MANCHESTER 1824
The University of Manchester

M\cr NA
Manchester Numerical Analysis

Matching Centrality Measures in Complex Networks

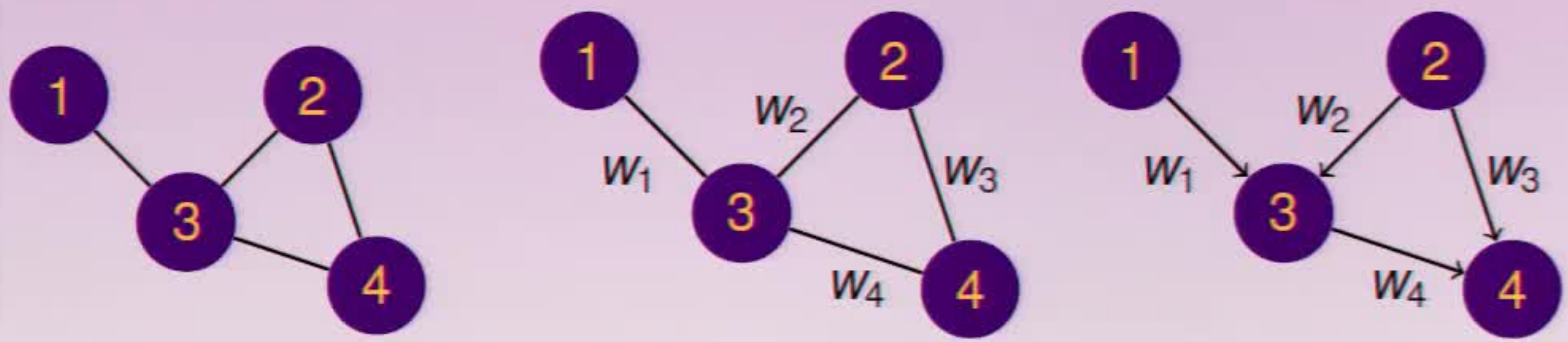
Mary Aprahamian
School of Mathematics
The University of Manchester

mary.aprahamian@manchester.ac.uk

SIAM Conference on Applied Linear Algebra 2015

Joint work with Nick Higham (U of Manchester) and Des Higham (U of Strathclyde)

Networks $\longrightarrow$ Linear Algebra



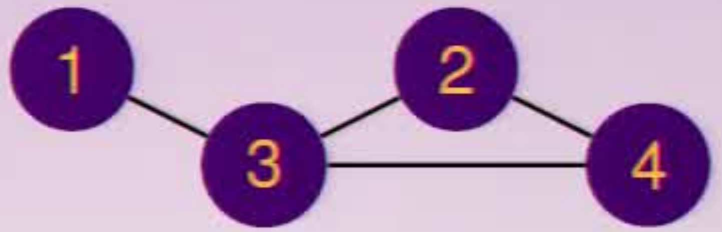
Adjacency matrices

$$\begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix} \qquad \begin{bmatrix} 0 & 0 & w_1 & 0 \\ 0 & 0 & w_2 & w_3 \\ w_1 & w_2 & 0 & w_4 \\ 0 & w_3 & w_4 & 0 \end{bmatrix} \qquad \begin{bmatrix} 0 & 0 & w_1 & 0 \\ 0 & 0 & w_2 & w_3 \\ 0 & 0 & 0 & w_4 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

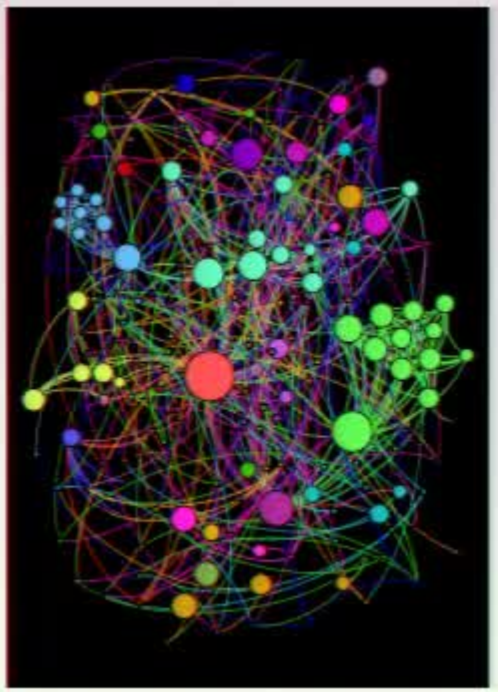
For some positive edge weights w_i.

Analysis of large networks is nontrivial

Analysis of the properties of small graphs may be trivial.



However, the analysis of large networks requires the use of advanced linear algebra.



Examples

- Human activity and relations.
- Infrastructure planning, e.g., transport networks and power grids.
- Biology, e.g., protein–protein interaction networks and disease spread.

Node ranking

Centrality

Numerical measure of the relative importance of a node in the whole network.

Importance is not unique, so there are many centrality measures, e.g., for each node

- Node degree - number of incident edges.
- Total number of closed walks.
- **Total number of open walks.**
- Betweenness.
- Eigenvector centrality.
- ...

Centrality from open walks

- $[A^k]_{ij}$ - number of open walks of length k, from node i to node j.
- $[A^k \mathbf{1}]_i = \sum_{j=1}^{n} [A^k]_{ij}$ - total number of open walks of length k, originating from node i.

$\mathbf{1}$ - vector of length n, all of whose elements are 1.

- $\sum_{k=0}^{\infty} [A^k \mathbf{1}]_i$ - total number of open walks from node i.

Problem Long and short walks are weighted equally.

Comparison

- **Exponential Centrality** $\mathbf{c}_e(A) = e^A\,\mathbf{1}$
  - \+ Very popular and successful.
  - − Computationally challenging.

  E.g., social networks, biochemical applications, anomaly detections in alarm systems, ...

- **Resolvent Centrality** $\mathbf{c}_\alpha(A) = (I - \alpha A)^{-1}\,\mathbf{1}$
  - \+ More convenient to compute.
  - − No agreed mechanism for selecting α.

  E.g., supply chain management, sports team rankings, musical influence analysis, ...

Katz parameter

Problem

Find a value of the Katz parameter α, so that the exponential and resolvent centralities give similar ordinal node rankings.

Choose the Katz parameter as the solution of

$$\min_{\alpha} \| e^{A}\mathbf{1} - (I - \alpha A)^{-1}\mathbf{1} \|_2^2,$$

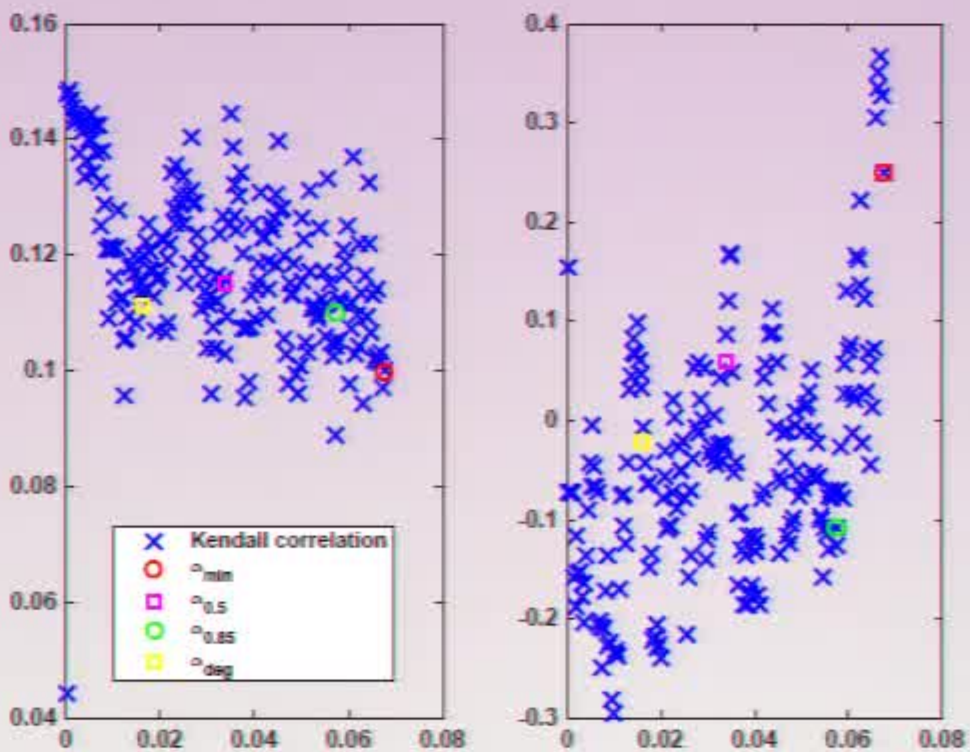
subject to $0 < \alpha < 1/\rho(A)$, where $\rho(A)$ is the spectral radius of A.

Numerical results

- Compare $\alpha_{\min} = (1 - e^{-\lambda_1})/\lambda_1$ with $\alpha_{0.5} = 0.5\dfrac{1}{\lambda_1}$, $\alpha_{0.85} = 0.85\dfrac{1}{\lambda_1}$, $\alpha_{deg} = \dfrac{1}{||A||_\infty + 1}$.
- We report the relative residual $res_{rel} = ||(e^A - (I - \alpha A)^{-1})\mathbf{1}||_2/||e^A\mathbf{1}||_2$ and Kendall's τ and Spearman's ρ correlation coefficients between the ordinal ranking obtained with $e^A\mathbf{1}$ and $(I - \alpha A)^{-1}\mathbf{1}$.
- We also report the correlation coefficients between only the top-ranked nodes.

Pajek/Erdos982

[Batagelj & Mrvar, 2006] A is symmetric, $n = 5822$, $\lambda_1 = 14.8194$.



Left: Kendall's τ correlation coefficients between the node rankings from full centrality vectors (left) and from the top 1% only (right) as a function of α.

Bottom: Kendall's τ and Spearman's ρ correlation coefficients between $e^A\mathbf{1}$ and $(I - \alpha A)^{-1}\mathbf{1}$.

Katz parameter	$\tau_{1\%}$	τ_{full}	$\rho_{1\%}$	ρ_{full}	res_{rel}
$\alpha_{min} = 0.0675$	0.2507	0.0997	0.3540	0.1346	0.0215
$\alpha_{0.5} = 0.0337$	0.0590	0.1153	0.0787	0.1545	0.9924
$\alpha_{0.85} = 0.0574$	−0.1081	0.1101	−0.1569	0.1498	0.9713
$\alpha_{deg} = 0.0161$	−0.0205	0.1115	−0.0490	0.1468	0.9954

Manchester United FC

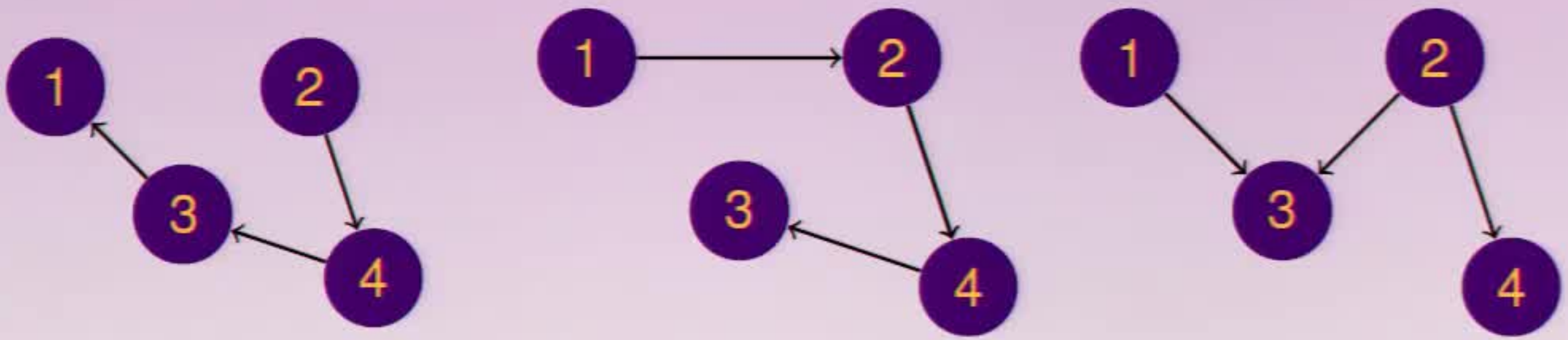
[D. J. Higham et al., 2014] A is weighted and nonsymmetric, $n = 148918$, $\lambda_1 = 41.1511$.

Top: Kendall's τ and Spearman's ρ correlation coefficients between $e^A \mathbf{1}$ and $(I - \alpha A)^{-1} \mathbf{1}$ (broadcaster scores).
Bottom: Kendall's τ and Spearman's ρ correlation coefficients between $e^{A^T} \mathbf{1}$ and $(I - \alpha A^T)^{-1} \mathbf{1}$ (receiver scores).

Katz parameter	$\tau_{1\%}$	τ_{full}	$\rho_{1\%}$	ρ_{full}	res_{rel}
$\alpha_{min} = 0.0242$	0.8489	0.6773	0.8959	0.7558	0.9995
$\alpha_{0.5} = 0.0121$	0.0247	0.4518	0.0287	0.5419	1.0000
$\alpha_{0.85} = 0.0205$	0.0516	0.4524	0.0620	0.5423	1.0000
$\alpha_{deg} = 0.0003$	0.0125	0.4496	0.0192	0.5395	1.0000

Katz parameter	$\tau_{1\%}$	τ_{full}	$\rho_{1\%}$	ρ_{full}	res_{rel}
$\alpha_{min} = 0.0242$	0.7026	0.6828	0.7735	0.7529	0.9997
$\alpha_{0.5} = 0.0121$	0.0226	0.5342	0.0340	0.6287	1.0000
$\alpha_{0.85} = 0.0205$	0.0408	0.5593	0.0617	0.6336	1.0000
$\alpha_{deg} = 0.0001$	0.0904	0.5385	0.1367	0.6354	1.0000

Temporal networks



$A_1 := A(t_1)$ $\quad$ $A_2 := A(t_2)$ $\quad$ $A_3 := A(t_3)$

Dynamic walks

- $[A_1^{k_1} A_2^{k_2} \cdots A_N^{k_N}]_{ij}$ - number of open walks of length $k = \sum_{i=1}^{N} k_i$, from node i to node j.
- $\sum_{k=0}^{\infty} [(A_1^{k_1} A_2^{k_2} \cdots A_N^{k_N}) \mathbf{1}]_i$ - total number of open walks from node i.

N. B. Walks respect the direction of time.

Centrality with multidamping

- **Exponential-based temporal centrality** uses $\alpha_{k_i}(t_j) = 1/k_i!$.
  The centrality vector is given by

$$\boldsymbol{c_e}(A_1, \cdots, A_N) := e^{A_1} e^{A_2} \cdots e^{A_N} \mathbf{1}\,.$$

  [Estrada, 2013]

- **Resolvent-based temporal centrality with multidamping** penalizes walks of length $k = k_1 + \cdots + k_N$ by $\alpha_1^{k_1} \alpha_2^{k_2} \cdots \alpha_N^{k_N}$.
  If $\alpha_j < 1/\rho(A_j)$, then the centrality vector is given by

$$\boldsymbol{c}_\alpha(A_1, \cdots, A_N) := (I - \alpha_1 A_1)^{-1} (I - \alpha_2 A_2)^{-1} \cdots (I - \alpha_N A_N)^{-1} \mathbf{1}\,.$$

Matching temporal centralities

Problem

When do the (temporal) exponential and Katz-like centralities give similar ordinal node rankings?

- What choice for the Katz-like parameter achieves this?
- What is a reasonable aggregation of the data into adjacency matrices?

At each time interval t_j we choose α_j similarly to the static case, i.e., $\alpha_j = (1 - e^{\lambda_1(A_j)})/\lambda_1(A_j)$. We also want $\rho(A_1) \approx \rho(A_2) \cdots \approx \rho(A_N)$.

Conclusions

- A value for the Katz parameter, which tries to match exponential- and resolvent-based centrality measures and which is suitable for most practical problems, $\alpha = \frac{1-e^{-\lambda_1}}{\lambda_1}$.
- Analysis is also suitable for temporal networks.
- In the future: establish how to meaningfully aggregate temporal data in adjacency matrices.

Conclusions

- A value for the Katz parameter, which tries to match exponential- and resolvent-based centrality measures and which is suitable for most practical problems, $\alpha = \frac{1-e^{-\lambda_1}}{\lambda_1}$.
- Analysis is also suitable for temporal networks.
- In the future: establish how to meaningfully aggregate temporal data in adjacency matrices.

THANK YOU!