important"?
- Can the structure of the Web, dominated by links, help us to find such "filters"?
  - first attempt: count words in documents
  - can we do better?

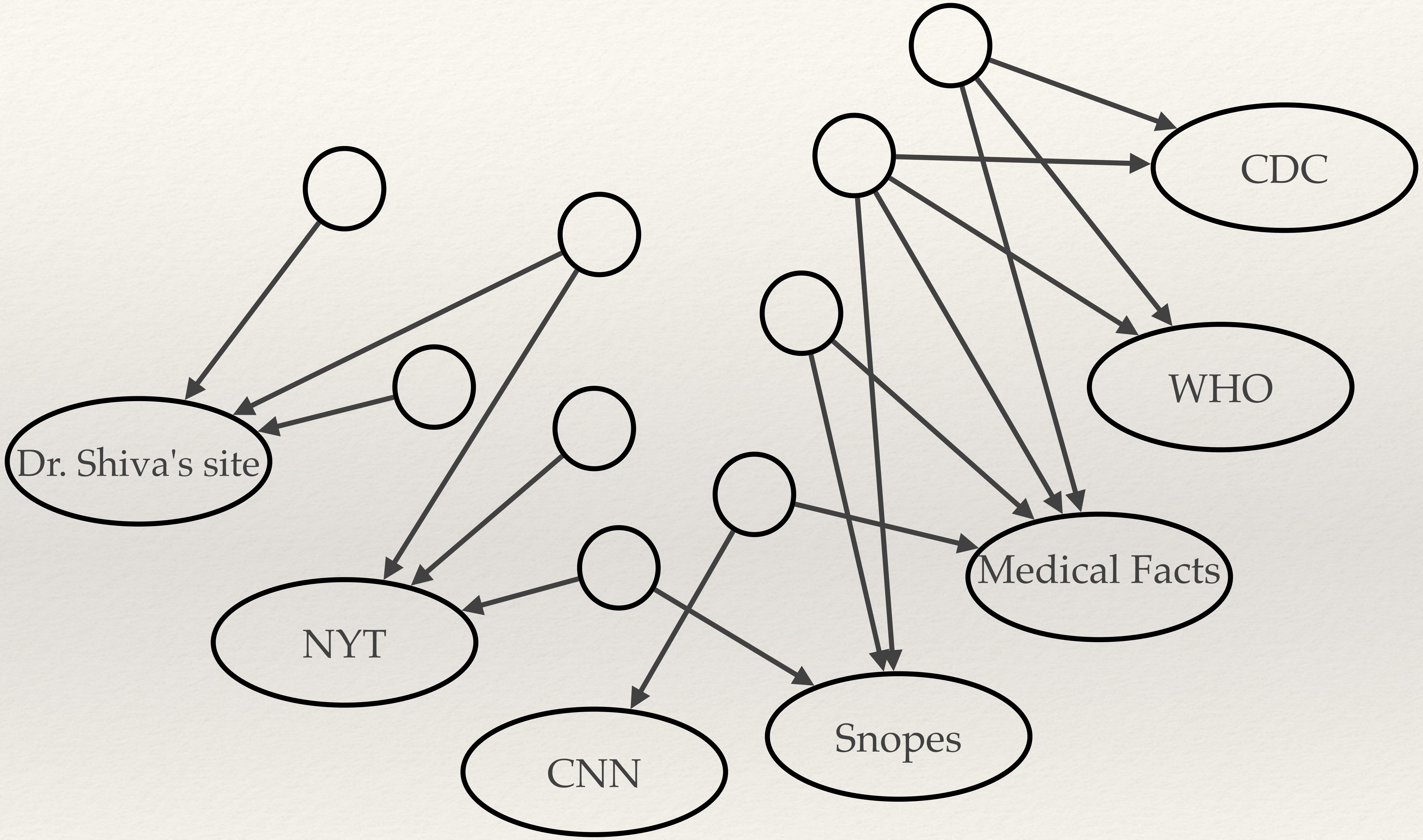
HITS

HITS: Link Analysis using Hubs and Authorities

- *Disclaimer: the term "Hub" is used slightly differently here w.r.t. previous lectures*
- Information contained "between" pages can be used as well
- Count "in-links":
  - select documents on a given topic
  - "in-links" are a measure of "authority" of a page on such a topic: it is an implicit "endorsement" from the community of web pages' authors

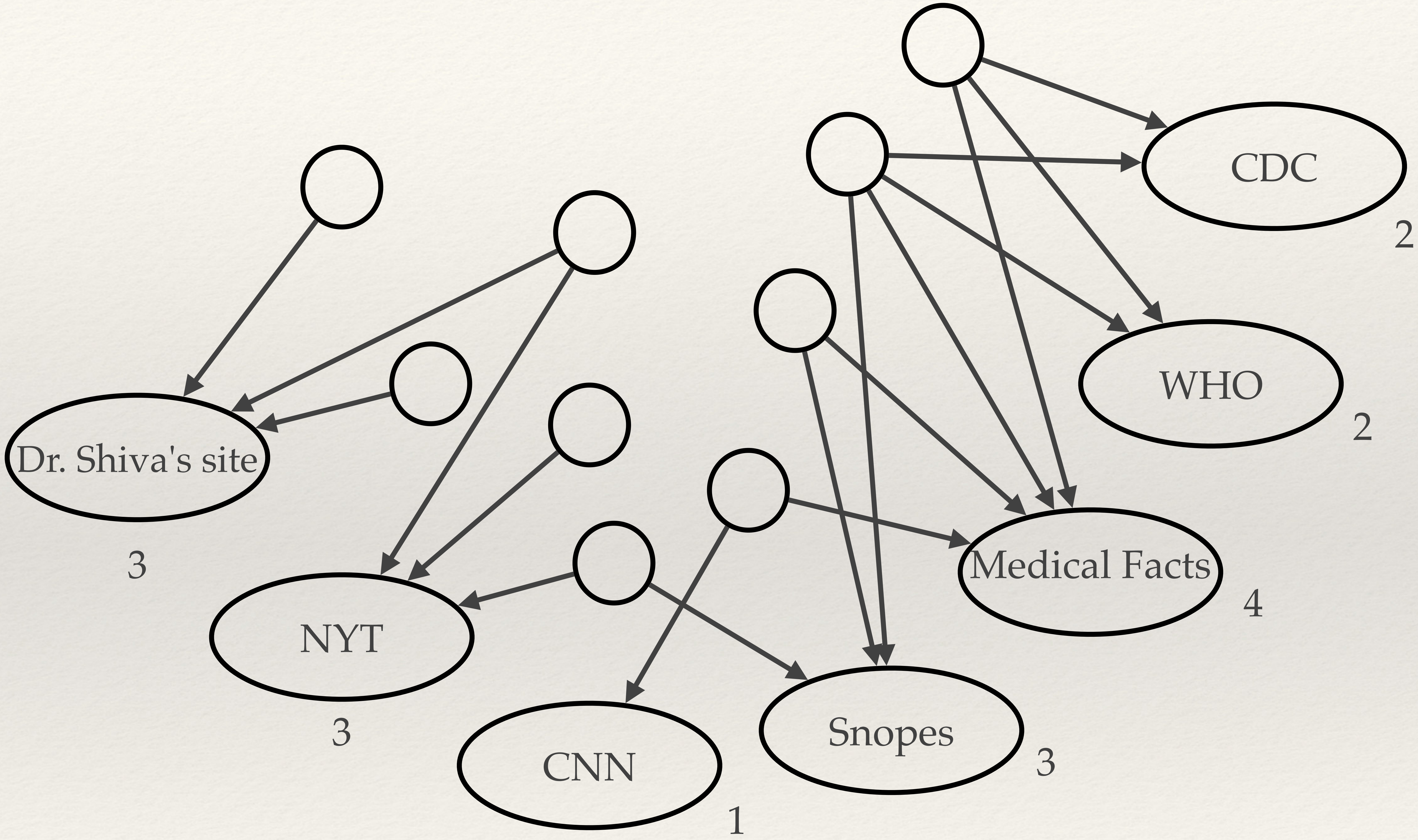
Finding lists

- query: "covid-19"
- we could have "intuitively correct" authority values with in-degrees
- "Lists": pages that provide many different out-links to other pages



List values

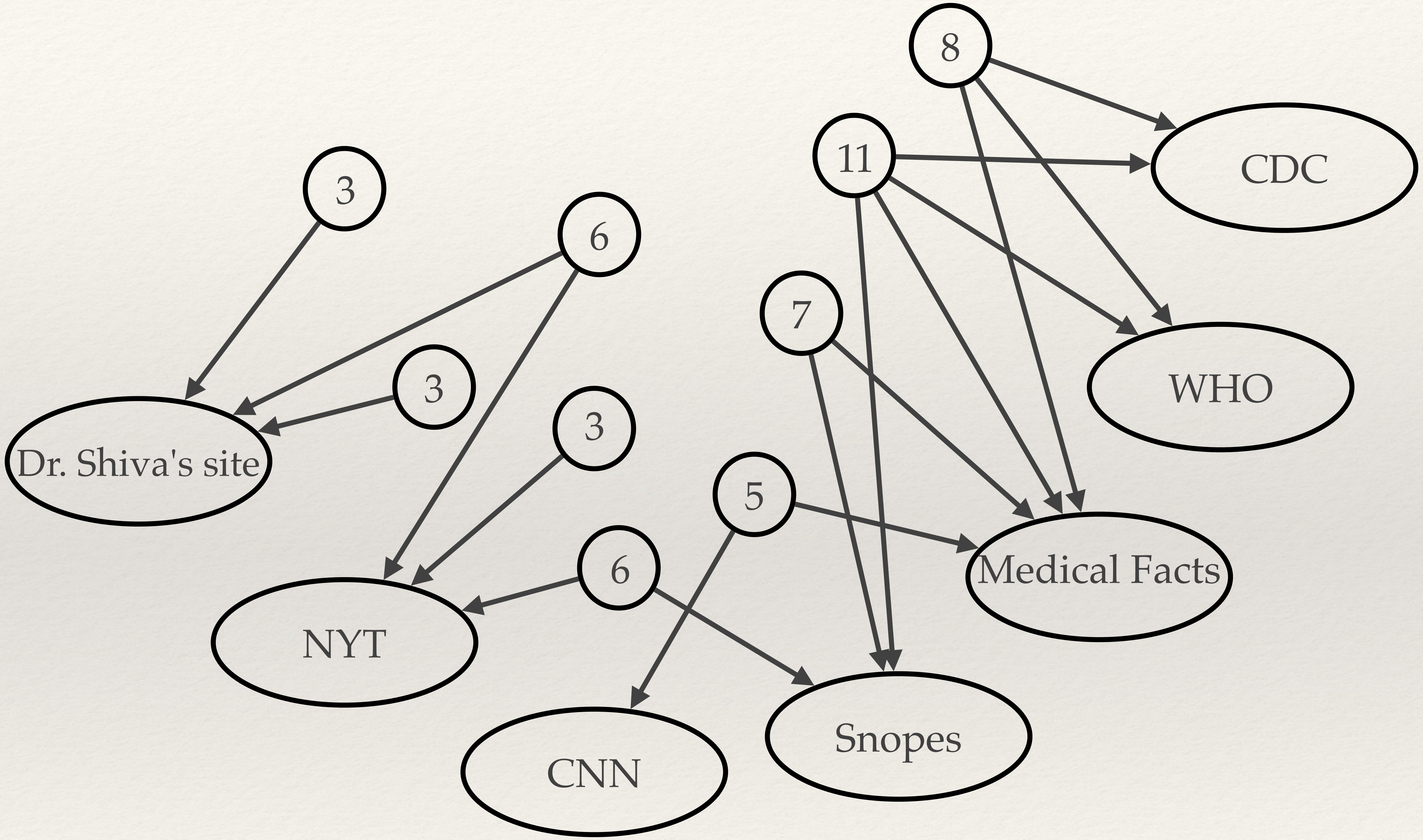
- Hub values for "list" pages: the sum of in-links received by all pages they link to
- Assumption: list pages have a better sense for where the "good" results are
  - "Authorities" are often competitors!



The principle of repeated improvement

- Now we can weight links from hubs more heavily

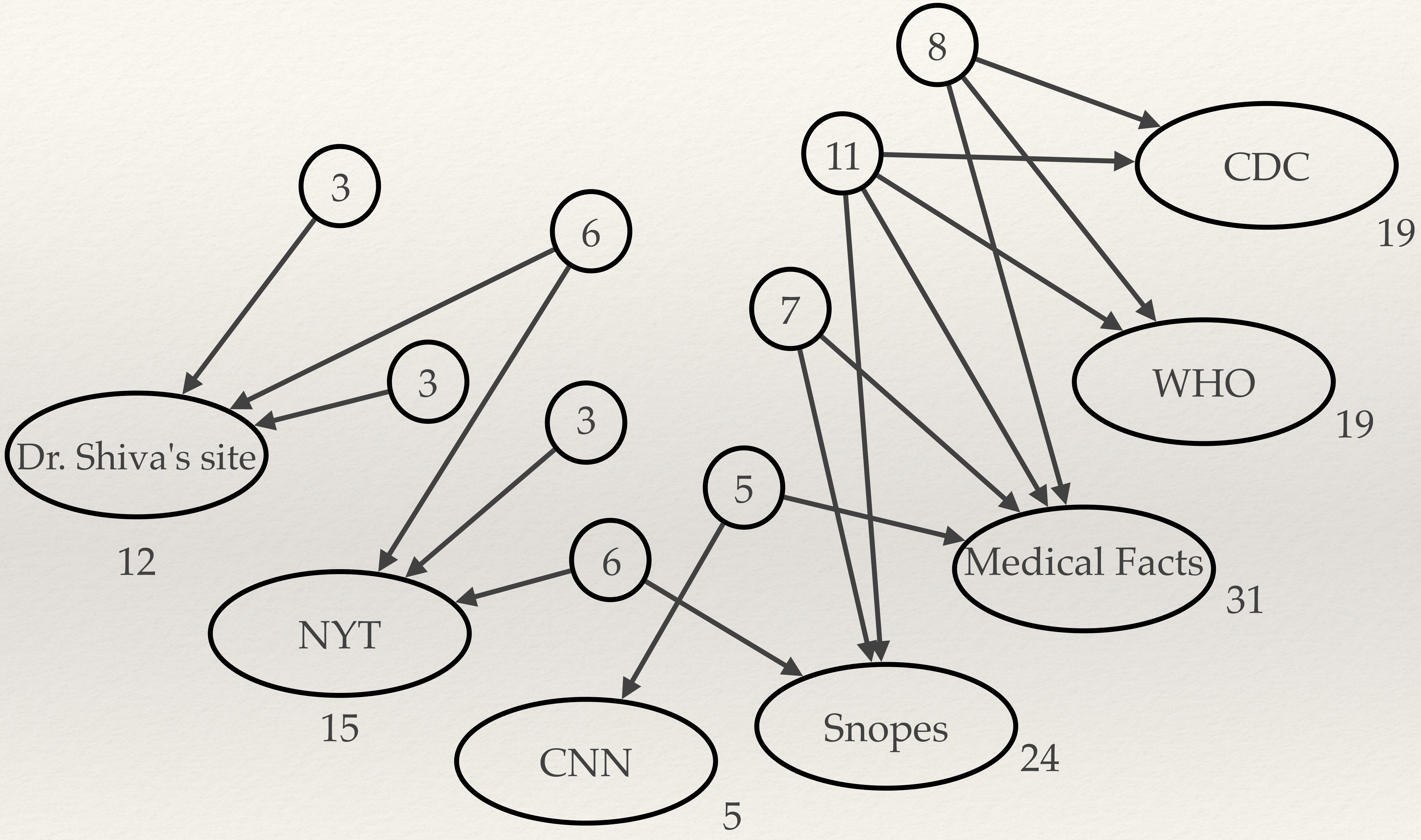
Recalculate =>



- Why stop here? we can refine values at both sides

Hubs and Authorities

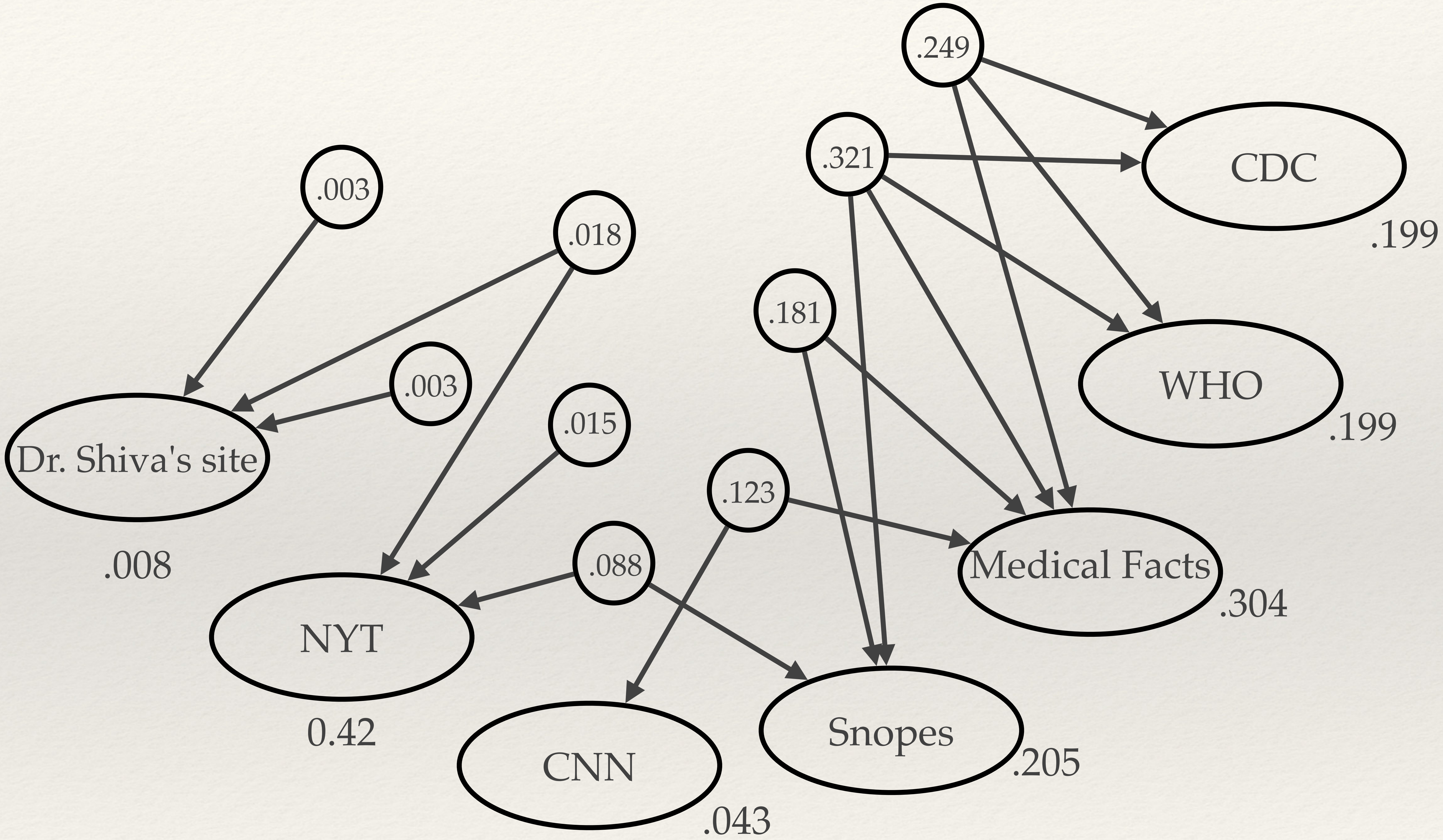
- step *0*: every node has both a hub and an *authority* value (all initialized to 1)
- step *k* (Update Rules):
  - authorities are updated after *k-1*'s hub values
  - hubs are updated on new authority values
  - normalize hub and authority values



Total (Authority): 125 Total (Hub): 52

Stabilization

- We will prove that normalized values **converge** when $i \to \infty$
- **Stabilization**: initial values are not important
  - limiting values for hubs and authorities are **properties of the links structure**
- Different form of game theoretical concept of **equilibrium**



Spectral Analysis of HITS

Introduction to spectral analysis

- We need to analyze the methods to compute **hubs and authorities** values
- Pre-requisites:
  - linear algebra
  - vector and matrix multiplication
- Limiting values are coordinates in eigenvectors for given eigenvalues in matrices derived from our graphs
- Eigenvalues/eigenvectors calculation to study the structure of networks => **spectral analysis**

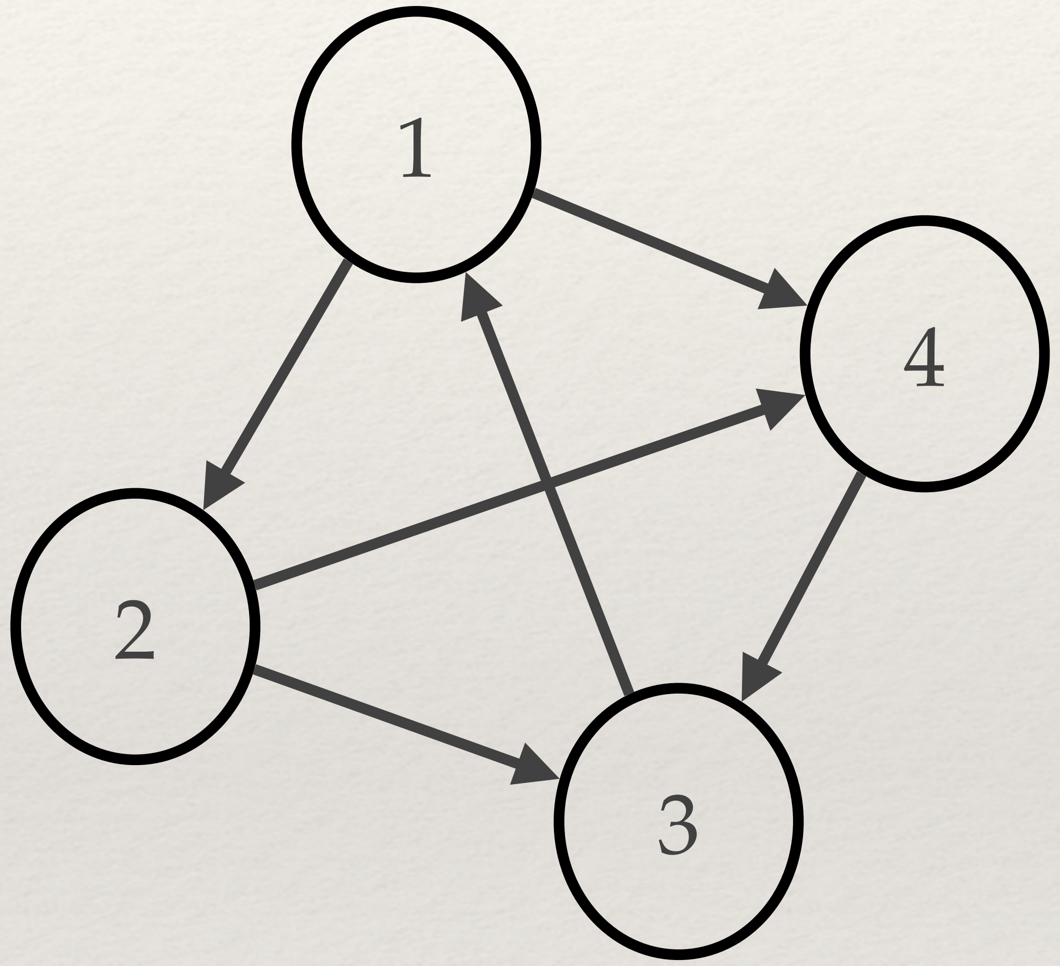
Spectral Analysis of Hubs and Authorities

def.: **Adjacency Matrix**(*)

nodes: *1, …, n*

M : n x n

$$M_{ij} = \begin{cases} 1, & \text{if } (i, j) \text{ is an edge} \\ 0, & \text{otherwise} \end{cases}$$

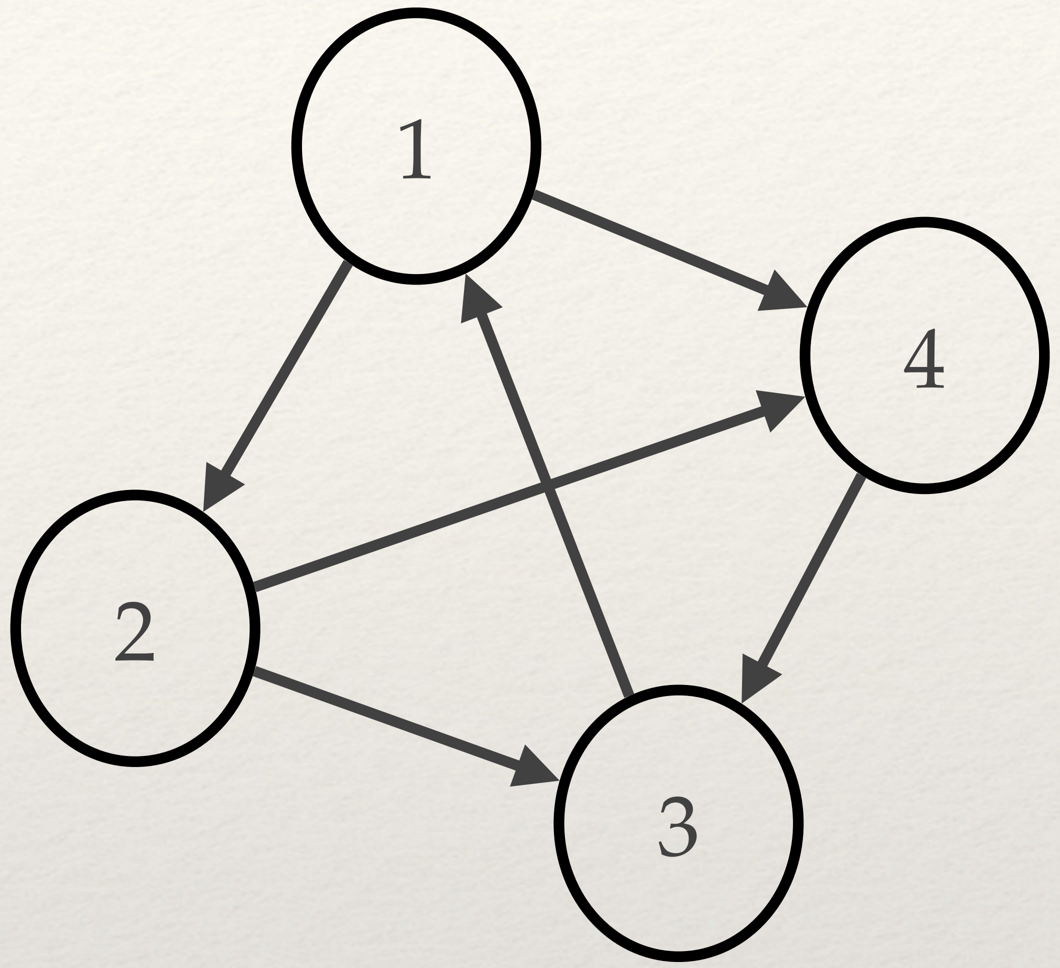


$$\begin{bmatrix} 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

(*) not necessarily efficient for computational representations (for practical use, consider adjacency lists or edge lists instead)

Update rules

h, **a**: n-dimensional vectors (resp., hub and authorities values)



Hub Update Rule

$$h_i \leftarrow \sum_{j=1}^{n} M_{ij} a_j$$

$$= M_{i1}a_1 + M_{i1}a_2 + \ldots + M_{in}a_n$$

$$\mathbf{h} \leftarrow \mathbf{M} \cdot \mathbf{a}$$

$$\underset{\mathbf{M}}{\begin{bmatrix} 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}} \cdot \underset{\mathbf{a}}{\begin{bmatrix} 1 \\ 1 \\ 2 \\ 2 \end{bmatrix}} = \underset{\mathbf{h}}{\begin{bmatrix} 3 \\ 4 \\ 1 \\ 2 \end{bmatrix}}$$

Authority Update Rule

$$a_i \leftarrow \sum_{j=1}^{n} M_{ji} h_j$$

$$\mathbf{a} \leftarrow \mathbf{M}^{\mathrm{T}} \cdot \mathbf{h}$$

Understanding the k-step hub-authority computation

init: $\mathbf{h}^{<0>} = \underbrace{(1,1,\ldots,1)}_{n}$

after k applications of update rules: $\mathbf{a}^{<k>}, \mathbf{h}^{<k>}$

1st: $\mathbf{a}^{<1>} = \mathbf{M}^{\mathbf{T}}\mathbf{h}^{<0>}$

$$\mathbf{h}^{<1>} = \mathbf{M}\mathbf{a}^{<1>} = \mathbf{M}\mathbf{M}^{\mathbf{T}}\mathbf{h}^{<0>}$$

2nd: $\mathbf{a}^{<2>} = \mathbf{M}^{\mathbf{T}}\mathbf{h}^{<1>} = (\mathbf{M}\mathbf{M}^{\mathbf{T}})\mathbf{M}^{\mathbf{T}}\mathbf{h}^{<0>}$

$$\mathbf{h}^{<2>} = \mathbf{M}\mathbf{a}^{<2>} = (\mathbf{M}\mathbf{M}^{\mathbf{T}})(\mathbf{M}\mathbf{M}^{\mathbf{T}})\mathbf{h}^{<0>} = (\mathbf{M}\mathbf{M}^{\mathbf{T}})^{2}\mathbf{h}^{<0>}$$

…

k-th: $\mathbf{a}^{<k>} = (\mathbf{M}\mathbf{M}^{\mathbf{T}})^{k-1}\mathbf{M}^{\mathbf{T}}\mathbf{h}^{<0>}$

$$\mathbf{h}^{<k>} = (\mathbf{M}\mathbf{M}^{\mathbf{T}})^{k}\mathbf{h}^{<0>}$$

a, h vectors: multiplication of an initial vector $\mathbf{h}^{<0>}$ by larger and larger powers of $\mathbf{M}^{\mathbf{T}}\mathbf{M}$ and $\mathbf{M}\mathbf{M}^{\mathbf{T}}$

Multiplications and Eigenvectors

normalization: we can find constants c and d s.t. $\dfrac{\mathbf{h}^{<\mathbf{k}>}}{c^k}$ and $\dfrac{\mathbf{a}^{<\mathbf{k}>}}{d^k}$

we want to prove that they **converge** for $k \to \infty$

Focus on hub vectors:

if $\dfrac{\mathbf{h}^{<\mathbf{k}>}}{c^k} = \dfrac{(\mathbf{M}\mathbf{M}^\mathbf{T})^k \cdot \mathbf{h}^{<0>}}{c^k}$ converges to a limit $\mathbf{h}^{<*>}$, then I can expect that:

$$c \cdot \mathbf{h}^{<*>} = (\mathbf{M}\mathbf{M}^\mathbf{T}) \cdot \mathbf{h}^{<*>}$$

Hence, we need to prove that the sequence of $\dfrac{\mathbf{h}^{<\mathbf{k}>}}{c^k}$ converges to the eigenvector of $\mathbf{M}\mathbf{M}^\mathbf{T}$

Eigenvectors and square matrices

- Observe that $\mathbf{MM^T}$ is a symmetric matrix
- **fact 1**: "Any symmetric matrix n x n has a set of n eigenvectors $\mathbf{z_1}, \cdots, \mathbf{z_n}$ that are orthogonal and all unit vectors - that is they form a basis for the space $\mathbb{R}^n$
  => $\mathbf{z_i} \cdot \mathbf{z_j} = 0$ and $\mathbf{z_i} \cdot \mathbf{z_i} = 1$
- That means that for our symmetric $\mathbf{MM^T}$ we can find:
  - n mutual orthogonal eigenvectors: $\mathbf{z_1}, \cdots, \mathbf{z_n}$ ← the *spectrum* of $\mathbf{MM^T}$
  - n corresponding eigenvalues: $c_1, \cdots, c_n$
  - Let's sort eigenvectors s.t. corresponding eigenvalues: $c_1 \geq c_2 \geq \cdots \geq c_n$
  - Assume (for now): $c_1 > c_2$

- Let's consider $\mathbf{x} \in \mathbb{R}^n$
  $\mathbf{x} = p_1\mathbf{z_1} + p_2\mathbf{z_2} + \ldots + p_n\mathbf{z_n}$ (for $p_1, p_2, \ldots, p_n$ coefficients)

- $$\begin{aligned}(\mathbf{MM^T})\mathbf{x} &= (\mathbf{MM^T})(p_1\mathbf{z_1} + p_2\mathbf{z_2} + \ldots + p_n\mathbf{z_n}) \\ &= p_1(\mathbf{MM^T})\mathbf{z_1} + p_2(\mathbf{MM^T})\mathbf{z_2} + \ldots + p_n(\mathbf{MM^T})\mathbf{z_n} \\ &= p_1c_1\mathbf{z_1} + p_2c_2\mathbf{z_2} + \ldots + p_nc_n\mathbf{z_n}\end{aligned}$$

- We will it use this equation to analyze multiplication by larger powers of $(\mathbf{MM^T})$

  $$(\mathbf{MM^T})^{\mathbf{k}}\mathbf{x} = p_1c_1^k\mathbf{z_1} + p_2c_2^k\mathbf{z_2} + \ldots + p_nc_n^k\mathbf{z_n}$$

Convergence of the Hub-Authority computation

1

vector of hub scores at step k:

$$\mathbf{h}^{<\mathbf{k}>} = (\mathbf{M}\mathbf{M}^{\mathbf{T}})^k \cdot \mathbf{h}^{<\mathbf{0}>}$$

$$\mathbf{h}^{<\mathbf{0}>} = q_1\mathbf{z_1} + q_2\mathbf{z_2} + \ldots + q_n\mathbf{z_n}$$

$$\mathbf{h}^{<\mathbf{k}>} = c_1^k q_1\mathbf{z_1} + c_2^k q_2\mathbf{z_2} + \ldots + c_n^k q_n\mathbf{z_n}$$

2

Let's divide both sides by c_1^k

$$\frac{\mathbf{h}^{<\mathbf{k}>}}{c_1^k} = \frac{c_1^k q_1\mathbf{z_1}}{c_1^k} + \frac{c_2^k q_2\mathbf{z_2}}{c_1^k} + \ldots + \frac{c_n^k q_n\mathbf{z_n}}{c_1^k}$$

assumption: $c_1 > c_2 \quad \Rightarrow \lim_{k\to\infty} \left(\frac{c_2}{c_1}\right)^k = 0$

$$lim_{k\to\infty}\frac{\mathbf{h}^{<\mathbf{k}>}}{c_1^k} = q_1\mathbf{z_1}$$

Wrapping up

(i) a limit in the direction of $\mathbf{z_1}$ is reached regardless of initial values of $\mathbf{h}^{<0>}$: let's suppose that $\mathbf{h}^{<0>} = \mathbf{x}$ and that is a positive vector:

$$\mathbf{x} = p_1\mathbf{z_1} + p_2\mathbf{z_2} + \ldots + p_n\mathbf{z_n}, \quad \Rightarrow (\mathbf{MM^T})^{\mathbf{k}}\mathbf{x} = c_1^k p_1\mathbf{z_1} + c_2^k p_2\mathbf{z_2} + \ldots + c_n^k p_n\mathbf{z_n}$$

$$lim_{k\to\infty}\frac{\mathbf{h}^{<\mathbf{k}>}}{c_1^k} = p_1\mathbf{z_1}$$

(ii) coefficient p_1 (or q_1) must be $\neq 0$: assuring that $p_1\mathbf{z_1}$ (or $q_1\mathbf{z_1}$) are non zero vectors, in the direction of $\mathbf{z_1}$ => textbook

(iii) relax assumption: $c_1 > c_2$

in general we can have $l > 1$ eigenvalues s.t. $c_1 = c_2 = \ldots = c_l$, until we find that $c_1 > c_{l+1}$

$$\frac{\mathbf{h}^{<\mathbf{k}>}}{c_1^k} = \frac{c_1^k q_1 \mathbf{z_1} + c_2^k q_2 \mathbf{z_2} + \ldots + c_l^k q_l \mathbf{z_l}}{c_1^k} + \frac{c_{l+1} n^k q_{l+1} \mathbf{z_{l+1}} + \ldots\ldots + c_n^k q_n \mathbf{z_n}}{c_1^k}$$

$$k \to \infty$$

$$= q_1 \mathbf{z_1} + q_2 \mathbf{z_2} + \ldots + q_l \mathbf{z_l} + 0$$

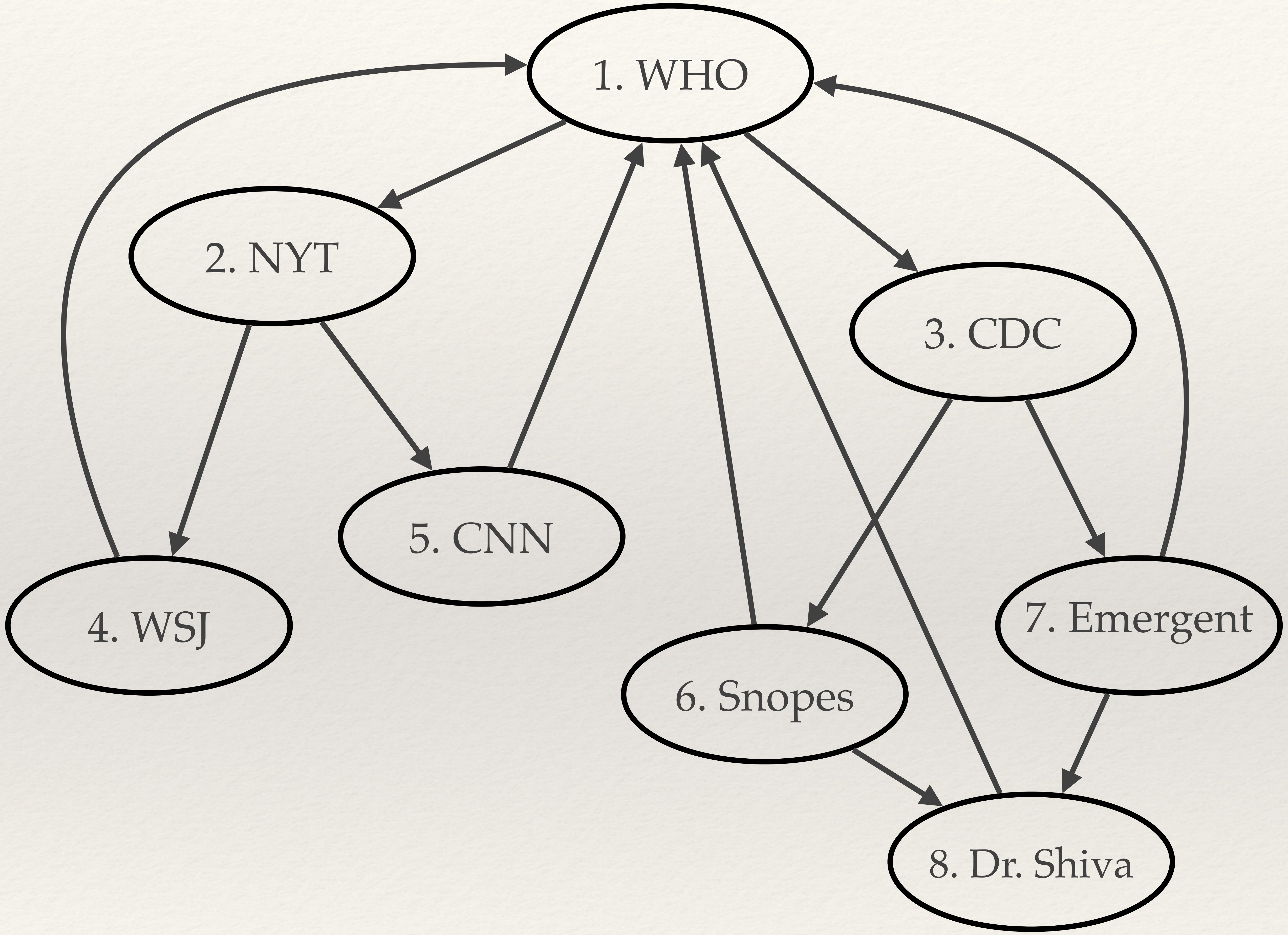
this is still a convergence

(iv) authority values: the argument is vary similar to hub values (multiplication by $\mathbf{M^T M}$)

Page Rank

Page Rank

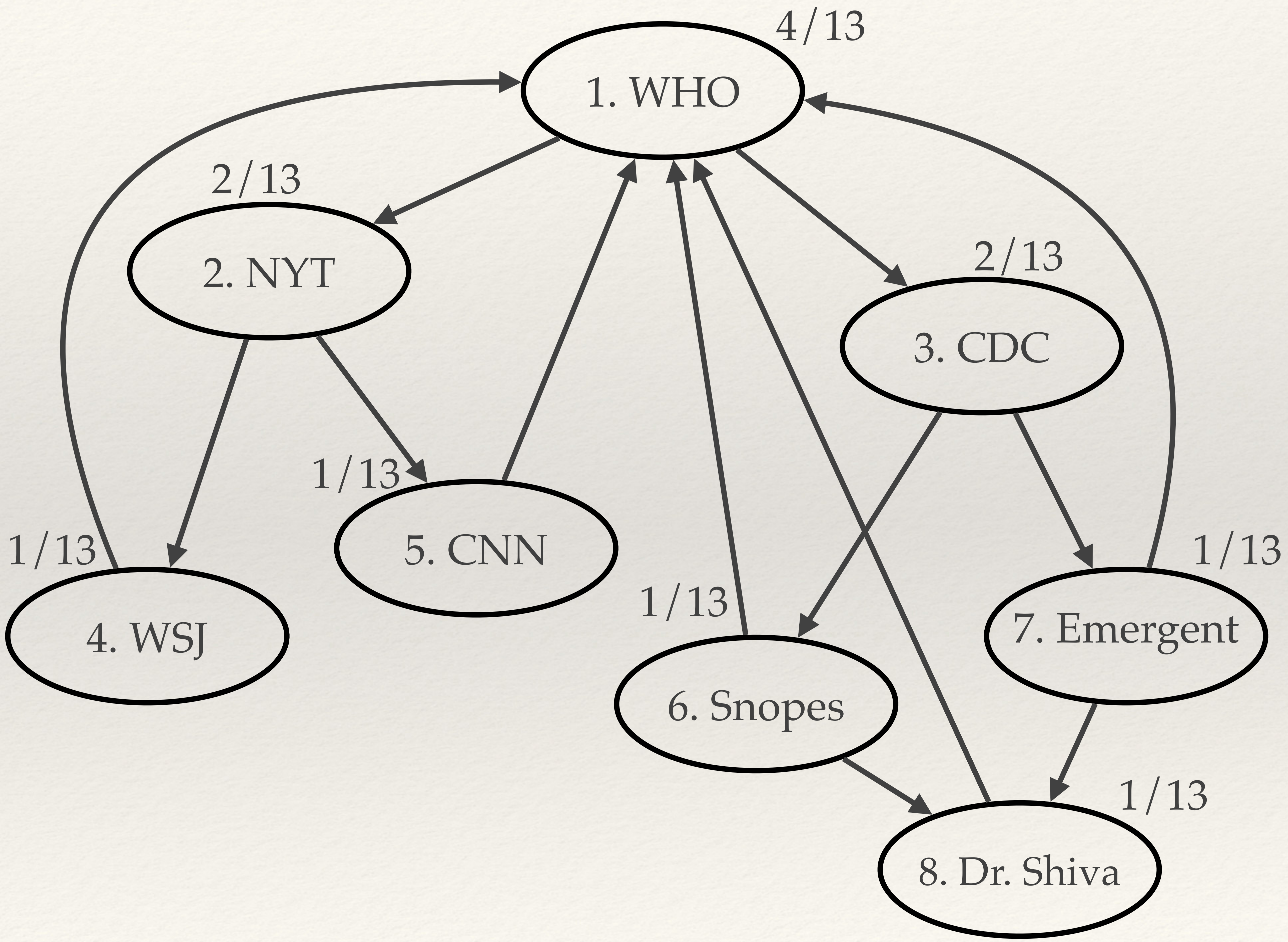
- "Endorsement" viewed as passing directly from one "important" node to another
  - endorsements received by in-links and passed across outgoing links
- Basic definition:
  - Step 0: Init all the pages p to a $PR(p) = \frac{1}{n}$, where n is the number of pages
  - Step k: Update all the $PR(p)$ to the sum of all the receiving PR values, normalized by out-links



	1	2	3	4	5	6	7	8
0	1/8	1/8	1/8	1/8	1/8	1/8	1/8	1/8
1	1/2	1/16	1/16	1/16	1/16	1/16	1/16	1/8
2	3/16	1/4	1/4	1/32	1/32	1/32	1/32	1/16
	…	…	…	…	…	…	…	…

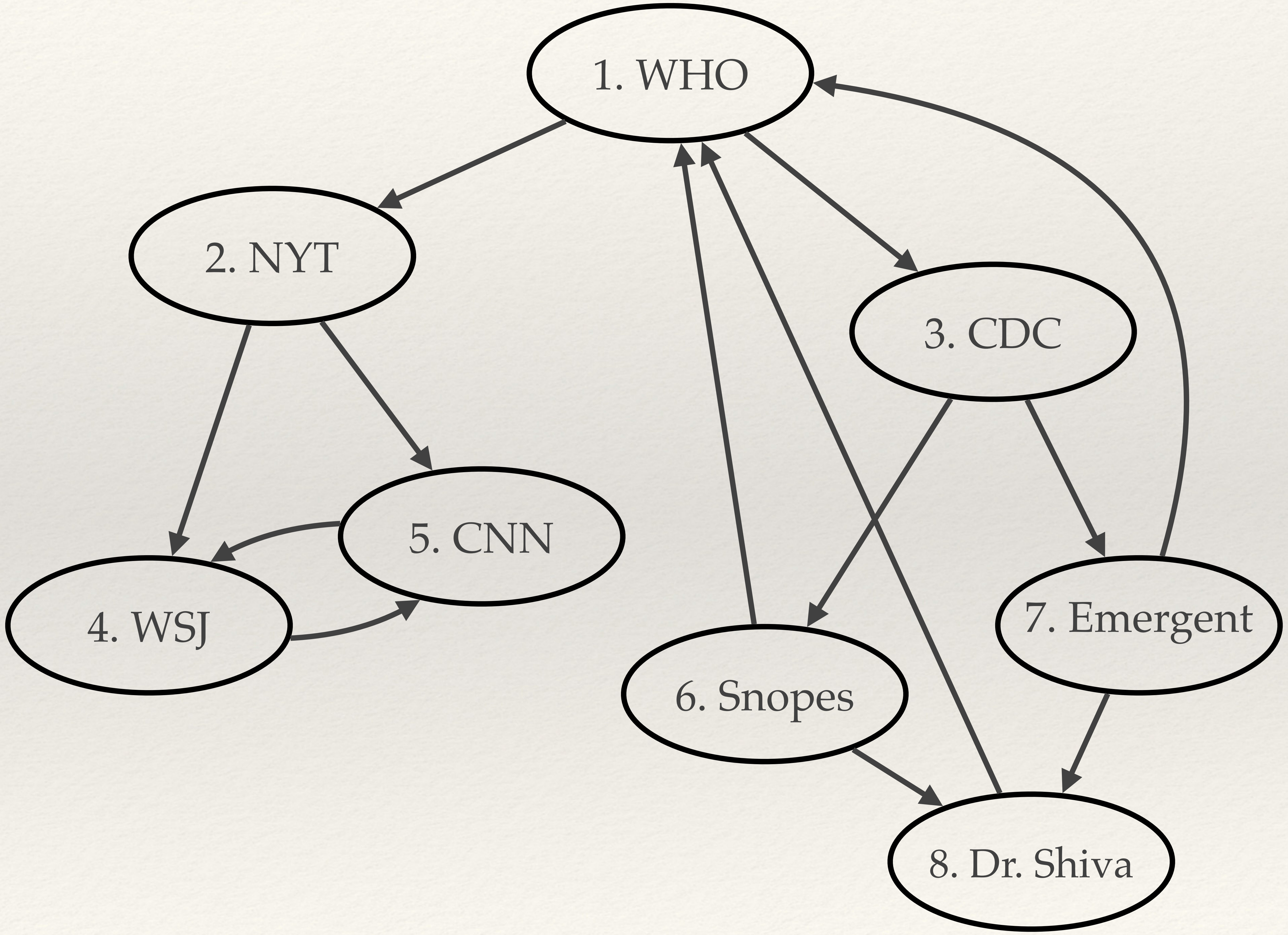
PageRank and stabilization

- PR values of all the nodes converge when $k \to \infty$ (but for some "degenerate cases")
- Equilibrium: if we apply our PR update rule, then our limiting values do not change



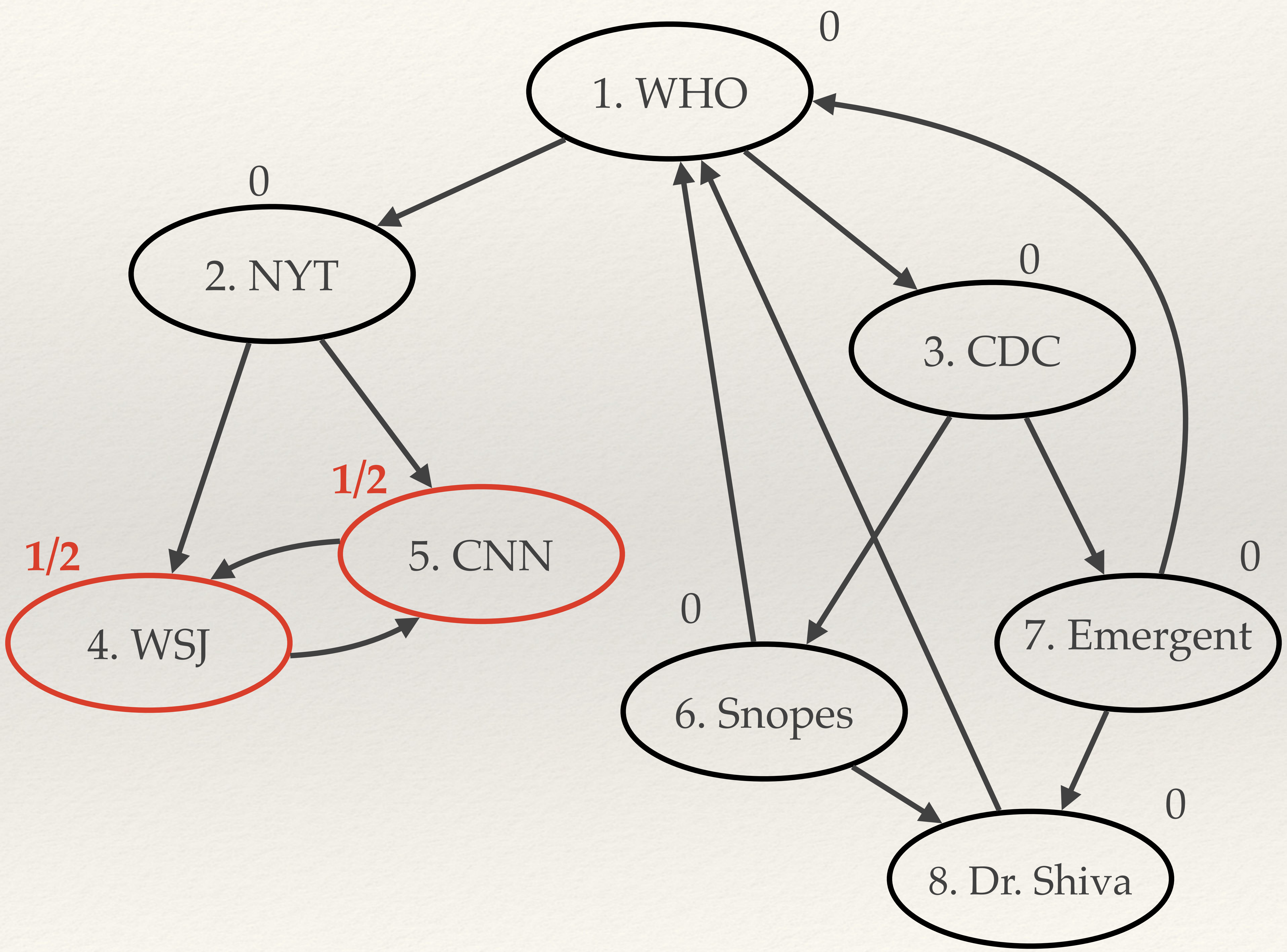
Scaling the definition of PageRank

- "degenerate cases", the problem: in some networks some nodes receive all the PR values of the the network



Scaling the definition of PageRank

- "degenerate cases", the problem: in some networks some nodes receive all the PR values of the the network
- Applying PR update rule until we get equilibrium =>
- I can have degenerate cases in the Out Component of the Web



- Why? We do not have path back to some other nodes
- Solution:
  let's force this "fluid" to stream back to other nodes "sometimes":
  - select a **scaling factor** (aka **damping factor**) $s : s \in [0,1]$
  - get a portion s of PR values from in-links and then add $(1-s)\frac{1}{n}$
  - Now we have convergence for $k \to \infty$
  - Observe: typically $s \in [0.8, 0.9]$

Spectral Analysis of Page Rank

Introduction to spectral analysis

- We need to analyze the methods to compute **page rank** values
- Pre-requisites:
  - linear algebra
  - vector and matrix multiplication
- Limiting values are coordinates in eigenvectors for given eigenvalues in matrices derived from our graphs
- Eigenvalues/eigenvectors calculation to study the structure of networks => **spectral analysis**

Page Rank (revisited)

At step 0 (init):

$$\forall i : r_i = \frac{1}{n};\ \text{n: \# pages}$$

$$r_i = PR(i)$$

At step k:

$$\forall i : r_i = \sum_{j=1}^{n} M_{ji} \frac{r_j}{k_j^{\text{out}}}$$

(basic PR update rule)

$$\forall i : r_i = s \cdot \sum_{j=1}^{n} M_{ji} \frac{r_j}{k_j^{\text{out}}} + (1 - s) \cdot \frac{1}{n}$$

(scaled PR update rule)

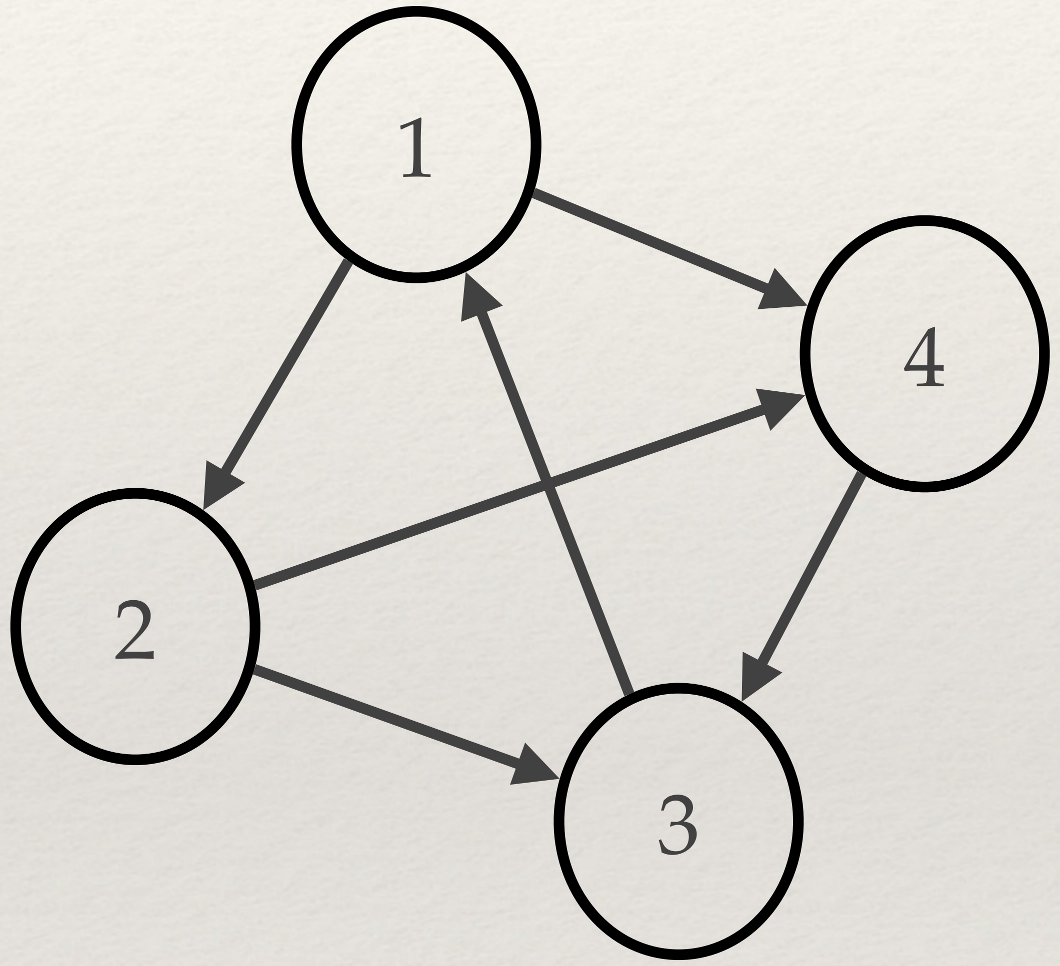
Using matrix notation

N: Matrix derived from M

nodes: *1, …, n*

N *: n* x *n*

$$N_{ij} = \begin{cases} \frac{1}{k_i^{\text{out}}}, & \text{if } (i, j) \\ 1, & \text{if } (i, j) \text{ is an edge, and } k_i^{\text{out}} = 0 \\ 0, & \text{otherwise} \end{cases}$$



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

N_{ij}: the share of *i*'s PR that *j* should get in one update step

Update rule (basic and scaled)

$\boxed{1}$ Basic update rule:

$$\forall i : r_i = \sum_{j=1}^{n} N_{ji} r_j$$

$$\leftarrow N_{1i} r_1 + N_{2i} r_2 + \ldots + N_{1n} r_n$$

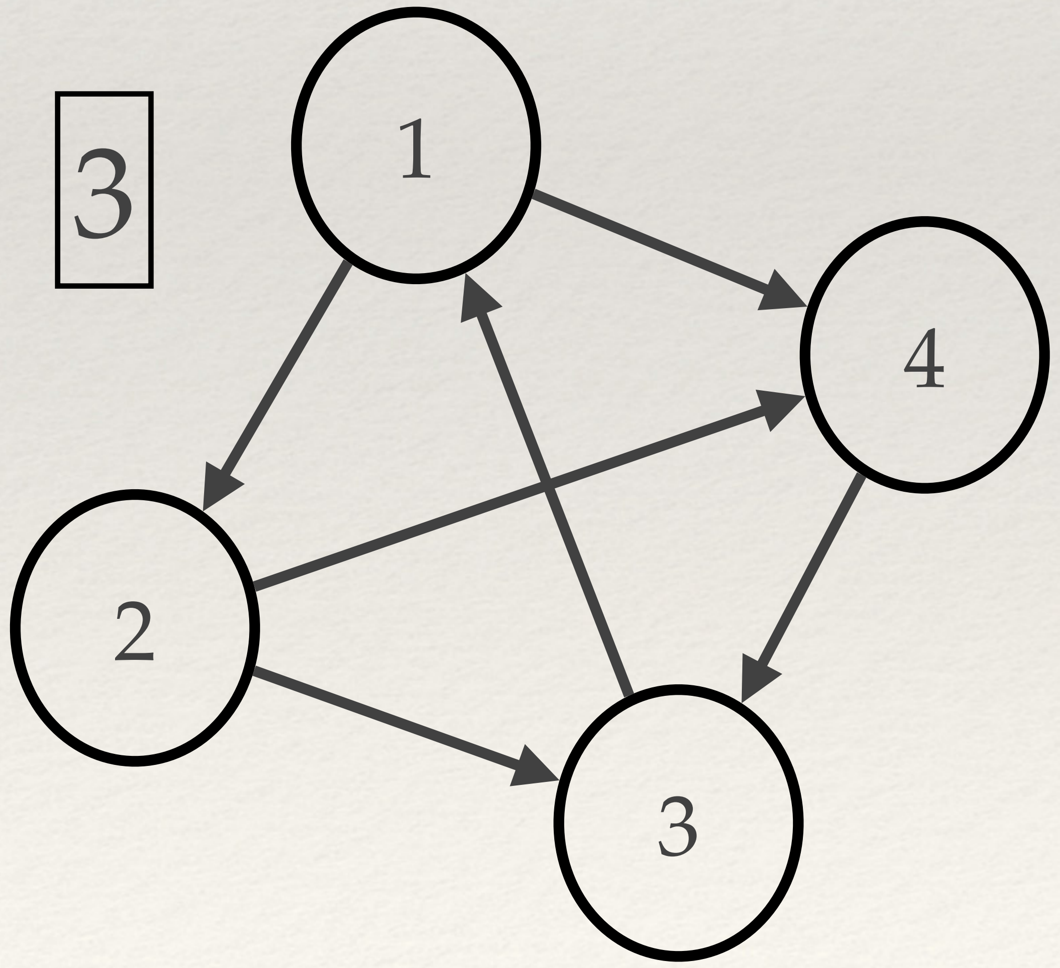
$$\mathbf{r} \leftarrow \mathbf{N}^{\mathbf{T}} \cdot \mathbf{r}$$

$\boxed{2}$ Scaled update rule (factor s):

$$\widetilde{N}_{ij} = s \cdot N_{ij} + (1 - s) \cdot \frac{1}{n}$$

$s = 0.85$

$\boxed{3}$



$$\begin{bmatrix} .05 & .45 & .05 & .45 \\ .05 & .05 & .45 & .45 \\ .85 & .05 & .05 & .05 \\ .05 & .05 & 0.85 & .05 \end{bmatrix}$$

$\boxed{4}$ Application of scaled update rule:

$$\forall i : r_i = \sum_{j=1}^{n} \widetilde{N}_{ji} r_j$$

$$\leftarrow \widetilde{N}_{1i} r_1 + \widetilde{N}_{2i} r_2 + \ldots + \widetilde{N}_{1n} r_n$$

$$\mathbf{r} \leftarrow \widetilde{\mathbf{N}}^{\mathbf{T}} \cdot \mathbf{r}$$

Repeated improvement

$\mathbf{r}^{<0>} = \left(\frac{1}{n}, \ldots, \frac{1}{n}\right)$, initial PR vector

$\mathbf{r}^{<\mathbf{k}>} = (\widetilde{\mathbf{N}}^{\mathbf{T}})^k \cdot \mathbf{r}^{<0>}$

Limiting vector $r^{<*>}$ satisfies $\widetilde{\mathbf{N}}^{\mathbf{T}} \cdot \mathbf{r}^{<*>} = 1 \cdot \mathbf{r}^{<*>}$

$\mathbf{r}^{<*>}$ should be an eigenvector of $\widetilde{\mathbf{N}}^{\mathbf{T}}$ with corresponding eigenvalue of 1 BUT $\widetilde{\mathbf{N}}^{\mathbf{T}}$ is not symmetric: this means that eigenvalues can be complex numbers and eigenvectors have no relationships to one another

Convergence of the scaled PR update rule

$\forall i,j : \widetilde{N}_{ij} > 0$

Perron's theorem

Matrix **P** (with entries > 0)

i) **P** has an eigenvalue $c > 0$ s.t. $c > c' \ \forall c'$ (with c' another eigenvalue)

ii) Exists an eigenvector **y** with real positive values corresponding to c, and **y** is unique (up to a multiplication constant)

iii) if $c = 1$, then for any starting vector $\mathbf{x} \neq 0$ with non negative coordinates, the sequence of vectors $p^k\mathbf{x}$ converges to a vector in the direction of **y** ($k \to \infty$)

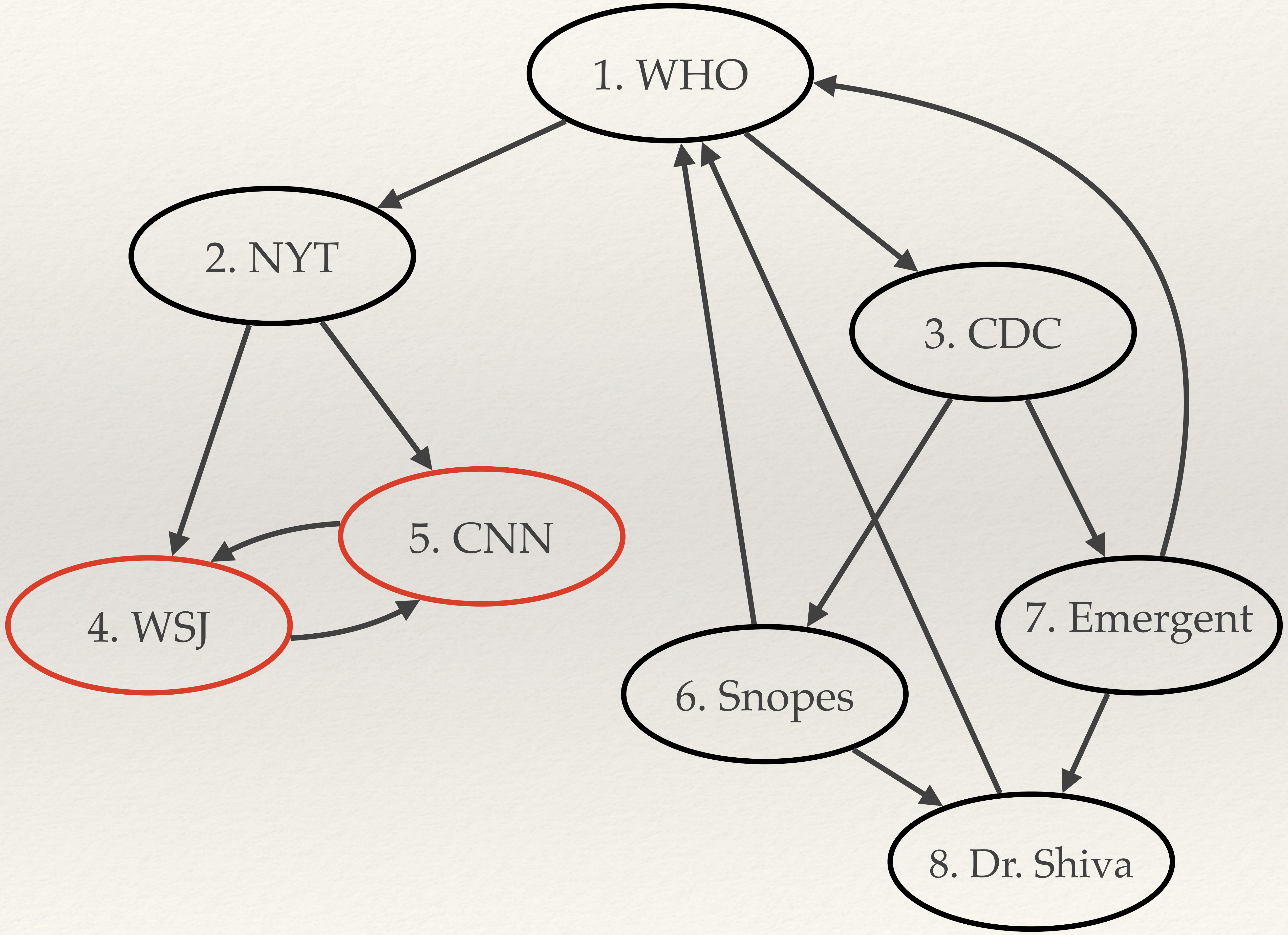
Random Walks and Page Rank

Random walks

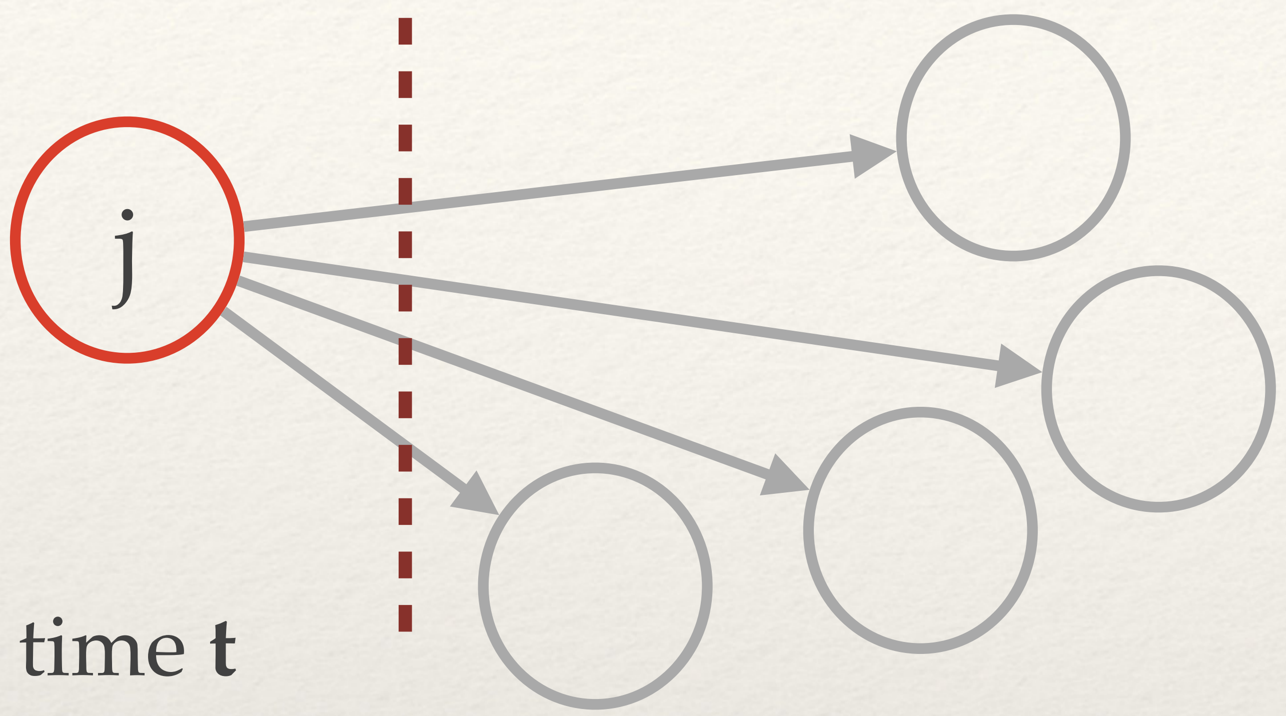
- randomly clicking from one page to another, picking each page with equal probability
- follow links for a sequence of length k
- **claim**: the probability of being at page x after k steps is the application of the basic PR update rule
- **additional intuition**: PR(x) is the limiting probability that a random walk across hyperlinks will end up at x as we sum the walk for larger and larger number of steps

Leakage

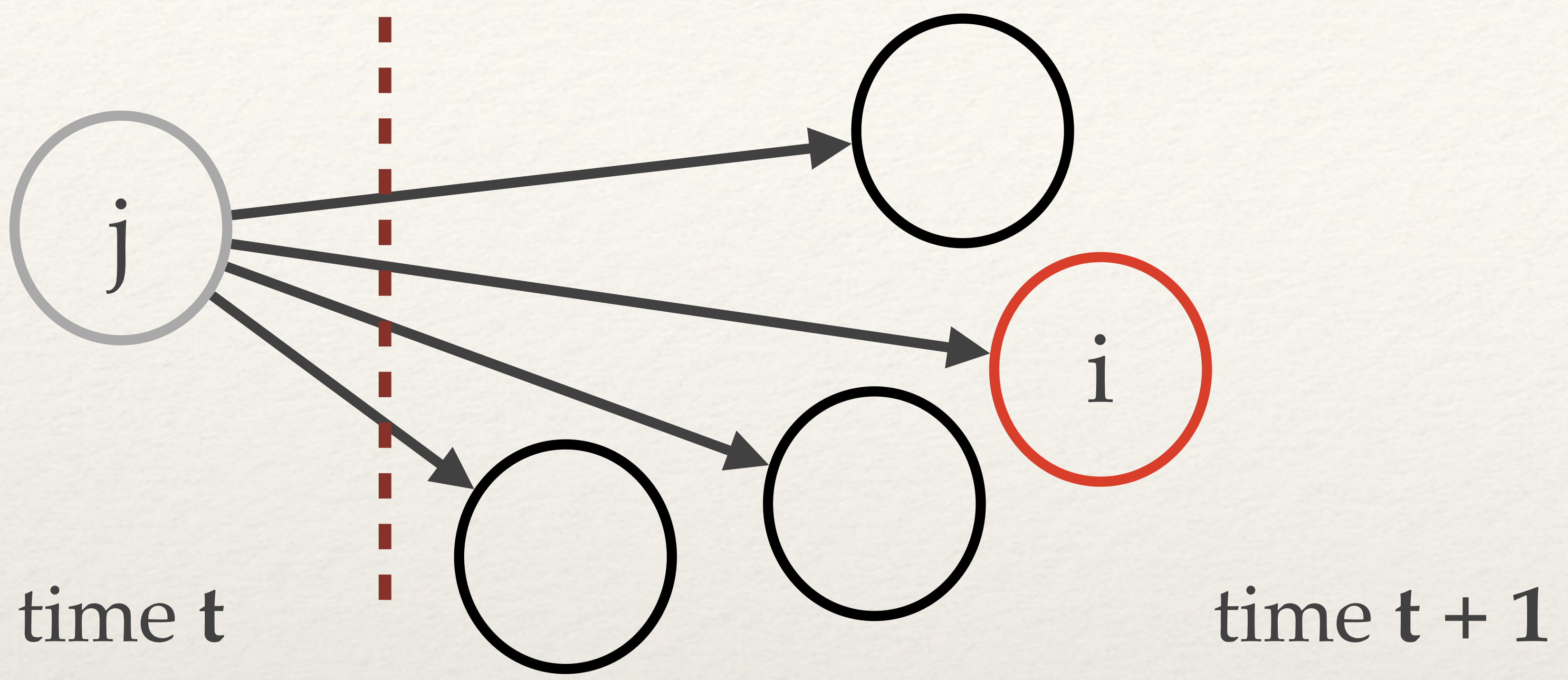
- The "leakage" of 4. and 5. has a natural interpretation: when the surfer reaches 4. or 5., then it is stuck forever
- Solution:
  - with prob. *s*, the random walker clicks on an hyperlink in the page
  - with prob. *1-s*, it jumps to a randomly selected node



Formulation of the PageRank using Random Walks



Formulation of the PageRank using Random Walks



Which is the probability of being at node *i* at time *t+1*?

$b_1, b_2, \ldots, b_n$: the probabilities of being at node *i* in a given step.

$b_i \leftarrow \sum_{j=1}^{n} M_{ji} \frac{b_j}{k_j^{out}}$: the probability of being at node *j* in the following step

Using matrix **N**: $b_i \leftarrow N_{1i}b_1 + N_{2i}b_2 + \ldots + N_{1n}b_n \quad \Rightarrow \mathbf{b} \leftarrow \mathbf{N}^\mathbf{T} \cdot \mathbf{b}$

claim: PR of page *i* is exactly the prob of being at node *i* after *k* step. (qed)$\square$

A scaled version of the random walk

For a given probability s : the walker follows a random outgoing edge

With prob *(1-s)*: the walker is teleported uniformly at random to another node

$$b_i \leftarrow s \cdot \sum_{j=1}^{n} M_{ji} \frac{b_j}{k_j^{out}} + \frac{(1-s)}{n}$$

Using matrix $\widetilde{N}$: $b_i \leftarrow \widetilde{N}_{1i} b_1 + \widetilde{N}_{2i} b_2 + \ldots + \widetilde{N}_{1n} b_n \Rightarrow \mathbf{b} \leftarrow \widetilde{\mathbf{N}}^{\mathbf{T}} \cdot \mathbf{b}$

claim: PR is equivalent to the scaled version of random walks. (qed) $\square$

Practical implications (also beyond the Web)

Modern Web search

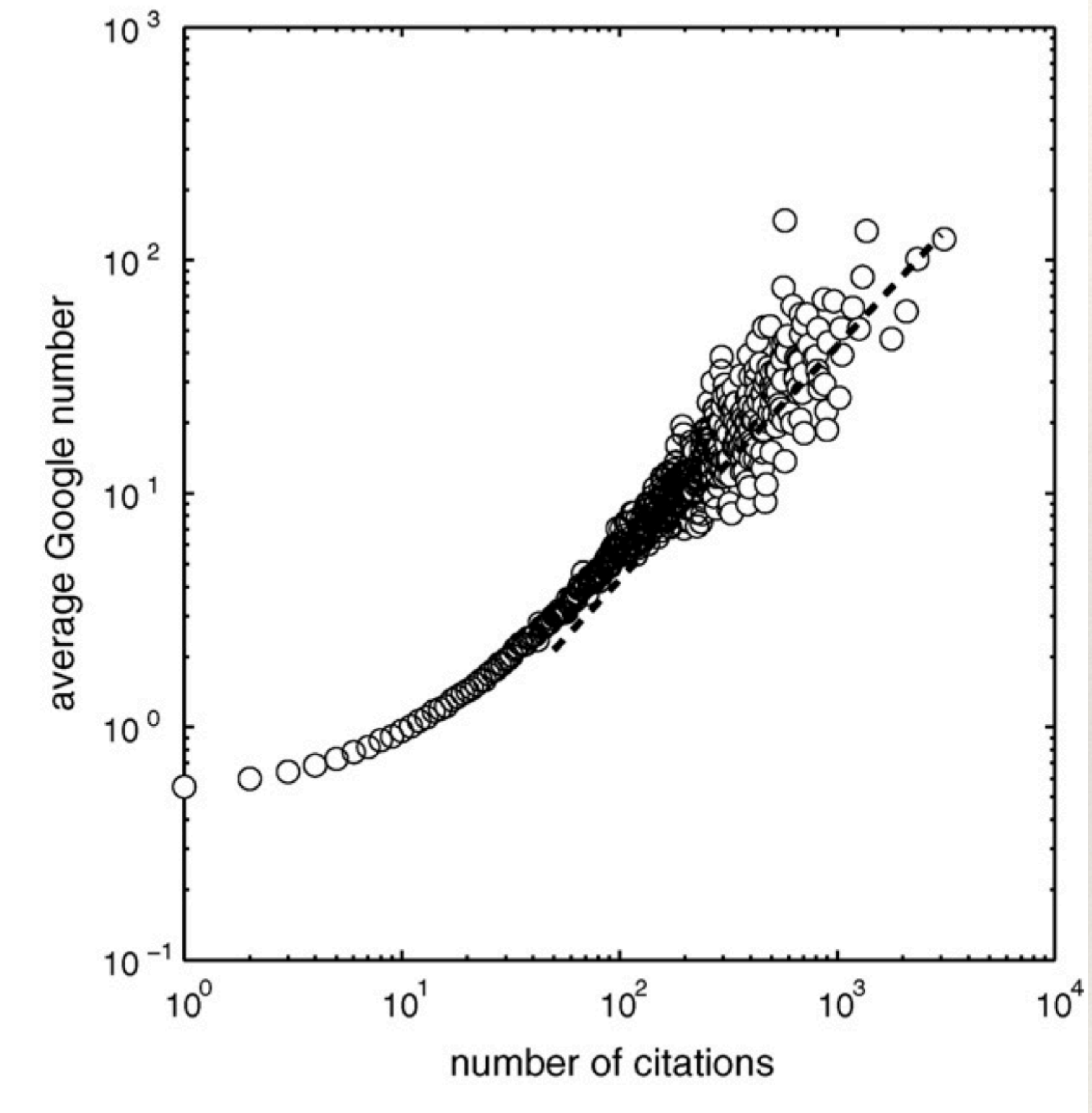
- Google today doesn't use PR anymore (original Page&Brin's paper: 2001…)
  - Hilltop (an extension of HITS) has been probably used for a while
  - anchor texts
  - clicking behavior
  - … much more (and who knows what they are actually using!)

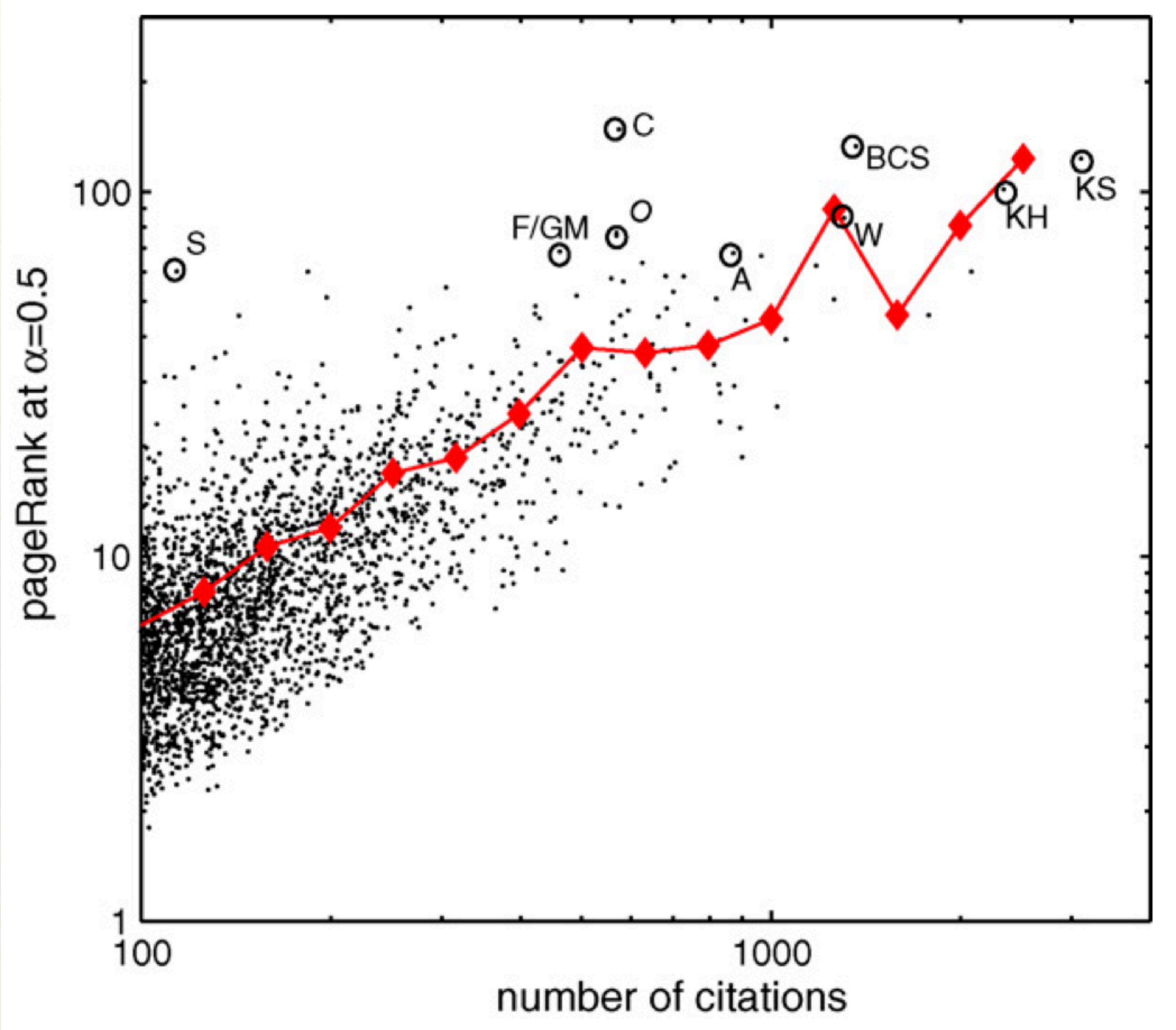
SEO vs Google

- SEO: Search Engine Optimization
- SEO companies: reverse engineering of search engine's ranking functions
- SE companies: define new measures
- … in loop!
- Feedback effects: perfect results are "moving targets"
- It is a game theoretic principle

Page Rank and citation analysis

- paper: "Finding Scientific Gems with Google Page Ranks" (2007)
- dataset: collection of scientific papers with their references
- positive correlation between number of citations and average PR values
- BUT **outliers** are papers with "limited" number of citations but highly influential anyhow





Google rank	Google # ($\times 10^{-4}$)	Cite rank	# cites	Publication				Title	Author(s)
1	4.65	54	574	PRL	10	531	1963	Unitary symmetry and leptonic...	N. Cabibbo
2	4.29	5	1364	PR	108	1175	1957	Theory of superconductivity	J. Bardeen, L. Cooper, and J. Schrieffer
3	3.81	1	3227	PR	140	A1133	1965	Self-consistent equations...	W. Kohn and L.J. Sham
4	3.17	2	2460	PR	136	B864	1964	Inhomogeneous electron gas	P. Hohenberg and W. Kohn
5	2.65	6	1306	PRL	19	1264	1967	A model of leptons	S. Weinberg
6	2.48	55	568	PR	65	117	1944	Crystal statistics I	L. Onsager
7	2.43	56	568	RMP	15	1	1943	Stochastic problems in...	S. Chandrasekhar
8	2.23	95	462	PR	109	193	1958	Theory of the Fermi interaction	R.P. Feynman and M. Gell-Mann
9	2.15	17	871	PR	109	1492	1958	Absence of diffusion in...	P.W. Anderson
10	2.13	1853	114	PR	34	1293	1929	The theory of complex spectra	J.C. Slater

Pros:
PR helps to find "gems"
in networks!

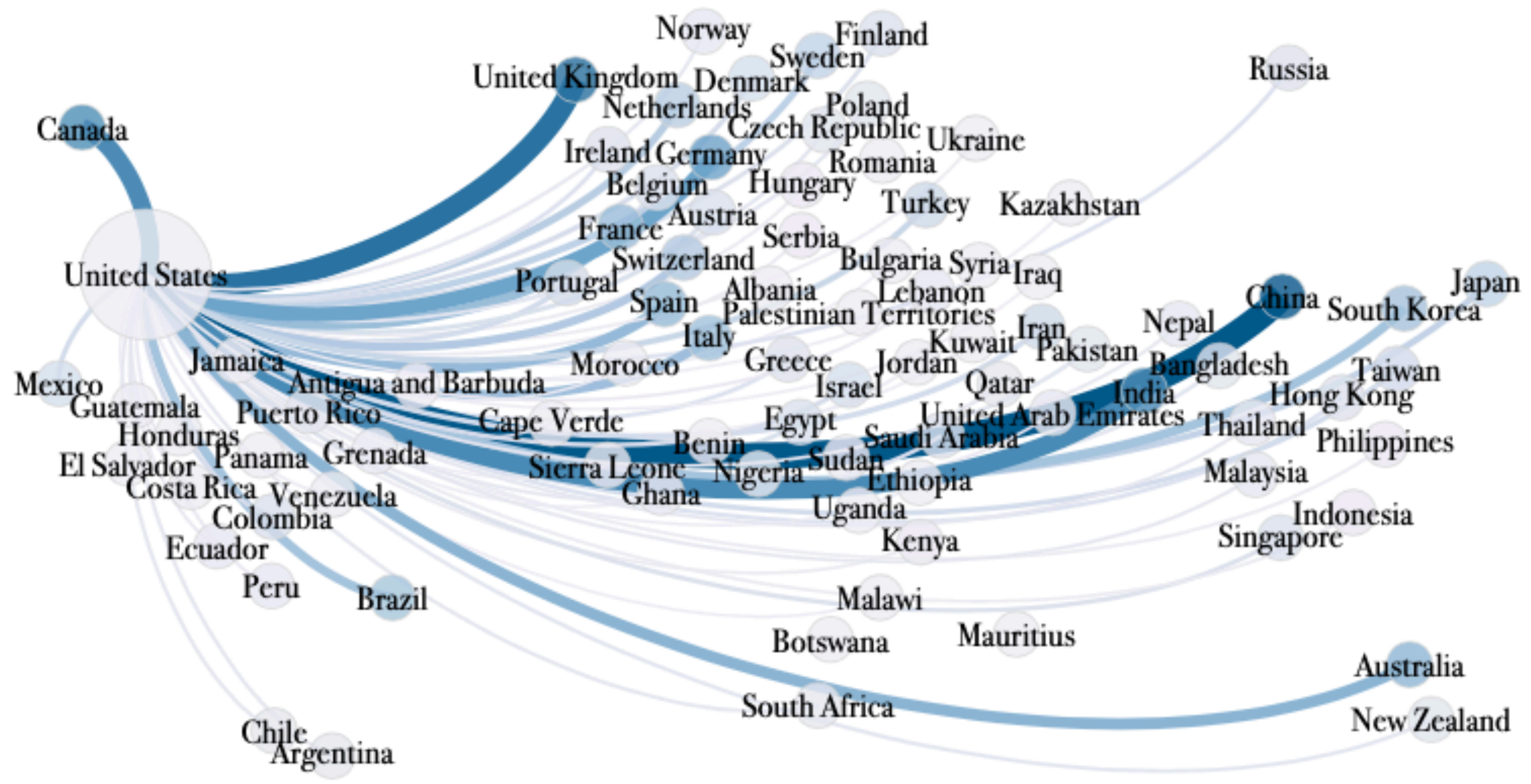
Cons:
indicators can change our behaviors…

Additional resources
research at ARC^{2}S groups

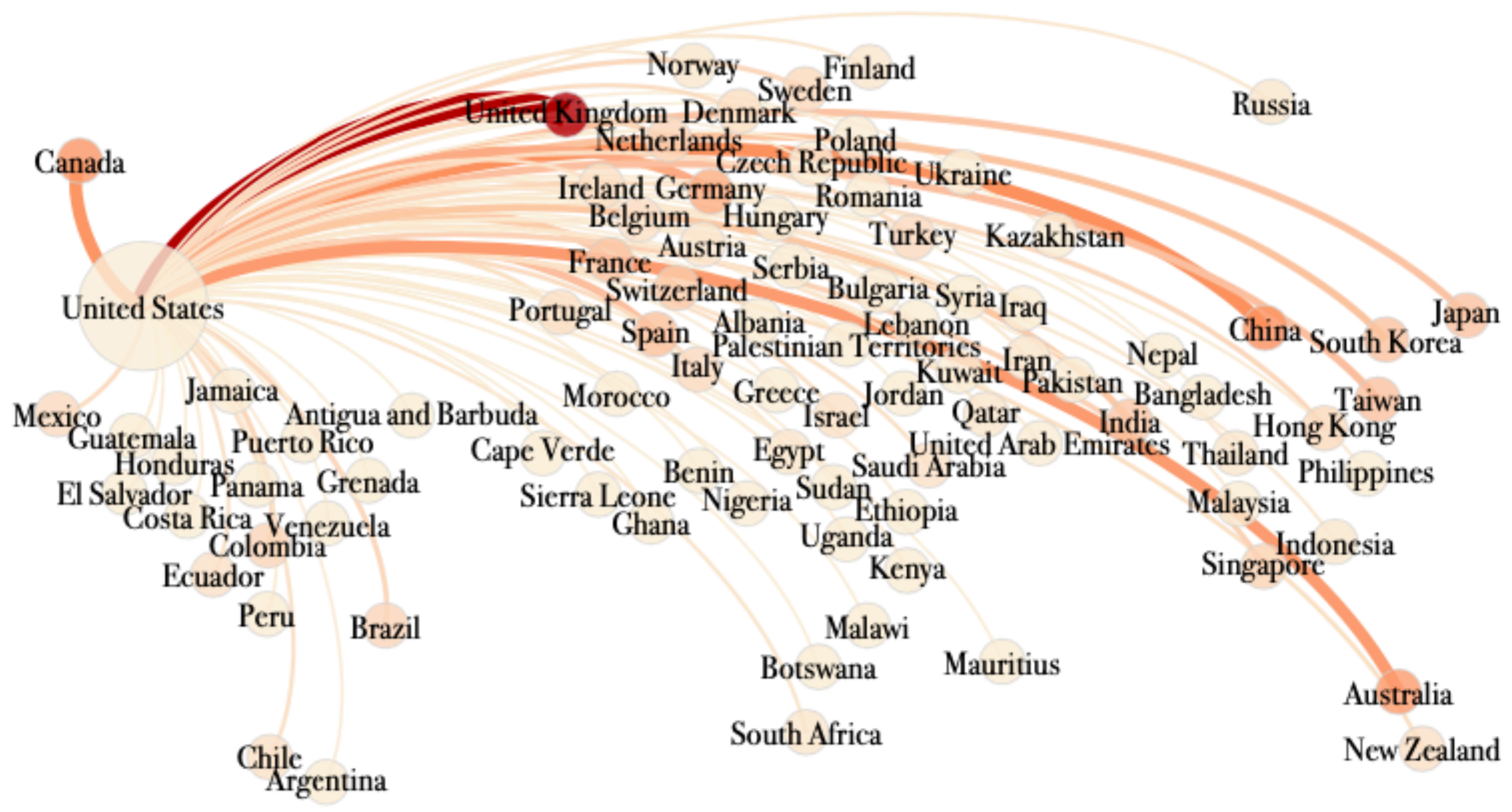
http://arcs.di.unito.it

Measuring the scientific brain drain: a country perspective

Our findings highlight the presence of a set of countries acting both as **hubs and authorities**, occupying a **privileged position** in the scientific migration network, and having similar local characteristics



(g) 2016



(h) 2016

Urbinati, A., Galimberti, E., Ruffo, G., *Measuring scientific brain drain with hubs and authorities: A dual perspective*, Online Social Networks and Media, Volume 26, 2021, 100176, https://doi.org/10.1016/j.osnem.2021.100176