AXIOMS FOR CENTRALITY

From a paper by Paolo Boldi, Sebastiano Vigna
Università degli Studi di Milano

Internet Math., 10(3-4):222-262, 2014.
Proc. 13th ICDMW. IEEE, 2013.
Proc. 4th WebScience. ACM, 2012.
Proc. 20th WWW. ACM, 2011.
More to come…

1. (Collecting) Consider the most important *centrality indices* proposed in the scientific literature(s)
2. (Assessing) Test them against *axioms* that all centrality indices should satisfy

What do these people have in common?

Ron Jeremy

Adolf Hitler

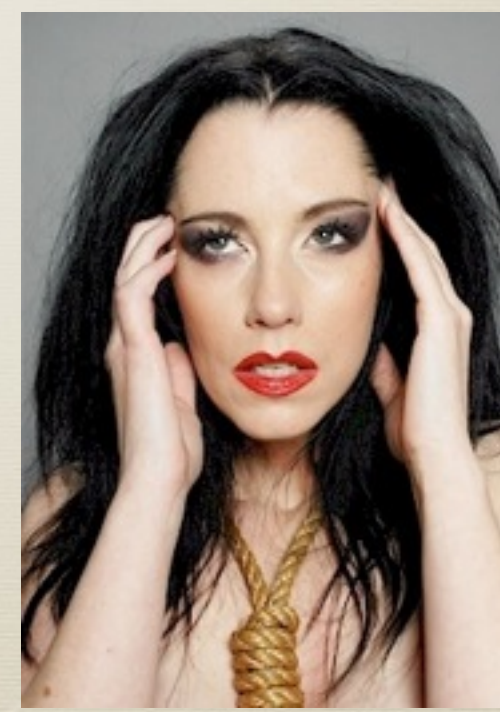
Lloyd Kaufman

George W. Bush

Ronald Reagan

Bill Clinton

Martin Sheen



Debbie Rochon

PageRank (believe it or not)

* These are the top-8 actors of the Hollywood graph according to PageRank
*

Who is going to tell 

that  is better?

And these?

William Shatner

Bess Flowers

Martin Sheen

Ronald Reagan

George Clooney

Samuel Jackson

Robin Williams

Tom Hanks

These were computed by using Harmonic Centrality instead.

Degree

* Definitely better, but who's Bess Flowers?

Bess Flowers

From Wikipedia, the free encyclopedia

Bess Flowers (November 23, 1898 – July 28, 1984) was an American actress. By some counts considered the most prolific actress in the history of Hollywood, she was known as "The Queen of the Hollywood Extras," appearing in over 700 movies in her 41 year career.

Bess Flowers

Born	November 23, 1898 Sherman, Texas, U.S.
Died	July 28, 1984 (aged 85)

Centrality in social sciences

- First works by Bavelas at MIT (1946)
- This sparked countless works (Bavelas 1951; Katz 1953; Shaw 1954; Beauchamp 1965; Mackenzie 1966; Burgess 1969; Anthonisse 1971; Czapiel 1974…) that Freeman (1979) tried to summarize concluding that:

several measures are often only vaguely related to the intuitive ideas they purport to index, and many are so complex that it is difficult or impossible to discover what, if anything, they are measuring

- Only few surveys
- Noteworthy (in the IR context): Craswell, Upstill, Hawking (ADCS 2003); Najork, Zaragoza, Taylor (SIGIR 2007); Najork, Gollapudi, Panigrahy (WSDM 2009)

Collecting

A brief survey of centrality measures

A tale of three tribes

* *Spectral indices*, based on some linear-algebra construction
* *Path-based indices,* based on the number of *paths* or shortest paths (geodesics) passing through a vertex
* *Geometric indices,* based on *distances* from a vertex to other vertices

Degree

- (In-)**Degree centrality**: the number of incoming links

$$c_{\mathrm{deg}}(x) = d^{-}(x)$$

- Or number of nodes at distance one
- Careful: when dealing with *directed* networks, some indices present two variants (e.g., in-degree vs. out-degree), the ones based on incoming paths being usually more interesting

The path tribe

Fraction of shortest paths from *y* to *z* passing through *x*

* **Betweenness** centrality (Anthonisse 1971):

$$c_{\mathrm{betw}}(x) = \sum_{y,z \neq x} \frac{\sigma_{yz}(x)}{\sigma_{yz}}$$

\# of paths of length *t* ending in *x*

* **Katz** centrality (Katz 1953):

$$c_{Katz}(x) = \sum_{t=0}^{\infty} \alpha^t \Pi_x(t) = \mathbf{1} \sum_{t=0}^{\infty} \alpha^t G^t$$

The distance tribe

- **Closeness** centrality (Bavelas 1946):

$$c_{\text{clos}}(x) = \frac{1}{\sum_y d(y,x)}$$

Distance from y to x

- The summation is over all *y* such that *d(y,x)*<∞
- **Harmonic centrality**:

$$c_{\text{harm}}(x) = \sum_{y \neq x} \frac{1}{d(y,x)}$$

- The denormalized reciprocal of the *harmonic mean of all distances* (even ∞), inspired by (Marchiori, Latora 2000)

The spectral tribe

- All based on the eigenstructure of some graph-related matrix
- Most obvious: left or right dominant eigenvector of some matrix derived from the graph adjacency matrix G. Most common: G_r, in which rows are normalized to row sum one
- All share the same issues of unicity and computability, mainly solved using Perron-Frobenius theory and the power method or more sophisticated approaches

Seeley index
(Seeley 1949)

* Basic idea: in a group of children, a **child is as popular** as the sum of the popularities of the children who like him, but popularities are divided evenly among friends:

$$c_{\mathrm{Seeley}}(x) = \sum_{y \to x} \frac{c_{\mathrm{Seeley}}(y)}{d^+(y)}$$

* In general it is a left dominant eigenvector of G_r

PageRank

(Brin, Page, Motwani, Winograd 1999)

* The idea is to start from Seeley's equation and add an adjustment to make it have a unique solution (and more)

$$c_{\mathrm{pr}}(x) = \alpha \sum_{y \to x} \frac{c_{\mathrm{pr}}(y)}{d^+(y)} + \frac{1-\alpha}{n}$$

* It is the dominant eigenvector of $\alpha G_r + (1-\alpha)\mathbf{1}^T\mathbf{1}/n$
* *Recall*: Katz (1953) is the dom. eigenv. of $\alpha G + (1-\alpha)\mathbf{e}^T\mathbf{1}/n$
* *Recall*: Seeley (1949) is the dom. eigenv. of G_r

HITS
(Kleinberg 1997)

- The idea is to start from the system:

$$c_{\mathrm{Hauth}}(x) = \sum_{y \to x} c_{\mathrm{Hhub}}(y)$$

$$c_{\mathrm{Hhub}}(x) = \sum_{x \to y} c_{\mathrm{Hauth}}(y)$$

- HITS centrality is defined to be the “authoritativeness” score

- It is a dominant eigenvector of $G^T G$, so it coincides with the dominant eigenvector on symmetric graphs

- Actually defined by Bonacich in 1991

Assessing

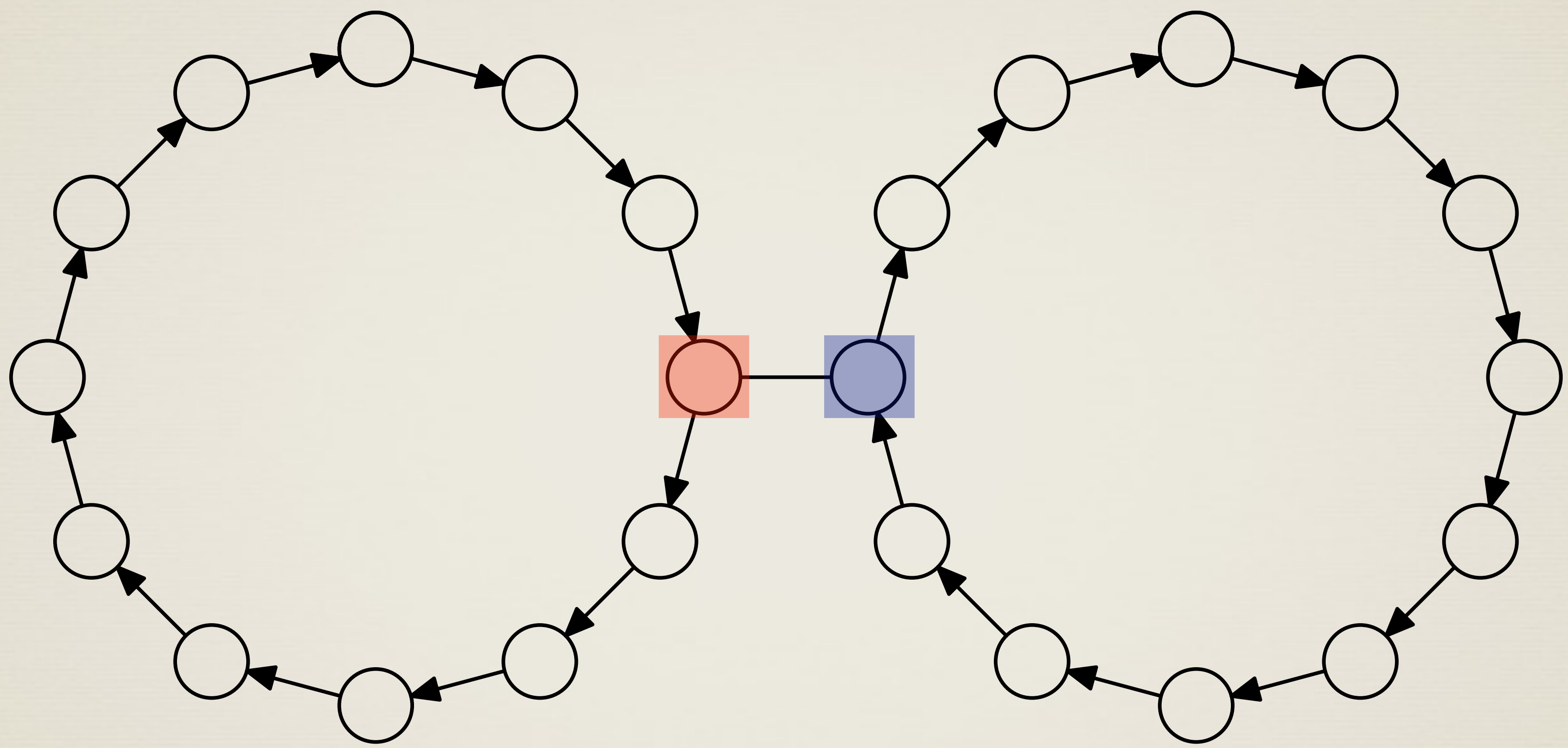
Axioms for centrality

Axioms for Centrality

* Let's try to isolate interesting properties
* Various precedents: Sabidussi [1966], Nieminen [1973], Chien *et al.* for PageRank [2004], Brandes *et al.* [2012]
* One of the axioms is indeed common (score monotonicity)

Axioms

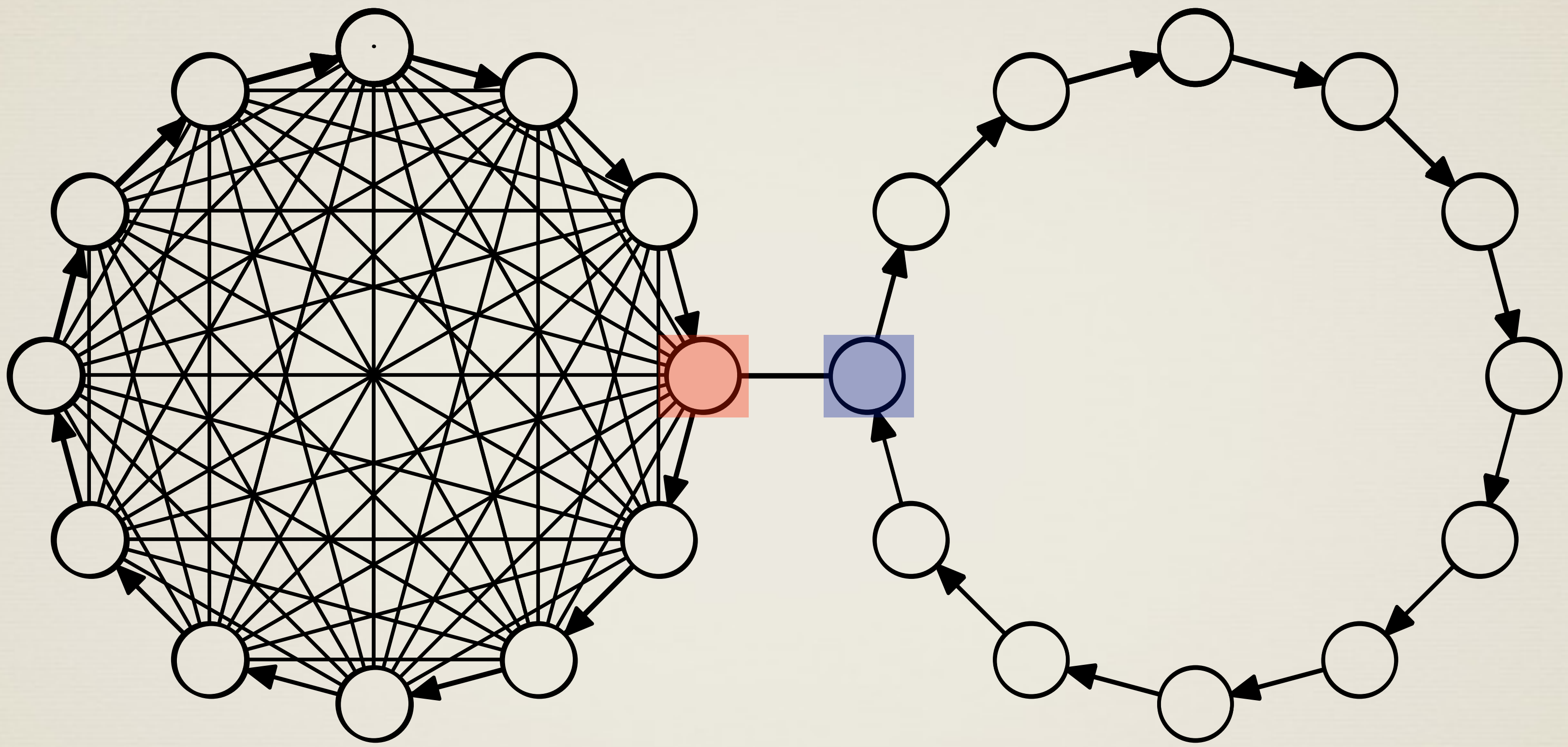
Sensitivity to density



The blue and the red node have the same importance (the two rings have the same size!)

Axioms

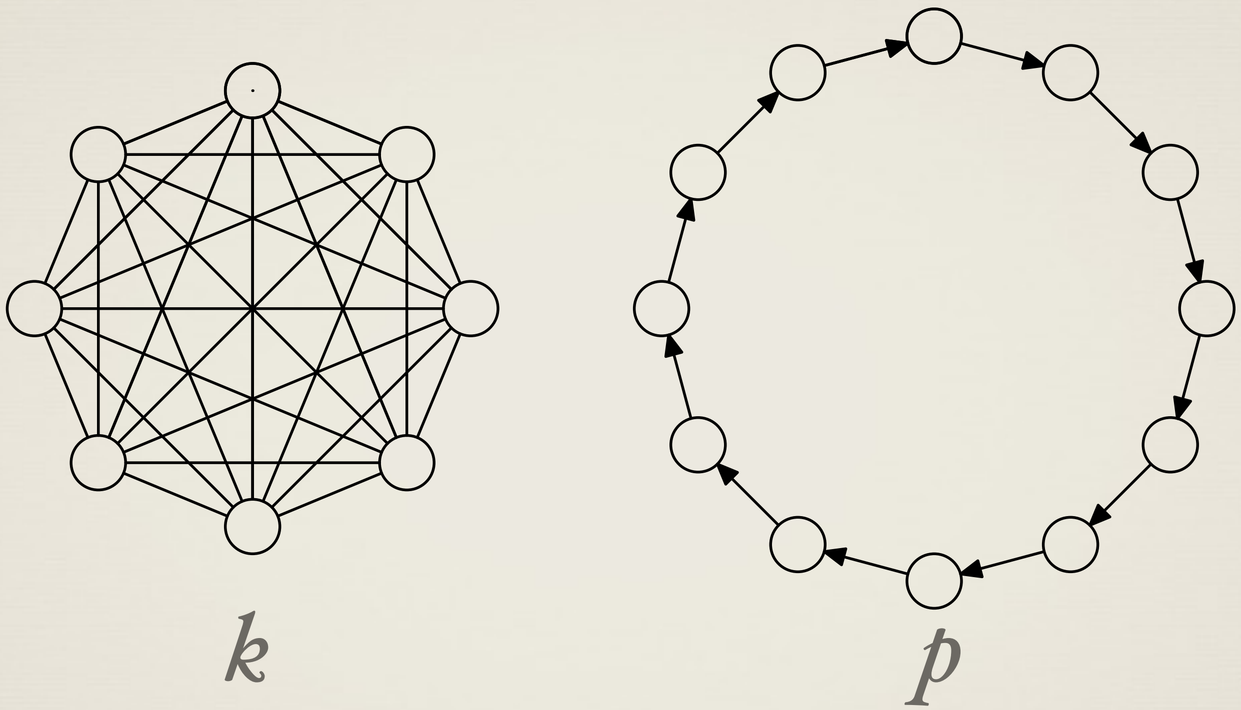
Sensitivity to density



Densifying the left-hand side, we expect the red node to become more important than the blue node

Axioms
Sensitivity to size

Two disjoint (i.e., “very far”) components of a single network



When k or p goes to ∞, the nodes of the corresponding subnetwork must become more important

Axioms
Score monotonicity

- Score monotonicity: if I add an arc towards y, the score of y must (strictly) increase (same as Sabidussi)

An axiomatic slaughter

	Density	Size	Monotone
Degree	yes	only *k*	yes
Betweenness	no (!)	only *p*	no
Dominant	yes	only *k*	no
Seeley	yes	no	no
Katz	yes	only *k*	yes
PageRank	yes	no	yes
HITS	yes	only *k*	no
Closeness	no (!)	no	no
Harmonic	**yes**	**yes**	**yes**

Conclusions

* Next time you need a centrality index… *try harmonic*!
* (You can compute it quickly using HyperBall)

Thanks!