Rank Monotonicity in Undirected Networks

Paolo Boldi, Flavio Furia, and Sebastiano Vigna, Università degli Studi di Milano, Italy

Is it good for me to get…

- A follower on Twitter?
- Formally, it is an arc towards me in a directed graph
- A friend on Facebook?
- Formally, an edge between me and someone else in an undirected graph
- To make the “good” part formal, in this talk we consider classic *geometric* and *spectral* measures of centrality
- For example: closeness centrality, eigenvector centrality, PageRank…

For Example, With (In)Degree…

- If I get a new follower or a new friend, my score increases
- Moreover, if someone has a smaller score (fewer friends or followers than me), their score remains smaller
- Even more is true, if someone else has the same score as me (i.e., number of followers of friends) they will have a lesser score afterwards
- This properties are obvious for (in)degree
- Can we show that they are still true while we switch to more “sophisticated” scores?

Score and Rank Monotonicity

- Score monotonicity: if you add an arc $x \to y$, the score of y increases (Sabidussi 1966)
- Rank monotonicity: if you add an arc $x \to y$, no vertex with a score lower than or equal y can get a score higher than y (Chien, Dwork, Kumar, Simon & Sivakumar 2004)
- Score monotonicity = something good happens
- Rank monotonicity = nothing bad happens
- *Strict* rank monotonicity: additionally, vertices with score equal to y get a score *lower* than y
- = nothing bad happens, something good happens
- Undirected case: must work at both ends (Boldi, Furia & Vigna 2021)

Geometric Centrality Measures

- Given a graph, we compute a score using some function of the distances
- Closeness centrality (Bavelas 1948): reciprocal of sum of all distances

$$\frac{1}{\sum_y d(y,x)}$$

- Harmonic centrality (Beauchamp 1965): sum of reciprocals of all distances

$$\sum_{y \neq x} \frac{1}{d(y,x)}$$

- Betweenness (Anthonisse, 1971; Freeman, 1977)

Spectral Centrality Measures

- Given a graph, we compute a score using some eigenvector associated with the adjacency matrix
- Eigenvector centrality (Landau 1895): just take the dominant eigenvector
- Motivation: If the graph is a voting graph, $\mathbf{1}A$ is majority voting (indegree)
- We can refine this: let's weight the voters using majority voting: $(\mathbf{1}A)A = \mathbf{1}A^2$
- Or refine again, but $\mathbf{1}A^k$ oscillates, so Landau proposes to find a positive $\boldsymbol{v}$ such that $\boldsymbol{v}A = \lambda\boldsymbol{v}$ (and indeed under mild hypotheses $\mathbf{1}A^k$ tends to such a $\boldsymbol{v}$)
- (Actually stated for chess tournaments using $A\mathbf{1}^T$)

Spectral Centrality Measures

- Seeley's index (1951): reputation is recursive
- Motivation: kindergarten data about child *i* liking child *j*
- Idea: global reputation should be defined recursively: $s_i = \sum_{j \to i} s_j \,/\, d_j$
- This is equivalent to $\boldsymbol{s}P = \boldsymbol{s}$, where P is A with ℓ_1-normalized rows (i.e., divide by the outdegree)
- Important idea: reputation is divided among people you endorse
- Equivalently, the steady state of the natural Markov chain on the graph (AKA "simplified PageRank"—PageRank without the damping factor)

Spectral Centrality Measures

- Katz's index (1953)
- Follows Landau's idea, but using a *summation* rather than a *limit*
- He computes $\mathbf{1} + \mathbf{1}\alpha A + \mathbf{1}\alpha^2 A^2 + \mathbf{1}\alpha^3 A^3 + \cdots = \mathbf{1}\sum_{n\geq 0} \alpha^n A^n = \mathbf{1}(\mathbf{1} - \alpha A)^{-1}$
- Note: α must be less than the inverse of the spectral radius; $\mathbf{1}A^k$ is also the number of incoming paths of length k, and that is the original formulation
- One can use a generic *border condition* $\boldsymbol{v}$ (Hubbell 1965)
- $\boldsymbol{v}(\mathbf{1} - \alpha A)^{-1}$ is equal to the dominant eigenvector of $\alpha A + (1 - \alpha)\boldsymbol{e}^T\boldsymbol{v}$, where $\boldsymbol{e}$ is a *right* dominant eigenvector of A (Vigna 2016, using Brauer 1952)

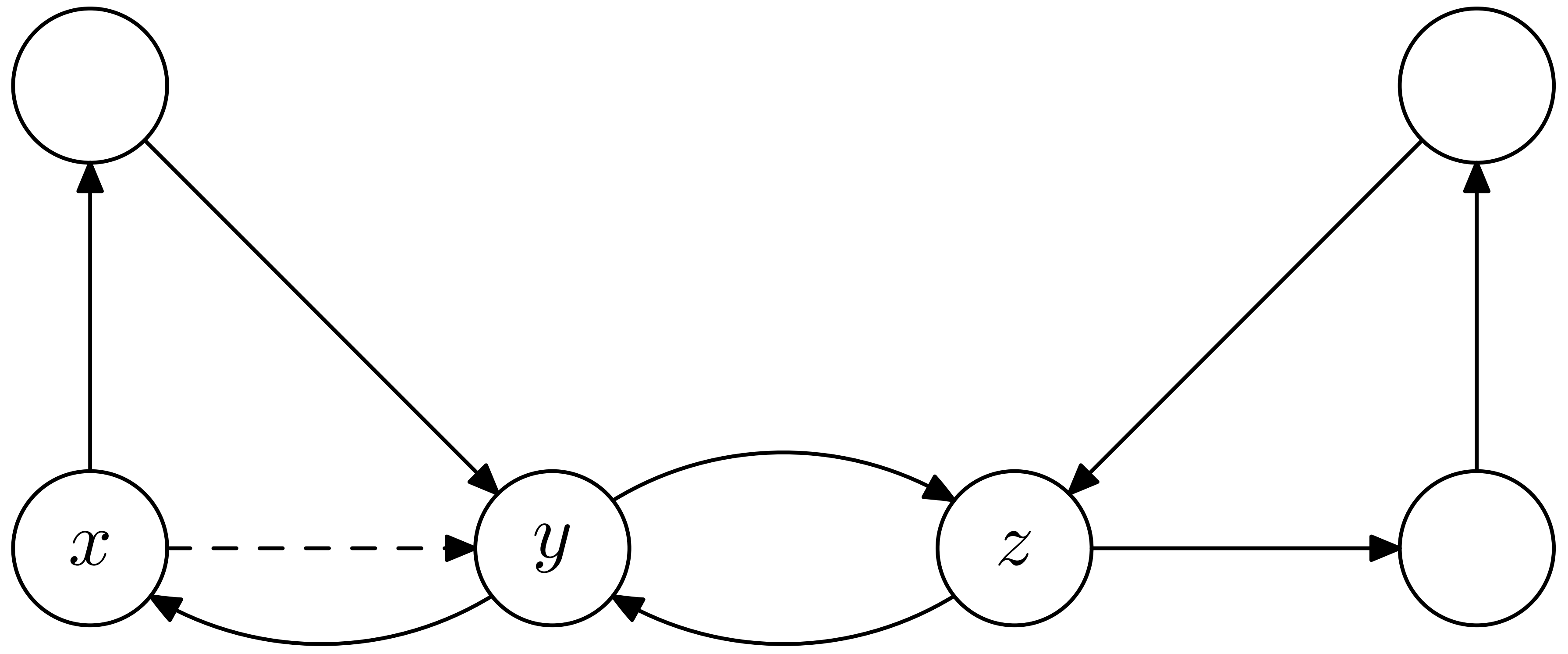
Spectral Centrality Measures

- PageRank (Page, Brin, Motwani & Winograd 1998); different formulations in time
- Steady state of a perturbed Markov chain: $\boldsymbol{p} = \boldsymbol{p}(\alpha P + (1 - \alpha)\mathbf{1}^T\boldsymbol{v})$
- Modulo a factor $(1 - \alpha)$, $\boldsymbol{v}(\mathbf{1} - \alpha P)^{-1} = \boldsymbol{v}\sum_{n\geq 0}\alpha^n P^n = \boldsymbol{v} + \boldsymbol{v}\alpha P + \boldsymbol{v}\alpha^2 P^2 + \boldsymbol{v}\alpha^3 P^3 + \cdots$
- So, Seeley's index is to PageRank as eigenvector centrality is to Katz's index
- In both cases, when α tends to its limiting value we go back

Getting a New Follower

The Directed Case: Geometric Centralities

- On strongly connected graphs, both score monotone and rank monotone
- Not surprisingly: we shorten a path
- More surprisingly: closeness is not strictly rank monotone



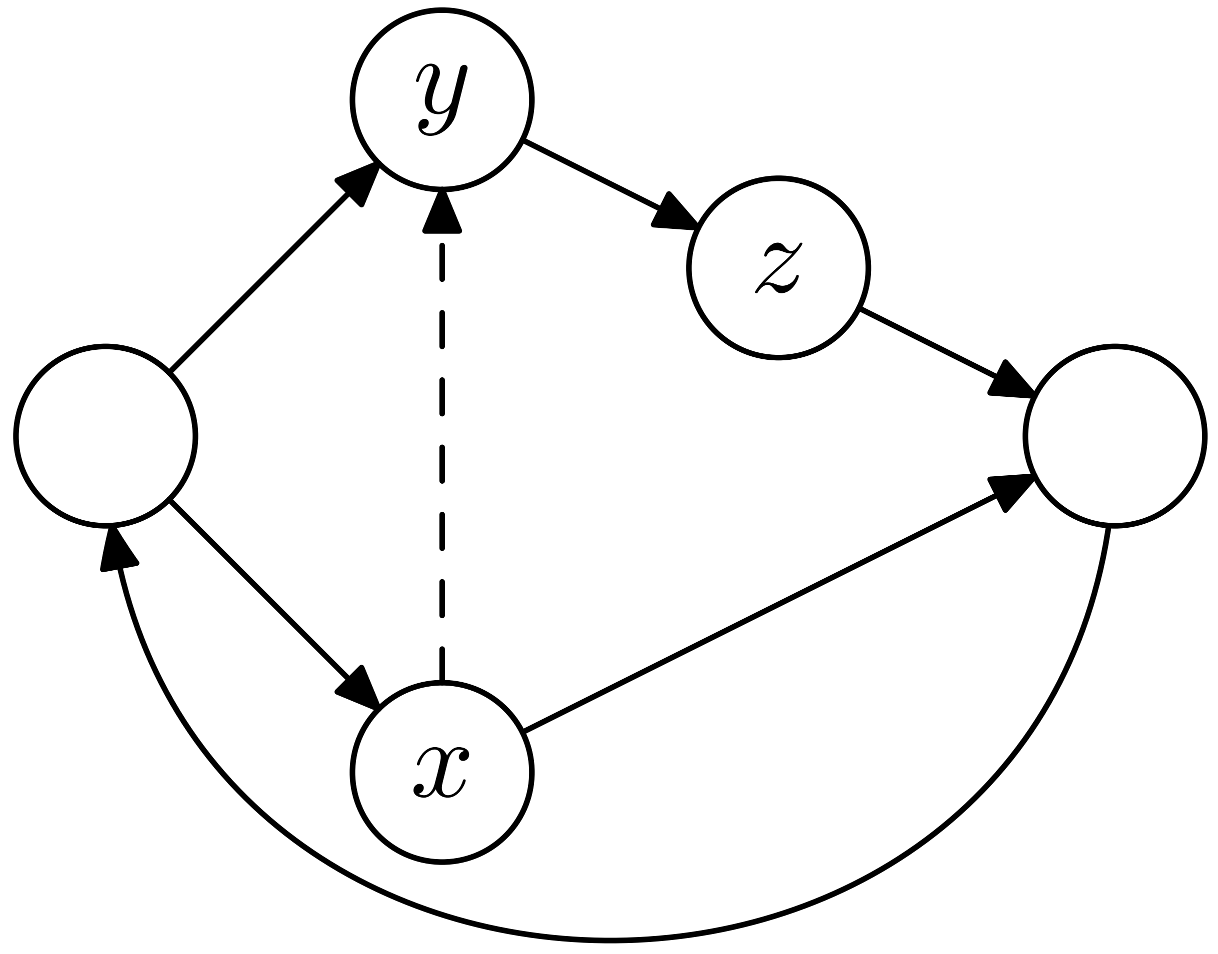
- (One more reason to ditch closeness in favor of harmonic)

The Directed Case: Eigenvector Centrality

- Eigenvector centrality is strictly rank monotone on strongly connected graphs (Boldi & Vigna 2004)
- Elsner, Johnson & Neumann 1982: any nonnegative (and overall nonzero) increase in row i of a nonnegative matrix results in coordinate i of the dominant eigenvector increasing more than any other coordinate
- So at least on strongly connected graphs the intuition is correct

The Directed Case: Seeley's Index

- Score and rank monotone only on strongly connected graphs
- Both results proved only for *regular* Markov chains in Chien, Dwork, Kumar, Simon & Sivakumar 2004
- First result does not need aperiodicity; second result can be proved using our results for PageRank and taking the limit
- Not strictly rank monotone: *y* and *z* maintain the same score



The Directed Case: PageRank

- Score monotonicity and loose rank monotonicity in Chien, Dwork, Kumar, Simon & Sivakumar 2004 under regularity assumptions
- We remove all hypotheses assuming just that the preference vector is positive, and prove *strict* rank monotonicity
- Or, loose rank monotonicity under the only hypothesis that x has nonzero score
- Our proofs cover a class of spectral centrality measures of the form $\boldsymbol{v}\sum_{n\geq 0}\alpha^n M^n$
- Same proof for PageRank and Katz's index
- Strategy: use the Sherman–Morrison formula to move the perturbation in the preference vector, and then argue using properties of M-matrices

The Directed Case: PageRank

- Let P and P' be the normalized adjacency matrices before and after adding an arc $x \to y$
- Then $P' - P = \boldsymbol{\chi}^{T}{}_{x}\,\boldsymbol{\delta}$, where contains $-1/d(d+1)$ in positions of successors of x and $1/(d+1)$ in position y
- Then the Sherman–Morrison formula expresses $(1 - \alpha P')^{-1}$ as an (ugly) correction to $(1 - \alpha P)^{-1}$, but we can gather the ugly stuff in a positive constant
- Then, $\boldsymbol{v}(1 - \alpha P')^{-1} = (\boldsymbol{v} + c\boldsymbol{\delta})(1 - \alpha P)^{-1}$

The Directed Case: PageRank

- But then we can rewrite the difference between the PageRank vectors as $\boldsymbol{v}(1 - \alpha P')^{-1} - \boldsymbol{v}(1 - \alpha P)^{-1} = (\boldsymbol{v} + c\boldsymbol{\delta})(1 - \alpha P)^{-1} - (1 - \alpha P)^{-1} = c\boldsymbol{\delta}(1 - \alpha P)^{-1}$

- McDonald, Neumann, Schneider & Tsatsomeros proved in 1995 that diagonal elements of the inverse of an M-matrix dominate strictly the off-diagonal elements, so if $B = (1 - \alpha P)^{-1}$:

$$\left[\delta\left(1 - \alpha P\right)^{-1}\right]_y = \frac{1}{d+1} b_{yy} - \sum_{x \to z} \frac{1}{d(d+1)} b_{zy} > \frac{1}{d+1} b_{yy} - \sum_{x \to z} \frac{1}{d(d+1)} b_{yy} = 0$$

The Directed Case: PageRank

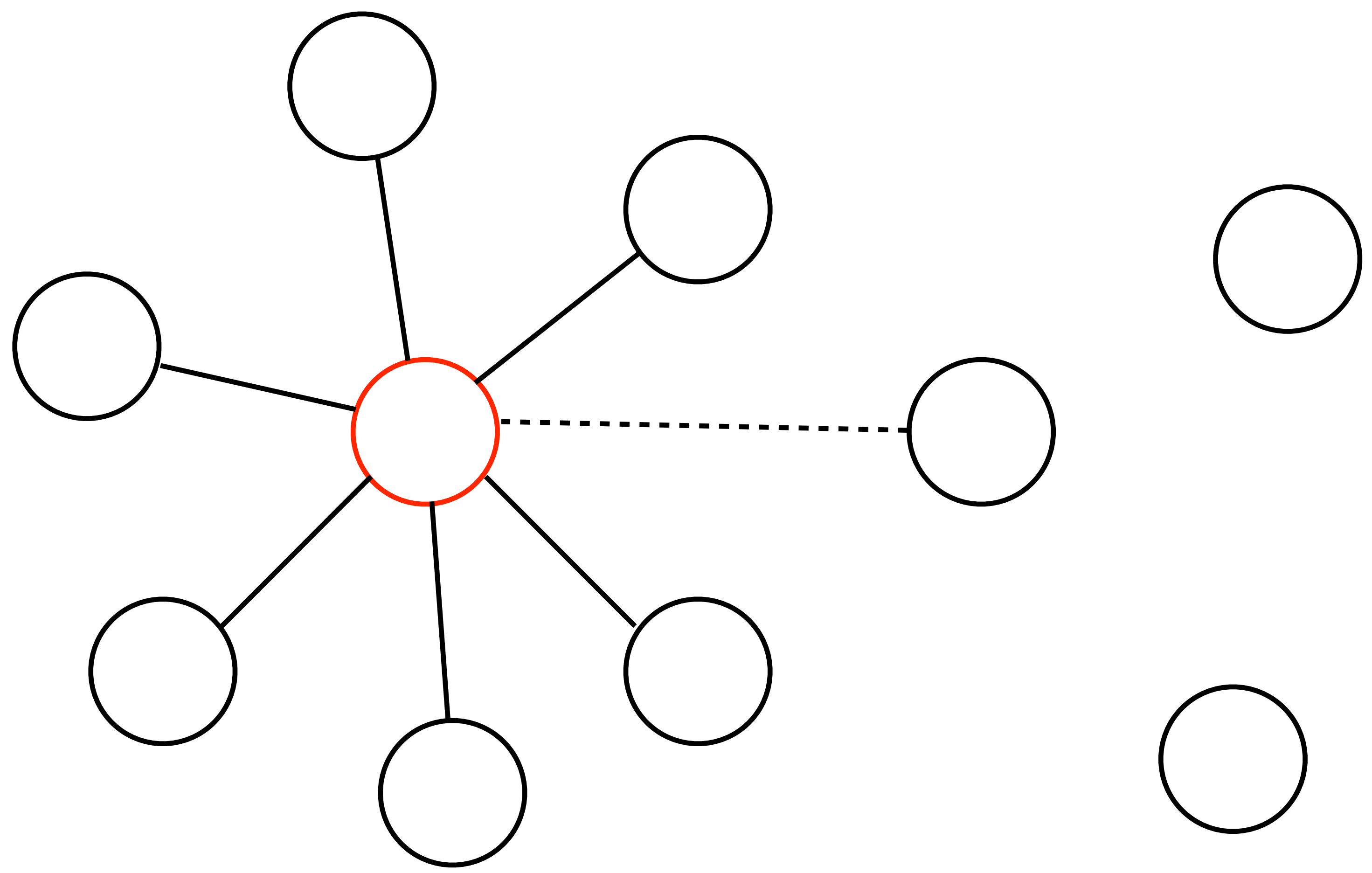
- A similar approach can be used to prove strict rank monotonicity
- There is no such proof in sight using Markov chains
- In general, properties that are not true on all Markov chains will not be within reach of proof techniques based on Markov chains
- Also, most strong results require regularity
- Linear algebra makes it possible to prove stronger statements using just irreducibility (strong connection)

Getting a New Friend

Seeley's: Score and Strictly Rank Monotone

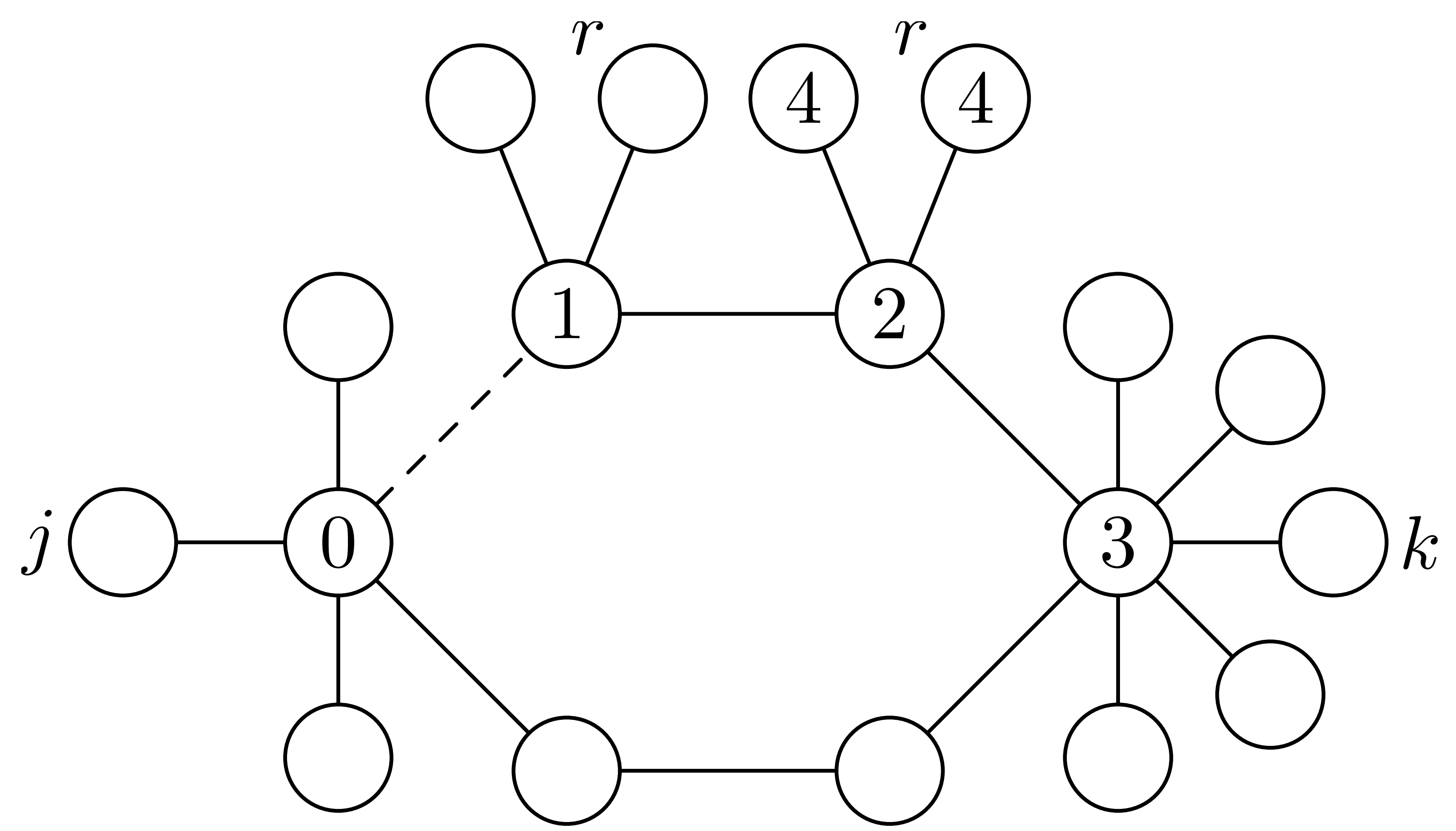
- Standardized Seeley's (ℓ_1-normalized degree)
- Obviously strictly rank monotone
- Score monotone except in the case of a star plus isolated nodes

$$\frac{d(x)+1}{2m+2} > \frac{d(x)}{2m} \implies d(x) < m$$



Nothing Else Works

- Geometric centralities



- Idea: connecting 0 and 1 benefits 1 a lot, because it gets closer to the *j* nodes around 0

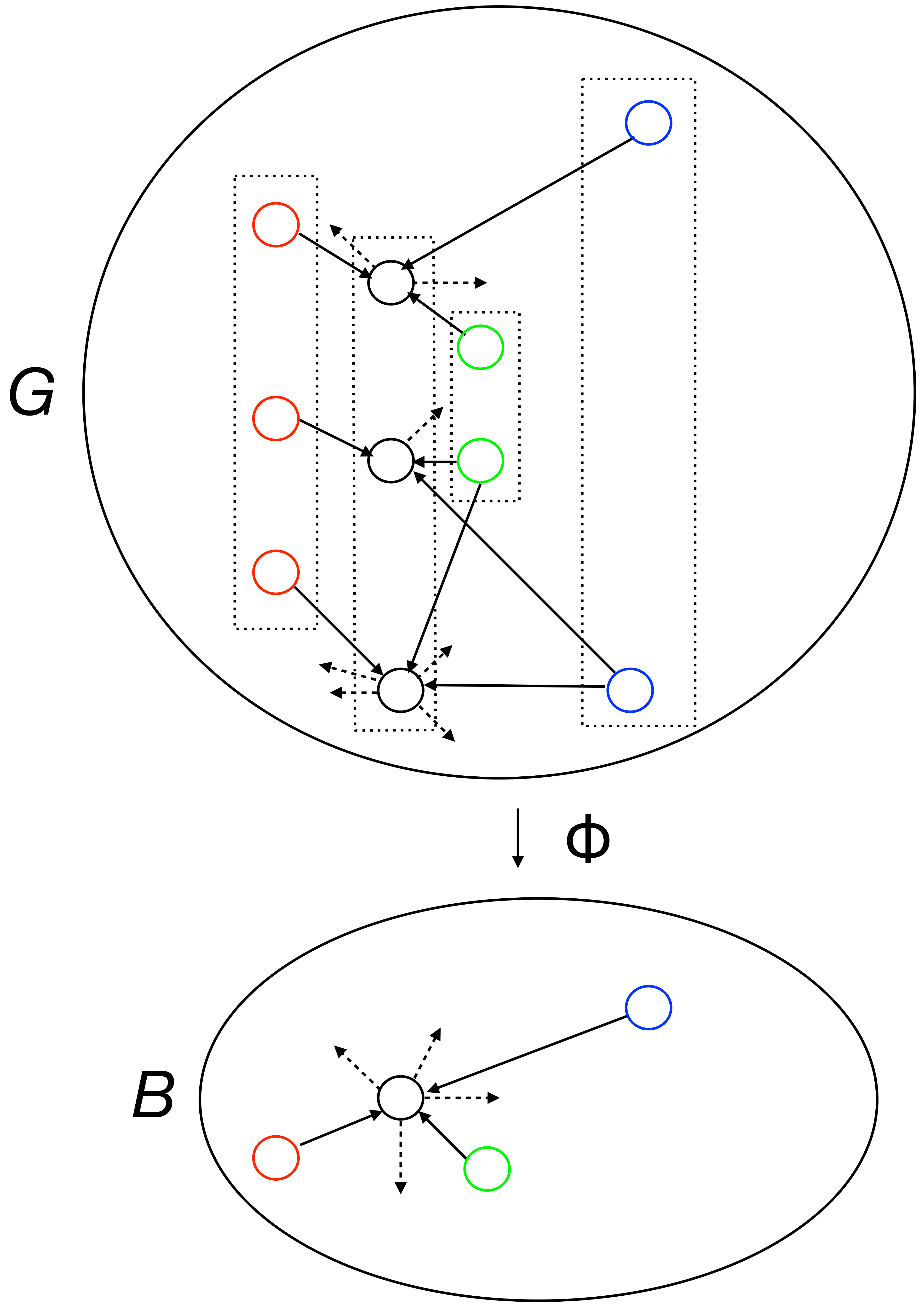
- The same doesn't happen for 0, which gains very little as *r* is small

PageRank: None of the Above, Again

- Much more complex situation
- We would like to have results for every value of the damping factor α
- For α sufficiently close to 1, adding an edge never violates rank monotonicity
- Thus, different values of α require different counterexamples
- This might suggest that in some right subinterval of the unit interval rank monotonicity works
- Surprisingly, this is not the case

Fibrations of Graphs

- Maps (morphisms) between graphs *G* and *B*
- Fiber: counterimage of a node in *B*
- Local in-isomorphism property: in-neighbourhoods are mapped bijectively
- Appeared under a plethora of names: equitable partitions and front divisors (spectral graph theory), left-resolving map (symbolic dynamics), Weisfeiler–Lehman test (graph isomorphism problems, graph neural networks), first stage of `nauty` (graph isomorphism), etc.

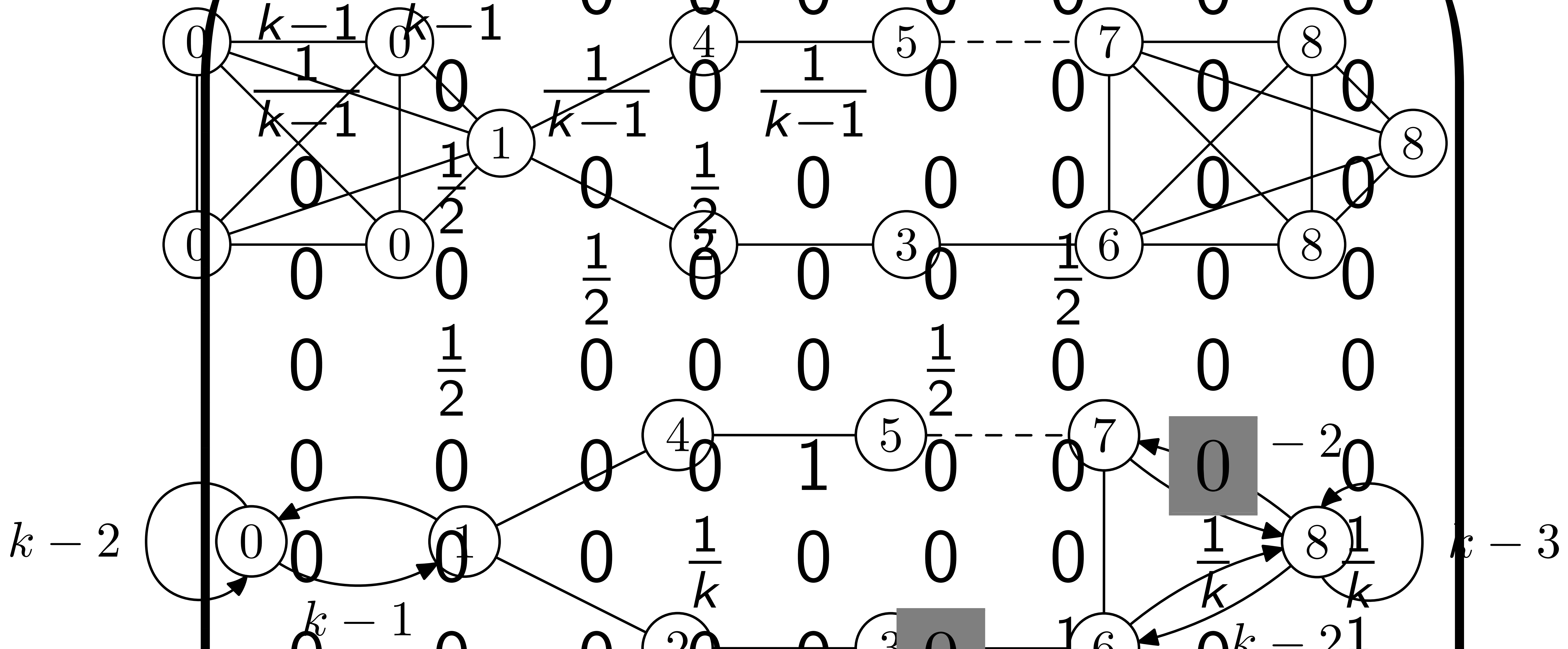


Fundamental Property

- Take a fibration $\phi: G \to B$
- If $\boldsymbol{v}$ is a vector on B, write $\boldsymbol{v}^{\Phi}$ for the *lifting of* $\boldsymbol{v}$ along ϕ, that is, the vector obtained by copying $\boldsymbol{v}$ along the fibers
- Then, $(\boldsymbol{v}B)^{\Phi} = \boldsymbol{v}^{\Phi}G$ (Sachs 1966)
- That is: if a vector is fiberwise constant, you can multiply by G by multiplying the fiber values by B and lifting
- Consequence: if G is strongly connected, the dominant eigenvector of G is the lifting of that of B, and the characteristic polynomial of B divides that of G (see the theory of *equitable partitions* and *graph divisors* in the '60s)
- More importantly, $\left(\boldsymbol{v}\sum_{n\geq 0} \alpha^n B^n\right)^{\Phi} = \boldsymbol{v}^{\Phi}\sum_{n\geq 0} \alpha^n G^n$ as lifting is linear

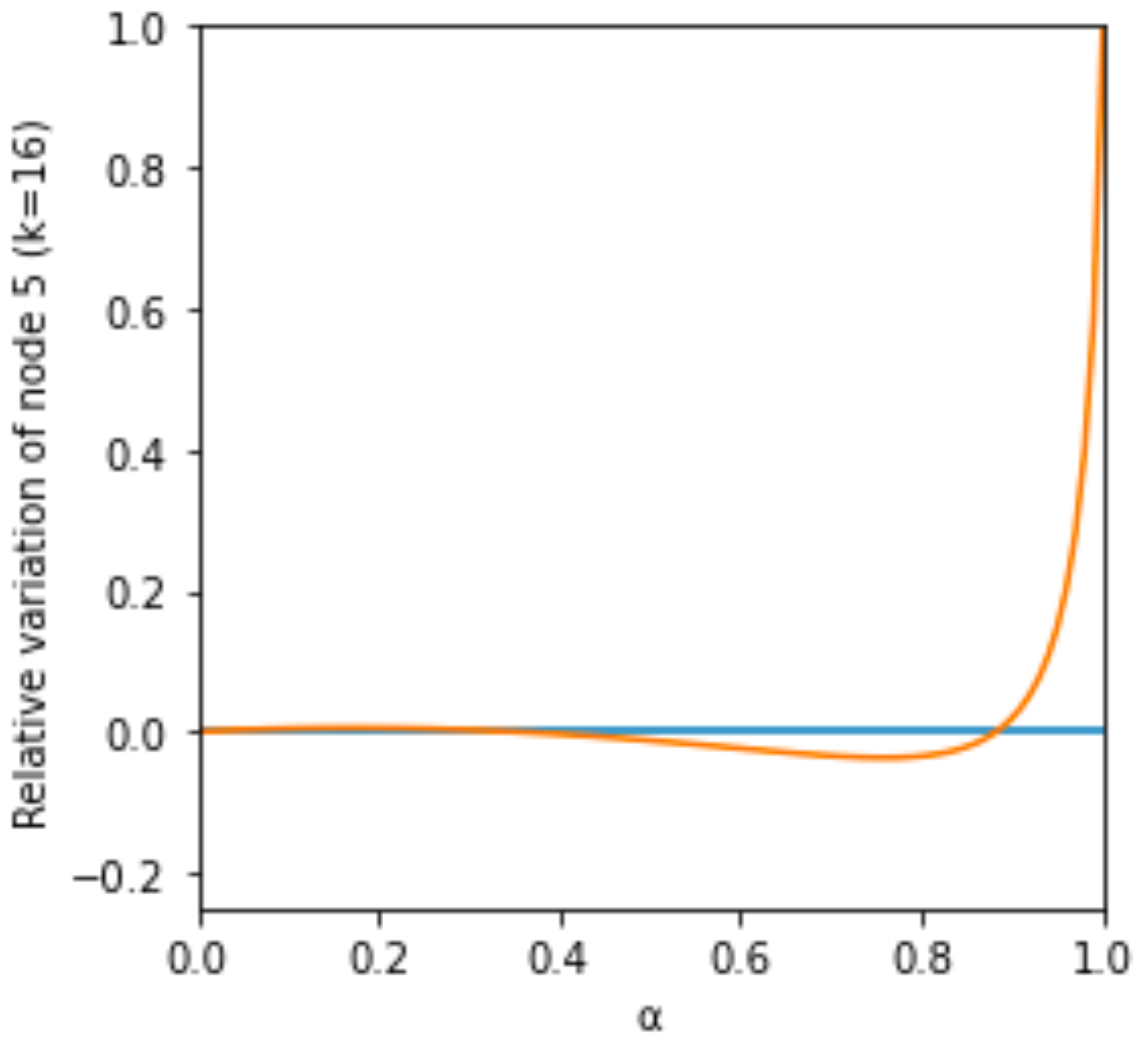
Idea: Parametric Counterexample

- We just need to find one that is fibered over a graph with a finite number of vertices



$$\begin{pmatrix}
\frac{k-2}{k-1} & \frac{k-1}{k-1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\frac{1}{k-1} & 0 & \frac{1}{k-1} & 0 & \frac{1}{k-1} & 0 & 0 & 0 & 0 \\
0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} & 0 & 0 \\
0 & \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & \boxed{0} & 0 \\
0 & 0 & 0 & \frac{1}{k} & 0 & 0 & 0 & \frac{1}{k} & \frac{1}{k} \\
0 & 0 & 0 & 0 & 0 & \boxed{0} & \frac{1}{k-1} & 0 & \frac{1}{k-1} \\
0 & 0 & 0 & 0 & 0 & 0 & \frac{k-2}{k-1} & \frac{k-2}{k-1} & \frac{k-3}{k-1}
\end{pmatrix}$$

Exact PageRank scores



- Now PageRank scores are rational functions in α whose coefficients are rational functions in *k*

- Degree is too high for exact solutions, but we can use Sturm polynomials

$$\frac{\left(\frac{\frac{1}{4}k^2-\frac{1}{2}k+\frac{1}{4}}{k^2+\frac{1}{2}k-3}\right)\alpha^7+\left(\frac{\frac{1}{4}k^3+\frac{1}{4}k^2-2k+\frac{3}{2}}{k^2+\frac{1}{2}k-3}\right)\alpha^6+\left(\frac{\frac{1}{4}k^4+\frac{1}{4}k^3-\frac{17}{4}k^2+5k-\frac{5}{4}}{k^2+\frac{1}{2}k-3}\right)\alpha^5+\left(\frac{-\frac{1}{4}k^4-3k^3+\frac{17}{4}k^2+\frac{11}{2}k-\frac{13}{2}}{k^2+\frac{1}{2}k-3}\right)\alpha^4+\left(\frac{-\frac{3}{2}k^4+2k^3+\frac{21}{2}k^2-16k+5}{k^2+\frac{1}{2}k-3}\right)\alpha^3+\left(\frac{\frac{3}{2}k^4+\frac{9}{2}k^3-\frac{27}{2}k^2+\frac{7}{2}k+4}{k^2+\frac{1}{2}k-3}\right)\alpha^2+\left(\frac{2k^4-6k^3+6k-2}{k^2+\frac{1}{2}k-3}\right)\alpha+\frac{-2k^4+2k^3+2k^2-2k}{k^2+\frac{1}{2}k-3}}{\alpha^7+\left(\frac{k^3+k^2-9k+\frac{13}{2}}{k-\frac{3}{2}}\right)\alpha^6+\left(\frac{\frac{1}{2}k^4+k^3-\frac{21}{2}k^2+3k+12}{k-\frac{3}{2}}\right)\alpha^5+\left(\frac{-\frac{1}{2}k^4-8k^3+\frac{17}{2}k^2+24k-18}{k-\frac{3}{2}}\right)\alpha^4+\left(\frac{-3k^4+3k^3+29k^2-27k-10}{k-\frac{3}{2}}\right)\alpha^3+\left(\frac{3k^4+11k^3-29k^2+k+6}{k-\frac{3}{2}}\right)\alpha^2+\left(\frac{4k^4-12k^3-4k^2+12k}{k-\frac{3}{2}}\right)\alpha+\frac{-4k^4+4k^3+4k^2-4k}{k-\frac{3}{2}}}$$

- We show that for α in the middle of [0..1) score monotonicity fails for *k*→∞

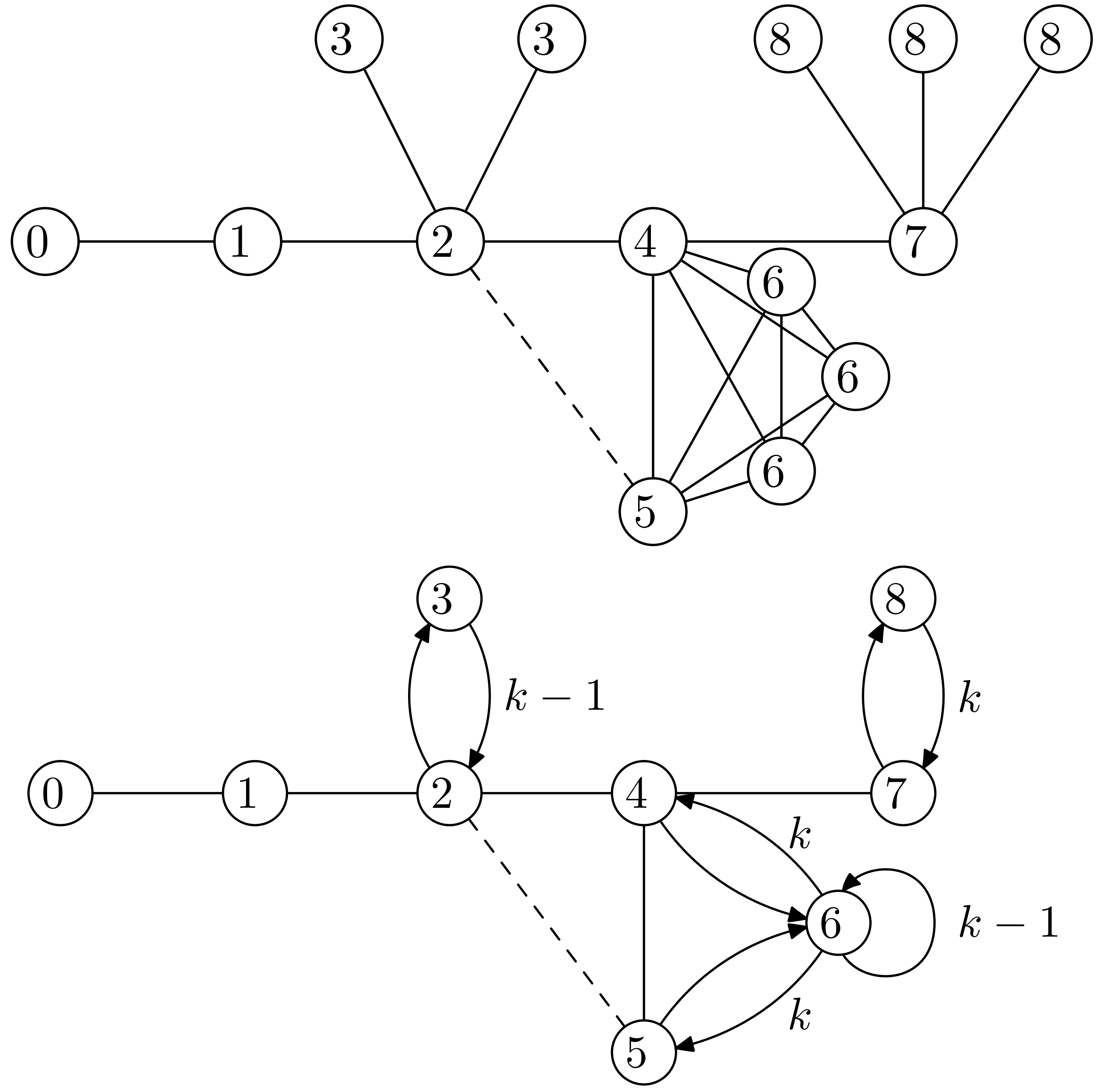
- Then we use Sturm polynomials to show that there are two sign changes around the middle

- We sandwich the sign changes between points going to 0 and 1 for *k*→∞

- We get a counterexample for each value of α

Rich Get Richer

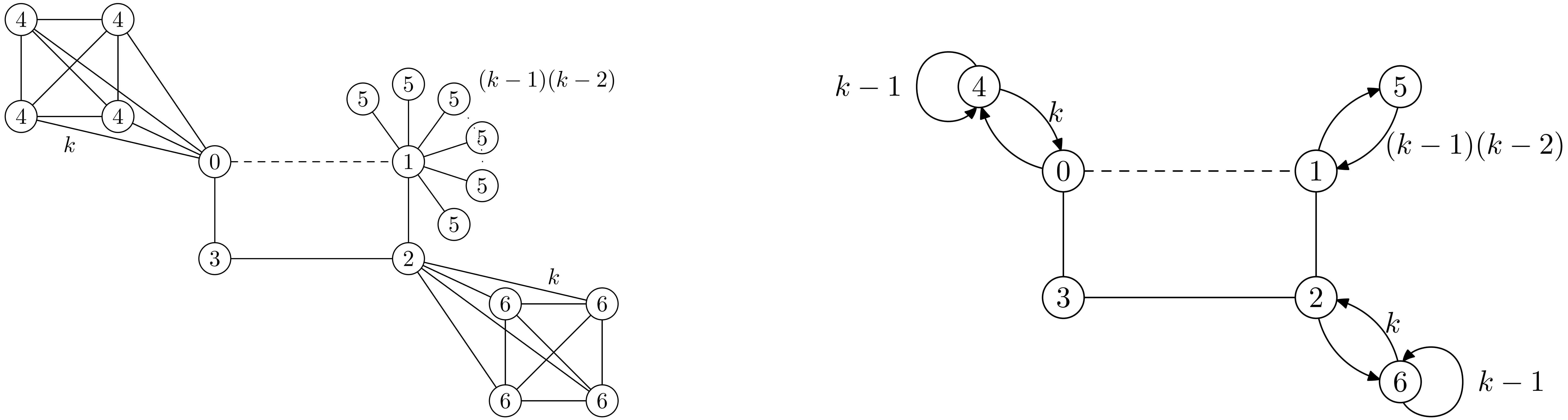
- In the previous example, the important node becomes more important ("vampire" behavior)
- We have a similar class of counterexamples for the opposite behavior



Eigenvector and Katz Centrality

- We can use the same technique, with a different graph, to prove similar results for eigenvector centrality and Katz centrality

- Eigenvector centrality is particularly tricky because we cannot have an exact description of the dominant eigenvector (unless you have less than five nodes, or you're very lucky)



- We solve the problem by proving the results for Katz's index in an interval around 1/ρ, where ρ is the spectral radius.

It Actually Happens (IMDB)!

Score increase	Score decrease	Rank violations
Meryl Streep	Yasuhiro Tsushima	Anne–Mary Brown, Jill Corso, …
Denzel Washington	Corrie Glass	Patrice Fombelle, John Neiderhauser, …
Sharon Stone	Mary Margaret (V)	Dolores Edwards, Colette Hamilton, …
John Newcomb	Robert Kirkham	Brandon Matsui, Evis Trebicka, …

Conclusion

- The natural insight into the behavior of popular spectral centrality measures on directed graph is correct
- General proof technique for monotonicity properties of spectral measures in the directed case: Sherman–Morrison formula plus *M*-matrices
- The same insight or proof techniques does not work anymore for undirected graphs
- General proof technique for spectral counterexamples: graph fibrations
- http://vigna.di.unimi.it/fibrations/