Communities in Multilayer Networks

Peter J. Mucha
University of North Carolina at Chapel Hill
Acknowledgements:

• NSF, NICHD, NIGMS, JSMF
• Jean Carlson, Skyler Cranmer, James Fowler, Scott Grafton, Jeff Henderson, Jim Moody, Mason Porter
• Dani Bassett, Kaveri Chaturvedi Parker, J.-P. Onnela, Saray Shai, Dane Taylor
• Elizabeth Menninga, Natalie Stanley, Andrew Waugh, James Wilson
• Kevin Macon, Ye Pei, Thomas Richardson
Outline

1. Community Detection
   - Political party polarization (arXiv:0907.3509, Network Science 2013)
   - “Network analysis reveals sex- and antibiotic resistance-associated antivirulence targets in clinical uropathogens” (r&r)

   - “Kantian fractionalization predicts the conflict propensity of the international system” (arXiv:1402.0126)


http://mucha.web.unc.edu/networks/
Philosophical Disclaimer

• Jim Moody (paraphrased): “I’ve been accused of turning everything into a network.”
• Me (in response): “I’m accused of turning everything into a network and a graph partitioning problem.”
Community Detection Reviews


• “Community Detection in Graphs,” S. Fortunato, Physics Reports 486, 75-174 (2010).
Voting as a network?

Scientific Coauthorship v. Roll Call Similarities
Polarization in Roll Call Networks
(Waugh, Pei, Fowler, Mucha & Porter, arXiv:0907.3509)
Network analysis reveals host sex associations and antibiotic resistance in uropathogenic *E. coli* (Chaturvedi, Wilson, Marschall, Mucha & Henderson, in r&r)
Community Structure in Multilayer Networks

Exploiting networks that vary across time, tie type, and community scale
“Multislice” Networks

- Intralayer identity arcs between node-layers representing same individual in different layers
- Harder: Include identity arcs in null model
  What does “relative to a random model” mean for identity arcs when they are definitional?
- Generalized Lambiotte et al. (2008) connection between modularity and autocorrelation in Laplacian dynamics

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

Ordered Categoric
Multilayer Modularity Derivation

\[ Q = \frac{1}{2W} \sum_{i,j} B_{ij} \delta(C_i, C_j), \]

\[ B_{ij} = A_{ij} - P_{ij} \]

\[ P_{ij} = \gamma \frac{k_i k_j}{2W} \]

- Generalized Lambiotte *et al.* to bipartite (Barber), directed (Leicht-Newman), and signed (Traag *et al.*)

\[ Q_{\text{multislice}} = \frac{1}{2\mu} \sum_{i,j,s} \left\{ \left( A_{ijs} - \gamma_s \frac{k_{is} k_{js}}{2m_s} \right) \delta_{sr} + \delta_{ij} C_{jsr} \right\} \delta(g_{is}, g_{jr}) \]

- intra-slice adjacency data and null
- inter-slice identity arcs
Mucha & Porter, *Chaos* 2010

Coupling = 0.2: 13 communities

Congress #
Mucha & Porter, *Chaos* 2010

**Coupling = 0.5: 8 communities**

- 3270R, 328D, 43W, 63other
- 1997D, 159R, 72other
- 384D, 162J, 72other
- 179W, 135DR, 97AJ, 78D, 49A, 69other
- 1194D, 99R, 41

**States:**

- AK, WA, CA, OR, WY, UT, NM, NV, MT, ID, CO, AZ, WY, TN, MD, MD, KY, TX, SC, NC, LA, FL, AR, AL, VA, NC, MO, MO, WI, IN, PA, NJ, VT, NH, DE, RI, MA, ME, VT, NH, ME

**Congress #:**

- 10-110

Mucha & Porter, *Chaos* 2010

**Coupling = 0.8: 6 communities**

- 2280D, 1260R, 223W, 97AJ, 68DR, 49A, 151other
- 2181R, 185D, 34other
- 424D, 286DR, 162J, 123other
- 151F, 50DR, 1PA
- 39PA, 20F, 7AA

![Graph showing communities across Congresses](Image)
Robust detection of dynamic community structure in networks

Danielle S. Bassett,1,2,a) Mason A. Porter,3,4 Nicholas F. Wymbs,5 Scott T. Grafton,5 Jean M. Carlson,1 and Peter J. Mucha6,7
CHAOS 23, 013142 (2013)

<table>
<thead>
<tr>
<th>Brain</th>
<th>Behavior</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q, Q^n</td>
<td>z, z^n</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Mean</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>γ</td>
<td>γ</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Q, Q^t</th>
<th>z, z^t</th>
<th>Q, Q^t</th>
<th>z, z^t</th>
</tr>
</thead>
<tbody>
<tr>
<td>x 10^{-3}</td>
<td></td>
<td>x 10^{-3}</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Mean</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>γ</td>
<td>γ</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Q, Q^t</th>
<th>z, z^t</th>
<th>Q, Q^t</th>
<th>z, z^t</th>
</tr>
</thead>
<tbody>
<tr>
<td>x 10^{-3}</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Mean</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>γ</td>
<td>γ</td>
</tr>
</tbody>
</table>

[Images of heat maps showing data distributions]
Multiplex example: “Kantian Fractionalization” (Cranmer, Menninga & Mucha, arXiv:1402.1026)

- Images from 1950, 1975, 2000
- Granger causal relationship to total conflicts in system
- Negligible effect from joint democracy
Multilayer SBM

arXiv:1410.8597v1
Consistent Estimation of Dynamic and Multi-layer Networks

Qiuyi Han1, Kevin S. Xu2, and Edoardo M. Airoldi1
1Department of Statistics, Harvard University, Cambridge, MA, USA, qiuyihan@fas.harvard.edu, airoldi@fas.harvard.edu
2Technicolor Research, Los Altos, CA, USA, kevinxu@outlook.com

November 3, 2014

Abstract

Dynamic networks where edges appear and disappear over time and multi-layer networks that deal with multiple types of connections arise in many applications. In this paper, we consider the multi-graph stochastic block model proposed by [Holland et al. (1983)], which serves as a foundation for both dynamic and multi-layer networks. We extend inference techniques in the analysis of single networks, namely maximum-likelihood estimation, spectral clustering, and variational approximation, to the multi-graph stochastic block model. Moreover we provide sufficient conditions for consistency of the spectral clustering and maximum-likelihood estimates. We verify the conditions for our results via simulation and demonstrate that the conditions are practical. In addition, we apply the model to two real data sets: a dynamic social network and a multi-layer social network, resulting in block estimates that reveal network structure in both cases.

arXiv:1411.1098v1
Multilayer stochastic block models reveal the multilayer structure of complex networks

Toni Vallès-Català,1 Francesco A. Massucci,1 Roger Guimerà,2,1 and Marta Sales-Pardo1
1Departament d’Enginyeria Química, Universitat Rovira i Virgili, 43007 Tarragona, Catalonia
2Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona 08010, Catalonia

In complex systems, the network of interactions we observe between system’s components is the aggregate of the interactions that occur through different mechanisms or layers. Recent studies reveal that the existence of multiple interaction layers can have a dramatic impact in the dynamical processes occurring on these systems. However, these studies assume that the interactions between systems components in each one of the layers are known, while typically for real-world systems we do not have that information. Here, we address the issue of uncovering the different interaction layers from aggregate data by introducing multilayer stochastic block models (SBMs), a generalization of single-layer SBMs that considers different mechanisms of layer aggregation. First, we find the complete probabilistic solution to the problem of finding the optimal multilayer SBM for a given aggregate observed network. Because this solution is computationally intractable, we propose an approximation that enables us to verify that multilayer SBMs are more predictive of network structure in real-world complex systems.
Strata MLSBM (sMLSBM)
(Stanley, Shai, Taylor & Mucha, arXiv:1507.01826)
Strata MLSBM (sMLSBM)

(Stanley, Shai, Taylor & Mucha, arXiv:1507.01826)

Initialization

Layer $l$ to stratum $s$

Iterative Process

Stratum $s$

1. Update number of strata to the number of unique clustering patterns according to (1) and (2)

2. K-means cluster $2L$ layers in $S$ strata

K-means cluster $L$ layers in to $S$ strata
sMLSBM on SparCC microbial interactions
(Stanley, Shai, Taylor & Mucha, arXiv:1507.01826)
Summary

• Communities are useful for exploring and analyzing network data
• Codes available: Louvain, GenLouvain, MapEquation.org, graph-tool
• Currently only very few options available for handling multilayer network data.
  – Especially if restriction to single layer should reduce to something you already understand
"Community Structure"
(Girvan & Newman, PNAS 2002)

Fig. 7. Hierarchical tree for the Chesapeake Bay food web described in the text.

Fig. 5. Hierarchical tree for the network reflecting the schedule of regular-season Division I college football games for year 2000. Nodes in the network represent teams, and edges represent games between teams. Our algorithm identifies nearly all of the conference structure in the network.
Modularity

\[ Q = \frac{1}{2W} \sum_{i,j} B_{ij} \delta(C_i, C_j), \]
\[ B_{ij} = A_{ij} - P_{ij}, \]
\[ P_{ij} = \gamma \frac{k_i k_j}{2W} \]

- **GOAL:** Assign nodes to communities to maximize Q
- NP-hard \(\sim\) enumerate possible partitions
  (Brandes et al. 2008)
- Numerous packages available
- Resolution parameter (Reichardt & Bornholdt 2006)
- Resolution limit (Fortunato & Barthelemy, PNAS 2007)
- Degenerate landscape (Good, de Montjoye & Clauset, PRE 2010)
- Partition (many authors!)
Zachary Karate Club

This partition optimizes *modularity*, which measures the weight of intra-community ties (relative to a random model).

“If your method doesn’t work on this network, then go home.”
Karate Club Example

Brought to you by Mason Porter at The Power Law Shop
http://www.cafepress.com/thepowerlawshop
“Cris Moore (left) is the inaugural recipient of the Zachary Karate Club Club prize, awarded on behalf of the community by Aric Hagberg (right). (9 May 2013)”
Facebook
(Traud, Kelsic, PJM & MAP, SIAM Review 2011; Traud, PJM & MAP, Physica A 2012)

Caltech:
Colors indicate residential “House” affiliations
Purple = Not provided
Facebook
(Traud, Kelsic, PJM & MAP,
SIAM Review 2011;
Traud, PJM & MAP,
Physica A 2012)

Caltech:
Colors indicate residential
“House” affiliations
Purple = Not provided
Facebook
(Traud, Kelsic, PJM & MAP, SIAM Review 2011; Traud, PJM & MAP, Physica A 2012)

Caltech:
Colors indicate residential “House” affiliations
Purple = Not provided
**Facebook**


Caltech Facebook (2005)  
N=762

Logistic Regression:

<table>
<thead>
<tr>
<th>Caltech 36</th>
<th>Edge</th>
<th>Year</th>
<th>Residence</th>
<th>High School</th>
<th>Major</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full</td>
<td>-3.6903(0.012891)</td>
<td>1.5382(0.018233)</td>
<td>2.4151(0.018644)</td>
<td>2.3789(0.14869)</td>
<td>0.53388(0.02881)</td>
</tr>
</tbody>
</table>
Facebook
(Traud, Kelsic, PJM & MAP,
*SIAM Review* 2011;
Traud, PJM & MAP,
*Physica A* 2012)

Caltech Facebook
(2005)
N=762
Logistic Regression:

<table>
<thead>
<tr>
<th>Z_{Rand}</th>
<th>Major</th>
<th>Residence</th>
<th>Graduation year</th>
<th>High School</th>
</tr>
</thead>
<tbody>
<tr>
<td>Institution and network</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Caltech 36 Full</td>
<td>4.0649</td>
<td>222.9566</td>
<td>8.5967</td>
<td>4.9078</td>
</tr>
</tbody>
</table>
Congressional Committees
(MAP, PJM, Newman, Warmbrand & Friend, 
Multilayer Networks


\[ M = (V_M, E_M, V, L) \]

- \( V \): set of nodes (as in ordinary graphs)
- \( L \): set of layers (categorical or ordered across 1 or more dimensions)
- \( V_M \): set of tuples that represent node-layers
- \( E_M \): multilayer edge set connecting these tuples (possibly with weights)
muxViz: Visualization and Analysis of Multilayer Networks
De Domenico, Porter & Arenas, *J. Complex Networks* (advance access)
Multilayer Map Equation
Multilayer Map Equation

FIG. 4. Community structure in the Pierre Auger Collaboration network. (a) The overlapping community structure revealed by the multiplex map equation with relax rate $r = 0.15$. Nodes for scientists are colored according to their module assignments, with node sizes proportional to the number of tasks in which they were active. Specifically, the area of a colored pie-chart slice is proportional to the number of tasks in which the corresponding scientist is active. (b) Subsets of nodes with direct comparison with the overlapping community structure obtained from dynamics with $r = 1.0$. 
Summary

- Communities are useful for exploring and analyzing network data
- Codes available: Louvain, GenLouvain, MapEquation.org, graph-tool
- Currently only very few options available for handling multilayer network data.
  - Especially if restriction to single layer should reduce to something you already understand