Communities in Networks

1. Overview
2. Methods
3. Temporal/Multiplex
4. Applications
5. Temporal Centrality

- Hub & Authority Scores
- MGP Exchange Network
- Temporal Authority
  - Redux: treat identity arc between slices differently
- SVD stabilizes strongly-coupled singular limit
- Perturbative expansion

With Sean Myers, Elizabeth Leicht, Aaron Clauset & Mason Porter
Thanks to Mitch Keller, NSF & JSMF
Hubs, Authorities, and HITS (Kleinberg)

• Let $A_{ij}$ indicate links from i to j
• A good hub points to good authorities
• A good authority is pointed to by good hubs
• Seeming circularity is an eigenvector problem: find the leading eigenvectors of $AA'$ and $A'A$
• “Hypertext-Induced Topic Search”
Mathematical genealogy and department prestige

Sean A. Myers,1 Peter J. Mucha,1 and Mason A. Porter2
1Department of Mathematics, University of North Carolina,
Chapel Hill, North Carolina 27599, USA
2Mathematical Institute, University of Oxford, OX1 3LB, UK

The Mathematics Genealogy Project (http://www.
genealogy.ams.org/) is a database of over 150,000 scholars
with advanced degrees in mathematics and related fields.
Entries include dissertation titles, adviser(s), graduation
years, degree-granting institutions, and advisees. The MGP
is popular among mathematicians, and it can be used to trace
academic lineages through luminaries like Courant, Hilbert,
and Wiener to historical predecessors such as Gauss, Euler,
and even Kant. For example, MGP data was used recently to
study the role of mentorship in protégé performance.1

We consider recent branches of this mathematical fam-
ily tree by projecting the MGP data for degrees granted since
1973 onto a network whose nodes represent academic insti-
tutions in the United States. An individual who earns a doc-
torate from institution A (during the selected period) and
later advises students at institution B is represented by a
directed edge of unit weight pointing from B to A. The total
edge weight from B to A counts the number of such advisers.

This network representation can be used to estimate the
mathematical prestige of each university using various
“centrality” scores of the corresponding node (see Fig. 1). We
represent “hub” and “authority” scores2 using node size
and color (red to blue), respectively. Institutions with high
authority scores have high-valued hubs pointing to them, and
high-valued hub nodes point to high-valued authorities. A
university with a high authority score is a strong source of
prestigious Ph.D. students and a university with a high hub
score is a strong destination. In the legend of Fig. 1, we list
the top 20 institutions in order of their authority scores.

We use a “geographically inspired” layout to balance node
locations and node overlap. Kamada-Kawai visualizations4
place the high-authority universities in the network’s center.
In Fig. 2, we compare authority scores with three rankings
of mathematics departments5-7 for the 58 universities that
appear in the top 40 of at least one of the rankings or have one
of the top-40 authority scores. As expected, higher authority
scores correlate with higher prestige (i.e., smaller rank numbers).

We thank Mitch Keller at the MGP for providing data.
This work was supported by the NSF (PJM: DMS-0645369)
and the James S. McDonnell Foundation (MAP: #220020177).

(2010).
2M. E. J. Newman, Networks: An Introduction (Oxford University Press,
nrctables.pdf.
6National Research Council 2010 (rank order of medians of the S-ranking
ranges; original release), http://graduate-school.phds.org/rankings/mathematics.
reviews.com/best-graduate-schools/top-mathematics-programs/rankings.

FIG. 1. (Color) Visualizations of a mathematics genealogy network.

FIG. 2. (Color) Rankings versus authority scores.

Ph.D. Exchange Networks

- Han Social Networks (2003): Lingua Franca’s annual Job
Tracks compilations, 1993-4 to 1999-2000, for English, History,
Mathematics, Economics, PoliSci, Psychology, Sociology.


hiring/placement capacity.
Kleinberg’s authority measures prestige ("placement capacity").


$A_{ij} = \#\text{Ph.D.s from school} \ j \text{ that later advise at school} \ i \text{ (i nominates j’s "product")}$.
\[ A_{ij} = \#\text{Ph.D.s from school j that later advise at school i (i nominates j’s “product”).} \]

Kleinberg’s authority measures prestige ("placement capacity")

---

**Mathematics Genealogy Project**

**Peter John Mucha**  
MathSciNet

- Ph.D. Princeton University 1998  
- Dissertation: *On Zero Reynolds Number Microhydrodynamics of Particulate Suspensions*
- Advisor 1: Isaac Goldhirsch  
- Advisor 2: Steven Alan Orszag

Students:  
Click [here](#) to see the students listed in chronological order.

<table>
<thead>
<tr>
<th>Name</th>
<th>School</th>
<th>Year</th>
<th>Descendants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mark Carlson</td>
<td>Georgia Institute of Technology</td>
<td>2004</td>
<td></td>
</tr>
<tr>
<td>Christel Hohenegger</td>
<td>Georgia Institute of Technology</td>
<td>2006</td>
<td></td>
</tr>
</tbody>
</table>

According to our current on-line database, Peter Mucha has 2 students and 2 descendants.  
We welcome any additional information.

If you have additional information or corrections regarding this mathematician, please use the [update form](#). To submit students of this mathematician, please use the [new data form](#).
Mathematics Genealogy Project: 1973-2010
Mathematics Genealogy Project: Absolute #s v. graduate normalization

Normalized authority alternative:

\[ A_{ij} = \text{fraction (cf. #) of granted doctorates from school j that later advise at school i} \]

[threshold out schools with few graduates]
Extend centrality to dynamic network

Motivation: Mucha et al. (2010), Modularity-based community detection (counting argument fails, identity arcs defined, null model obtained by Laplacian dynamics a la Lambiotte et al.).

1. Identity arcs connect same actor across slices
2. Strength of identity coupling controls temporal smoothing
3. What behavior do we get if we apply Hub/Authority directly?

Community Structure in Time-Dependent, Multiscale, and Multiplex Networks

Peter J. Mucha,1,2,3, Thomas Richardson,1,2 Kevin Macom,2 Mason A. Porter,4,5 Jukka-Pekka Onnela1,6,7

Network science is an interdisciplinary endeavor, with methods and applications drawn from across the natural, social, and information sciences. A prominent problem in network science is the algorithmic detection of tightly connected groups of nodes known as communities. We developed a generalized framework of network quality functions that allowed us to study the community structure of arbitrary multislice networks, which are combinations of individual networks coupled through links that connect each node in one network slice to itself in other slices. This framework allows studies of community structure in a general setting encompassing networks that evolve over time, have multiple types of links (multiplexity), and have multiple scales.

The study of graphs, or networks, has a long tradition in fields such as sociology and mathematics, and it is now ubiquitous in academic and everyday settings. An important tool in network analysis is the detection of mesoscopic structures known as communities (or cohesive groups), which are defined intuitively as groups of nodes that are more tightly connected to each other than they are to the rest of the network (1–3). One way to quantify communities is by a quality function that compares the number of intracommunity edges to what one would expect at random. Given the network adjacency matrix $A$, where the element $A_{ij}$ details a direct connection between nodes $i$ and $j$, one can construct a quality function $Q(A, S)$ for the partitioning of nodes into communities as $Q = \sum_i \sum_j (A_{ij} - P_{ij}) \delta(g_i, g_j)$, where $\delta(g_i, g_j) = 1$ if the community assignments $g_i$ and $g_j$ of nodes $i$ and $j$ are the same and 0 otherwise, and $P_{ij}$ is the expected weight of the edge between $i$ and $j$ under a specified null model. The choice of null model is a crucial consideration in studying network community structure (2). After selecting a null model appropriate to the network and application at hand, one can use a variety of computational heuristics to assign nodes to communities to optimize the quality $Q$ (2, 3). However, such null models have not been available for time-dependent networks; analyses have instead depended on ad hoc methods to piece together the structures obtained at different times (6–9) or have abandoned quality functions in favor of such alternatives as the Minimum Description Length principle (10). Although tensor decompositions (11) have been used to cluster network data with different types of connections, no quality-function method has been developed for such multiplex networks.

We developed a methodology to remove these limits, generalizing the determination of community structure via quality functions to multislice networks that are defined by coupling multiple adjacency matrices (Fig. 1). The connections encoded by the network slices are flexible; they can represent variations across time, variations across different types of connections, or even community detection of the same network at different scales. However, the usual procedure for establishing a quality function as a direct count of the intracommunity edge weight minus that expected at random fails to provide any contribution from these interlice couplings. Because they are specified by common identifications of nodes across slices, interlice couplings are either present or absent by definition, so when they do fall inside communities, their contribution in the count of intracommunity edges exactly cancels that expected at random. In contrast, by formulating a null model in terms of stability of communities under Laplacian dynamics, we have derived a principled generalization of community detection to multislice networks.

1 Carolina Center for Interdisciplinary Applied Mathematics, Department of Mathematics, University of North Carolina, Chapel Hill, NC 27599, USA. Institute for Advanced Materials, Nanoscience and Technology, University of North Carolina, Chapel Hill, NC 27599, USA. Operations Research, North Carolina State University, Raleigh, NC 27695, USA. Oxford Centre for Industrial and Applied Mathematics, Mathematical Institute, University of Oxford, Oxford OX1 3LB, UK. CABDyN Complexity Centre, University of Oxford, Oxford OX1 3PH, UK. Department of Health Care Policy, Harvard Medical School, Boston, MA 02115, USA. Harvard Kennedy School, Harvard University, Cambridge, MA 02138, USA.
2 To whom correspondence should be addressed. E-mail: mucha@unc.edu

Fig. 1. Schematic of a multislice network. Four slices $s = \{1, 2, 3, 4\}$ represented by adjacency matrices $A_{ij}$ encode intraslice connections (solid lines). Interslice connections (dashed lines) are encoded by $C_{ij}$, specifying the coupling of node $i$ to itself between slices $s$ and $s'$. For clarity, interslice couplings are shown for only two nodes and depict two different types of couplings: (i) coupling between neighboring slices, appropriate for ordered slices; and (ii) all-to-all interslice coupling, appropriate for categorical slices.

Fig. 2. Multislice community detection of the Zachary Karate Club network (22) across multiple resolutions. Colors depict community assignments of the 34 nodes (renumbered vertically to group similarly assigned nodes) in each of the 16 slices with resolution parameters $\alpha = \{0.25, 0.5, \ldots, 4\}$, for $\alpha = 0$ (top), $\alpha = 0.1$ (middle), and $\alpha = 1$ (bottom). Dashed lines bound the communities obtained using the default resolution ($\alpha = 1$).
Mathematics Genealogy Project: 1946-2010 as “multislice” network (“Take 1”)

\[
\tilde{A} = \begin{pmatrix}
 cA^{(1)} & I & 0 & 0 \\
 I & cA^{(2)} & I & 0 \\
 0 & I & cA^{(3)} & I \\
 0 & 0 & I & cA^{(4)}
\end{pmatrix}
\]

\[
M(c) = \tilde{A}' \tilde{A}
\]

\[A^{(y)}_{ij} = \# \text{ graduated in year } y \text{ from school } j \text{ that later advise at school } i\]

Calculate authority from the NT-by-NT adjacency matrix as the leading eigenvector of \(M(c)\)  

[N universities, T time slices = individual years]

Large c: Emphasis on current/recent years
Small c: Temporal smoothing of actor roles
\(c \to 0\): Singular limit as multislice network disconnects into N chains
MGP: 1946-2010 ("Take 1")
Total authority in each slice/year
MGP: 1946-2010 ("Take 1")
Node authority relative to that year’s total

 authority rela’ve to that year’s total
MGP: 1946-2010 ("Take 1")

- The strongly-coupled singular limit is $N$ identical T-node lines

$A_{\text{line}} = \begin{pmatrix} 0 & 1 & 0 & \cdots \\ 1 & 0 & 1 & \ddots \\ 0 & 1 & 0 & \ddots \\ \vdots & \ddots & \ddots & \ddots \end{pmatrix}$

$A'_{\text{line}}A_{\text{line}} = \begin{pmatrix} 1 & 0 & 1 & 0 & \cdots \\ 0 & 2 & 0 & 1 & \ddots \\ 1 & 0 & 2 & 0 & \ddots & 0 \\ 0 & 1 & 0 & 2 & \ddots & 1 & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \end{pmatrix}$
MGP: 1946-2010 (“Take 1”)

- The strongly-coupled singular limit is N identical T-node lines

\[
A_{\text{line}} = \begin{pmatrix}
0 & 1 & 0 & \cdots \\
1 & 0 & 1 & \ddots \\
0 & 1 & 0 & \ddots \\
\vdots & \ddots & \ddots & \ddots
\end{pmatrix}
\]

\[
A'_{\text{line}}A_{\text{line}} = \begin{pmatrix}
1 & 0 & 1 & 0 & \cdots \\
0 & 2 & 0 & 1 & \ddots \\
1 & 0 & 2 & 0 & \ddots & 0 \\
0 & 1 & 0 & 2 & \ddots & 1 & 0 \\
\vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\
0 & 1 & 0 & 2 & 0 & \cdots \\
0 & 1 & 0 & 1 & \cdots & \cdots
\end{pmatrix}
\]

GOOD: Authority on a line is analytical calculation
BAD: $C \rightarrow 0$ limit decouples odd and even nodes
MGP: 1946-2010 ("Take 2")
A better definition of temporal authority

- Eigenvector Centrality of $A_{line} = \begin{pmatrix}
0 & 1 & 0 & \cdots \\
1 & 0 & 1 & \ddots \\
0 & 1 & 0 & \ddots \\
\vdots & \ddots & \ddots & \ddots
\end{pmatrix}$

$$\lambda = 2 \cos \frac{\pi}{T+1}, \quad V_t \propto \sin \frac{\pi t}{T+1}$$

![Graph of authority over time](image)
MGP: 1946-2010 ("Take 2")
A better definition of temporal authority

- Eigenvector Centrality of $A_{\text{line}}$

$$\lambda = 2 \cos \frac{\pi}{T+1} \quad V_t \propto \sin \frac{\pi t}{T+1}$$

$$\tilde{M}(c) = \begin{pmatrix}
cA^{(1)'}A^{(1)} & I & 0 & 0 \\
I & cA^{(2)'}A^{(2)} & I & 0 \\
0 & I & cA^{(3)'}A^{(3)} & I \\
0 & 0 & I & cA^{(4)'}A^{(4)}
\end{pmatrix}$$
MGP: 1946-2010 ("Take 2")
A better definition of temporal authority

\[ V_t \propto \sin \frac{\pi t}{T + 1} \]
SVD stabilization of singular limit

$c \to 0$: Singular limit as multislice network disconnects into $N$ chains

Reshaping values into $N$ (schools) by $T$ (years) matrix of “prestige” $P$, expect

$P \to$ outer product of time-averaged prestige with temporal variability along chain

(analytically obtained sine waves) as $c \to 0$. **First component of SVD at small $c$.**
Perturbative solution around $c=0$

- Essential issue of the singular nature: at $c=0$, the connected temporal network of $N$ actors across $T$ times degenerates into $N$ lines of length $T$, each with the sine wave time dependence; but the authorities of chains relative to one another are only obtained in the limit as $c \to 0^+$

$$V_t \propto \sin \left(\frac{\pi t}{T + 1}\right)$$
Perturbative solution around $c=0$

\[
\lambda v = \tilde{M}(\epsilon)v = (C + \epsilon G)v
\]

\[
C = \begin{pmatrix}
0 & I & 0 & \cdots \\
I & 0 & I & \ddots \\
0 & I & 0 & \ddots \\
\vdots & \ddots & \ddots & \ddots
\end{pmatrix}
\]

\[
G = \begin{pmatrix}
A^{(1)'A^{(1)}} & 0 & \cdots \\
0 & A^{(2)'A^{(2)}} & \ddots \\
\vdots & \ddots & \ddots
\end{pmatrix}
\]

\[
\lambda = \lambda_0 + \epsilon \lambda_1 + \cdots
\]

\[
v = v^{(0)} + \epsilon v^{(1)} + \cdots
\]
Base solution around $c=0$

\[ \lambda v = \tilde{M}(\epsilon)v = (C + \epsilon G)v \]

\[ C = \begin{pmatrix}
  0 & I & 0 & \cdots \\
  I & 0 & I & \ddots \\
  0 & I & 0 & \ddots \\
  \vdots & \ddots & \ddots & \ddots \\
\end{pmatrix} \]

\[ \lambda = \lambda_0 + \epsilon \lambda_1 + \cdots \]

\[ v = v^{(0)} + \epsilon v^{(1)} + \cdots \]

\[ (\lambda_0 I - C)v^{(0)} = 0 \]

\[ \Rightarrow v_{i+}^{(0)} = a_i V_t \]
Base solution around $c=0$

$$\lambda v = \tilde{M}(\epsilon)v = (C + \epsilon G)v$$

$$C = \begin{pmatrix}
0 & I & 0 & \cdots \\
I & 0 & I & \ddots \\
0 & I & 0 & \ddots \\
\vdots & \ddots & \ddots & \ddots 
\end{pmatrix}$$

$$G = \begin{pmatrix}
A^{(1)'} & A^{(1)} & 0 & \cdots \\
0 & A^{(2)'} & A^{(2)} & \ddots \\
\vdots & \ddots & \ddots & \ddots 
\end{pmatrix}$$

$$\lambda = \lambda_0 + \epsilon \lambda_1 + \cdots$$

$$v = v^{(0)} + \epsilon v^{(1)} + \cdots$$

$$(\lambda_0 I - C)v^{(0)} = 0$$

$$\Rightarrow v^{(0)}_{it} = a_i V_t$$

Solvability condition in next order:

$$(\lambda_0 I - C)v^{(1)} = (G - \lambda_1 I)v^{(0)}$$
Base solution around $c=0$

\[
\lambda v = \tilde{M}(\epsilon) v = (C + \epsilon G) v
\]
\[
(\lambda_0 I - C)^{v^{(0)}} = 0
\]
\[
\Rightarrow v^{(0)}_{it} = a_i V_t
\]

Solvability condition in next order:

\[
(\lambda_0 I - C)^{v^{(1)}} = (G - \lambda_1 I)^{v^{(0)}}
\]
\[
L_0 \equiv (\lambda_0 I - C)
\]
\[
L_1 \equiv (G - \lambda_1 I)
\]

\[
L_0 v^{(1)} = L_1 v^{(0)}
\]
\[
\forall i : V \cdot [L_1 v^{(0)}]_i = 0
\]
\[
\left[ \sum_t \sin^2 \left( \frac{\pi t}{T+1} \right) G_{ij} \right] a_j = \lambda_1 \left[ \sum_t \sin^2 \left( \frac{\pi t}{T+1} \right) \right] a_i
\]
Base solution around c=0

\[ \lambda v = \tilde{M}(\epsilon)v = (C + \epsilon G)v \]

\[ (\lambda_0 I - C)v^{(0)} = 0 \]

\[ \Rightarrow \quad v^{(0)}_{it} = a_i V_t \]

Solvability condition in next order:

\[ (\lambda_0 I - C)v^{(1)} = (G - \lambda_1 I)v^{(0)} \]

\[ L_0 \equiv (\lambda_0 I - C) \]

\[ L_1 \equiv (G - \lambda_1 I) \]

\[ L_0 v^{(1)} = L_1 v^{(0)} \]

\[ \forall i : \mathbf{V} \cdot [L_1 v^{(0)}]_i = 0 \]

\[ \sum_t \sin^2 \left( \frac{\pi t}{T+1} \right) G_{ij} a_j = \lambda_1 \left[ \sum_t \sin^2 \left( \frac{\pi t}{T+1} \right) a_i \right] \]

\[ \hat{L}a = 0 \]
First correction around $c=0$

\[
\lambda v = \tilde{M}(\epsilon)v = (C + \epsilon G)v
\]

\[
L_0 \equiv (\lambda_0 I - C)
\]

\[
L_1 \equiv (G - \lambda_1 I)
\]

\[
L_0 v^{(1)} = L_1 v^{(0)}
\]

\[
v^{(1)} = L_0^+ L_1 v^{(0)} + "b_i V_t"\]

Solvability condition in next order:

\[
L_0 v^{(2)} = L_1 v^{(1)} - \lambda_2 v^{(0)}
\]

\[
\forall i : \mathbf{V} \cdot [L_1 v^{(1)} - \lambda_2 v^{(0)}]_i = 0
\]

\[
\hat{L}b = \lambda_2 \left[ \sum_t \sin^2 \left( \frac{\pi t}{T + 1} \right) \right] a - \mathbf{V} \cdot [L_1 L_0^+ L_1 v^{(0)}]_i
\]

$\lambda_2$ is given directly by requirement to zero out component in null space: the solvability condition has its own solvability condition.
Perturbative solution around $c=0$
Summary: Temporal Centrality

• Authority (“placement capacity”) of Ph.D. exchange to measure academic “prestige”

• Hub/Authority scores applied in multislice setting by block matrix coupling of authority matrices
  – While some other eigenvector centralities may be applied in multislice setting without modification, temporal hub and authority again calls to need to treat identity arcs differently than links

• Vary identity coupling to control temporal scale

• Strong identity limit is mathematically singular
  – Singular value decomposition extrapolation to limit
  – Perturbative expansion, with the wrinkle of a solvability condition inside another solvability condition

• Other centrality measures for temporal network data: e.g., Grindrod, Higham, Parsons & Estrada PRE 2011
Communities in Networks

Overview

Methods I: Clustering Network Data
Methods II: Temporal & Multiplex Network Data

Applications I: Modeling Network Dynamics
Applications II: Centrality in Temporal Network Data

THANK YOU!