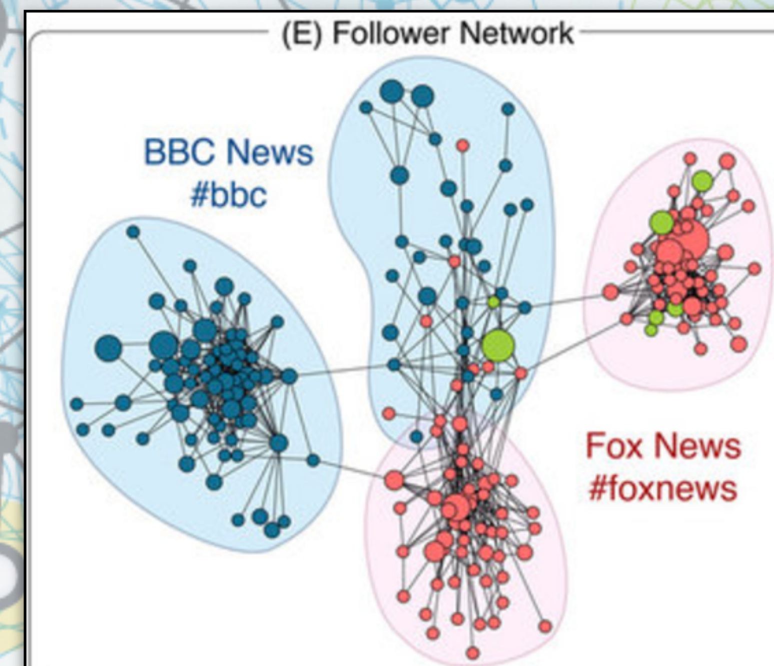
Data science with multilayer networks: Mathematical foundations and applications

CDSE Days
University at Buffalo, State University of New York
Monday April 9, 2018

Dane Taylor
Assistant Professor of Mathematics
University at Buffalo, State University of New York

Large-scale networks arise in numerous research fields to describe social networks, biological systems, technologies, economics, …

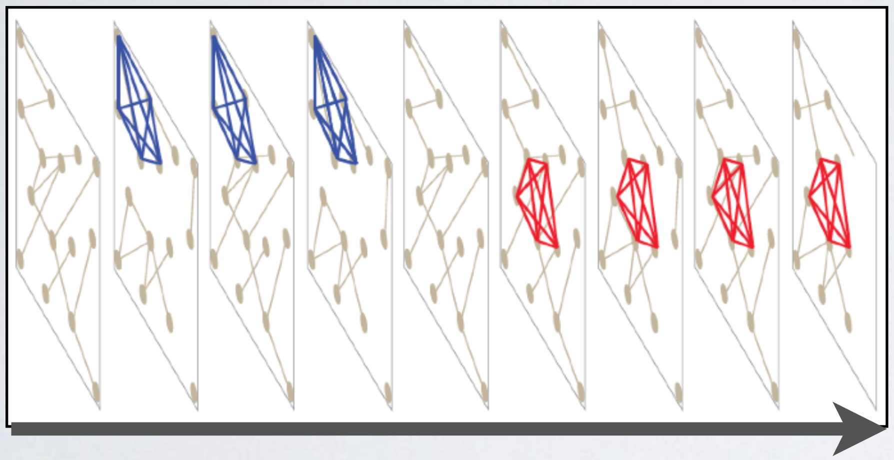


There are myriad techniques for extracting insights from structural patterns

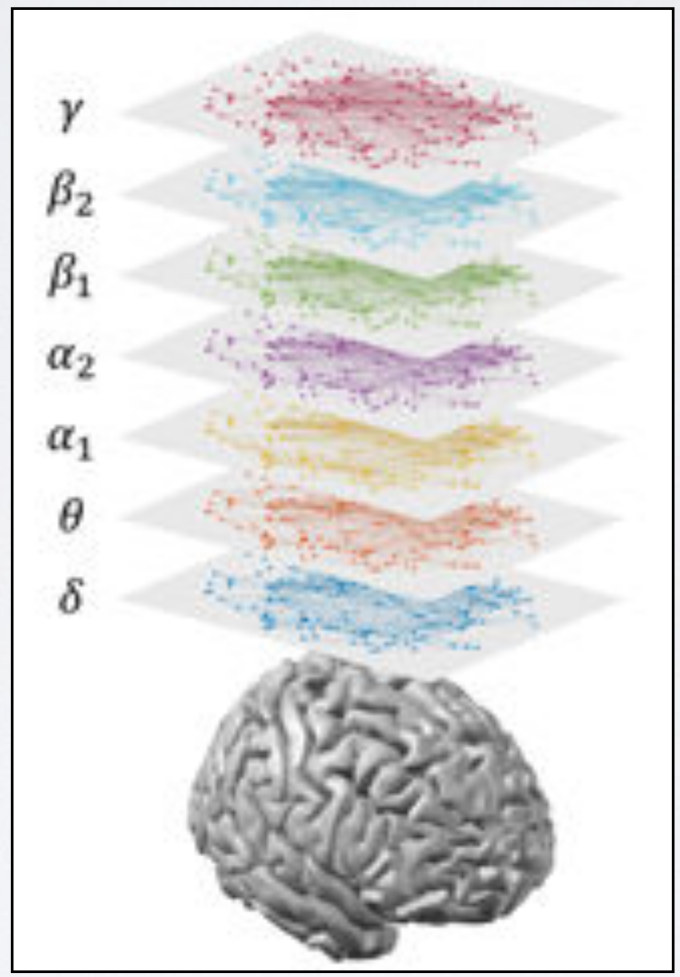
MULTILAYER NETWORKS

- A more comprehensive modeling framework

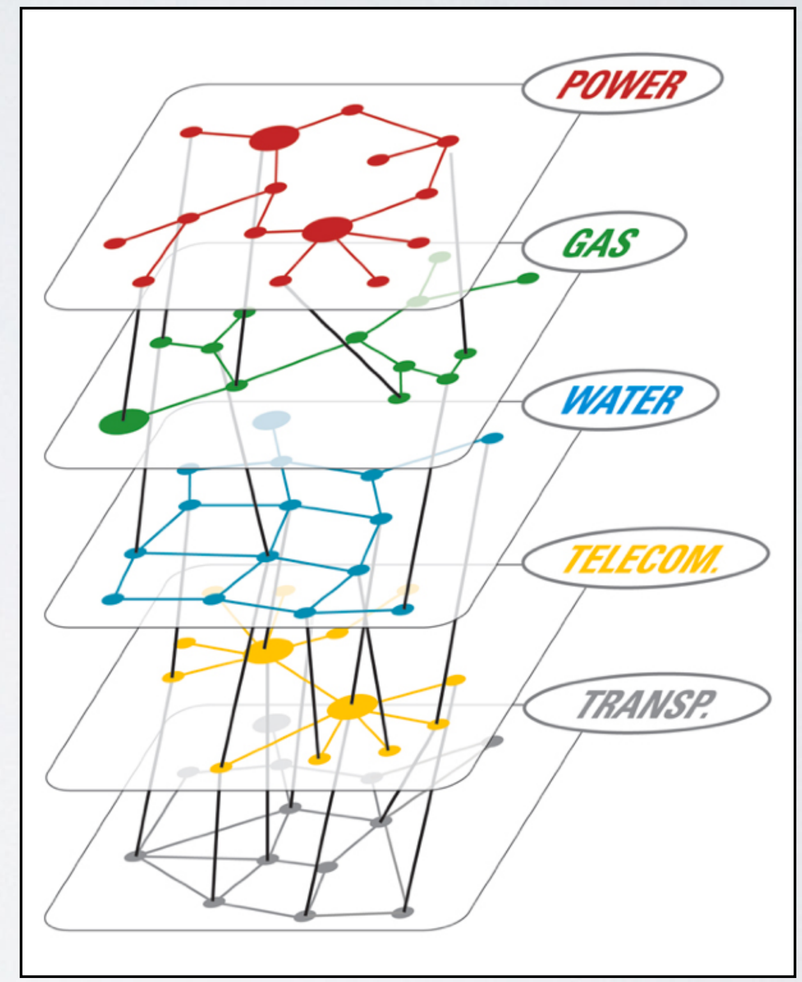
time-varying networks



complementary data



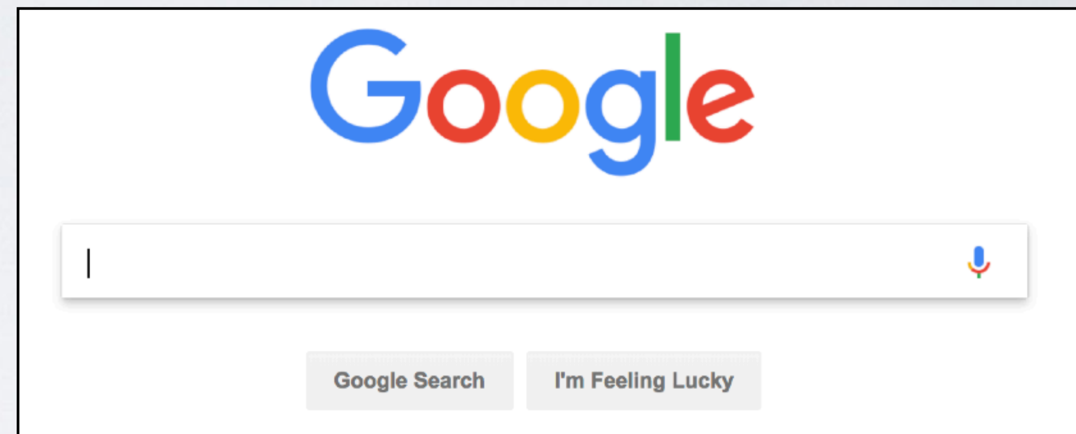
interdependent systems

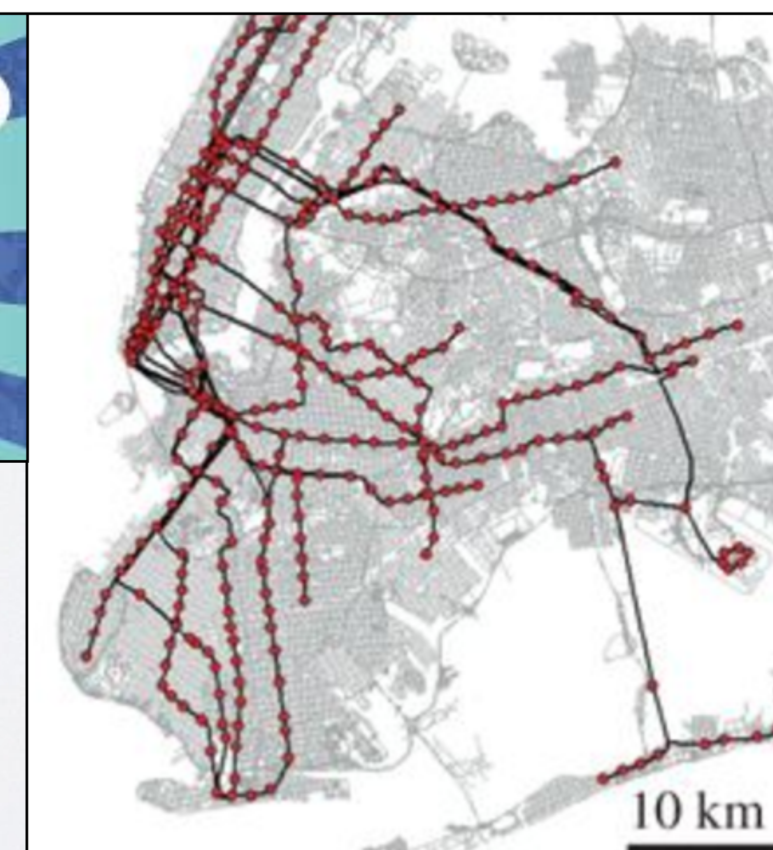


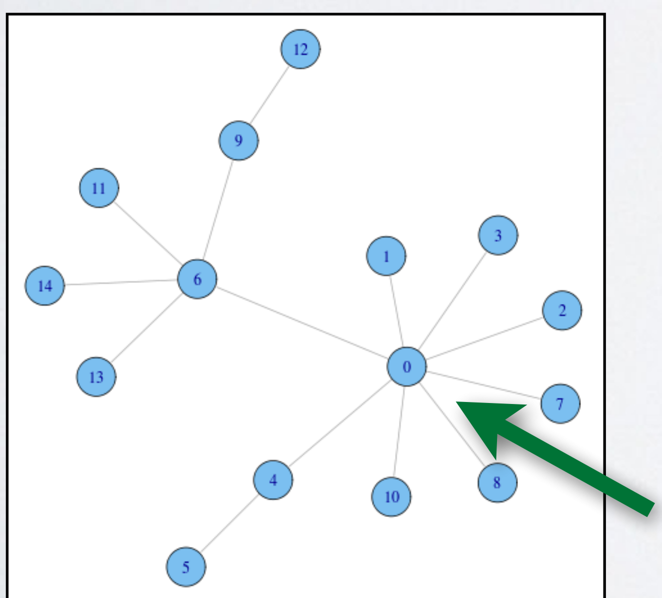
- See reviews
  - Kivelä et al. (2014) Multilayer networks. J. of Complex Networks 2(3), 203-271.
  - Boccaletti et al. (2014) The structure and dynamics of multilayer networks. Physics Reports 544(1), 1-122.

CENTRALITY AND RANKING

- **Centrality Analysis -** ranking nodes according to their importances
  - Google PageRank for web search
  - Identifying influential persons
  - Points of fragility in complex systems
  - Ranking universities, academics, etc.
  - Drug targets in biological networks



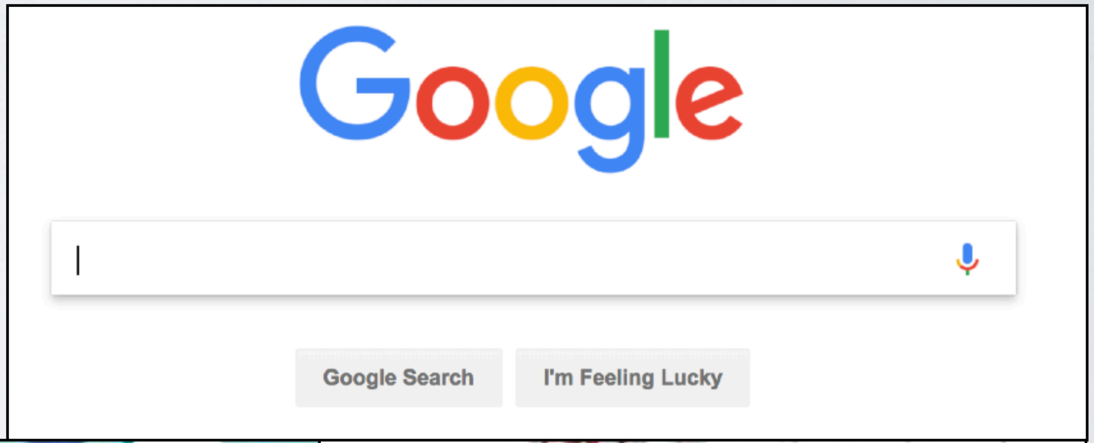


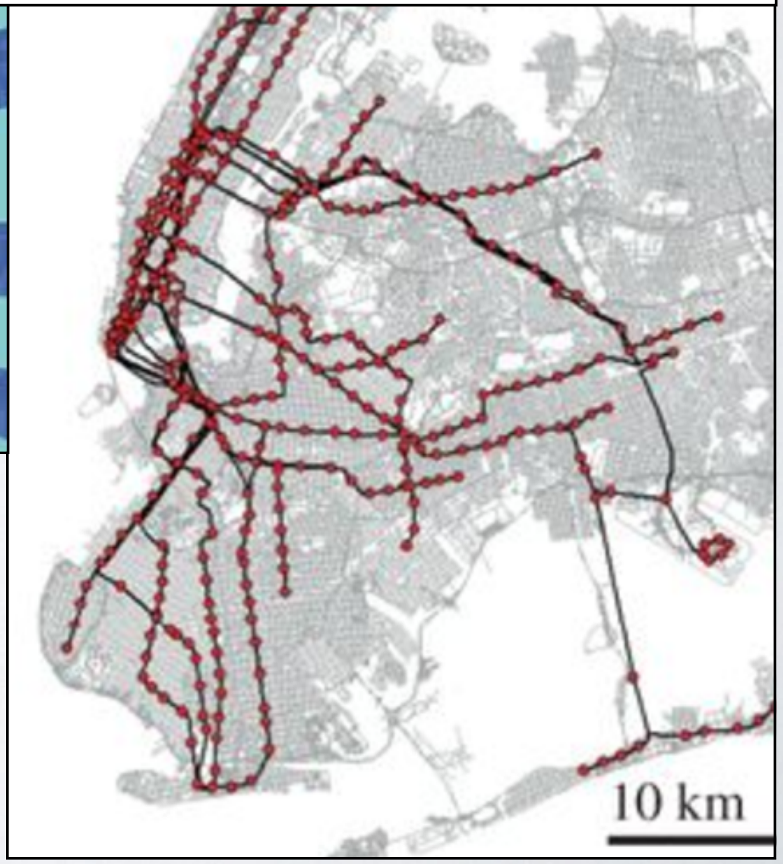


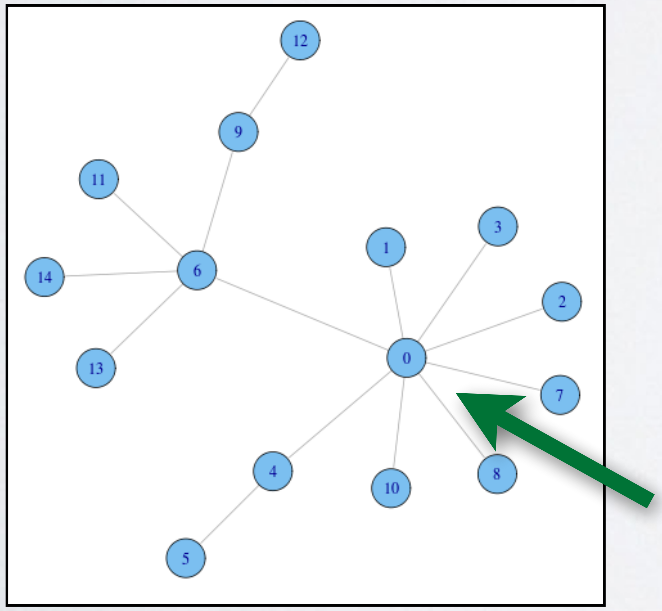
most important

CENTRALITY AND RANKING

- **Centrality Analysis -** ranking nodes according to their importances
  - **Google PageRank** for web search
  - Identifying influential persons
  - Points of fragility in complex systems
  - Ranking universities, academics, etc.
  - Drug targets in biological networks





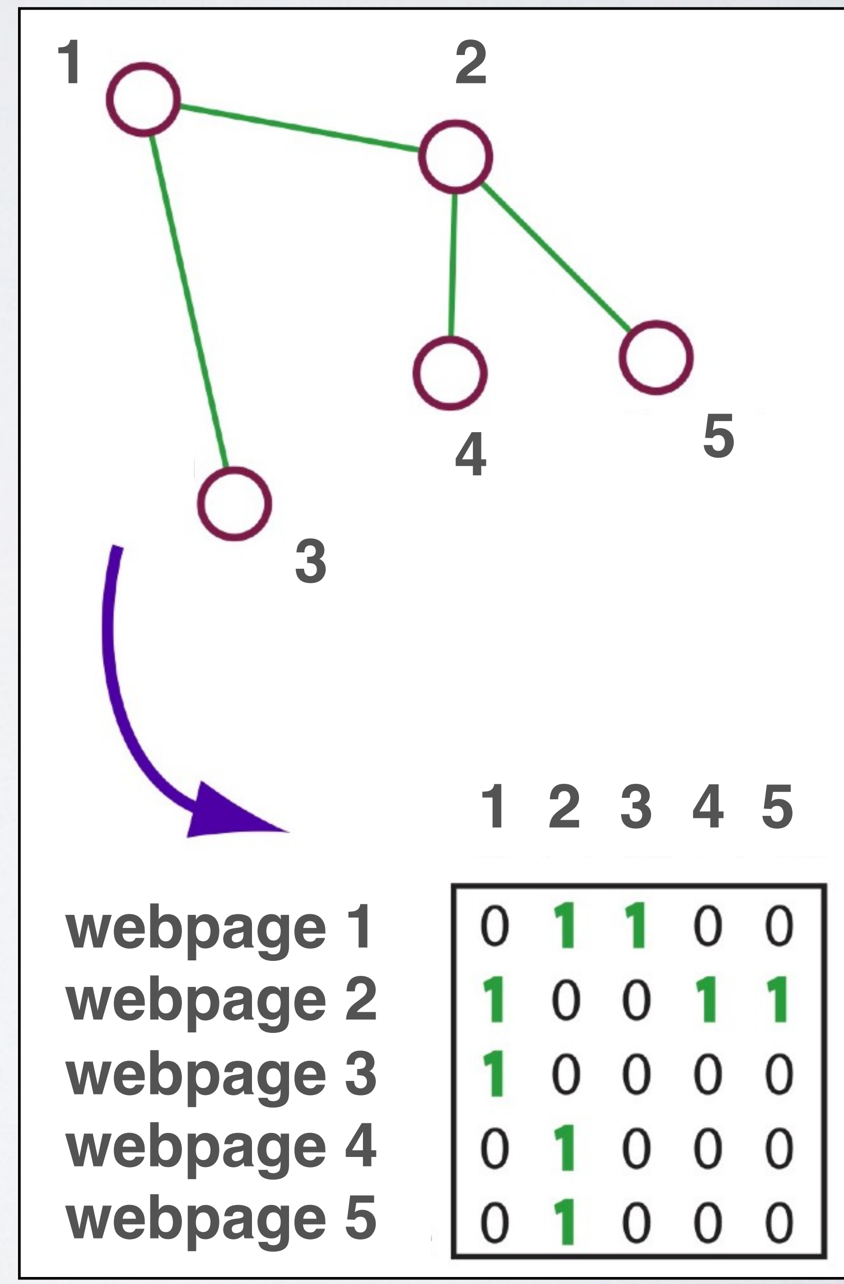


most important

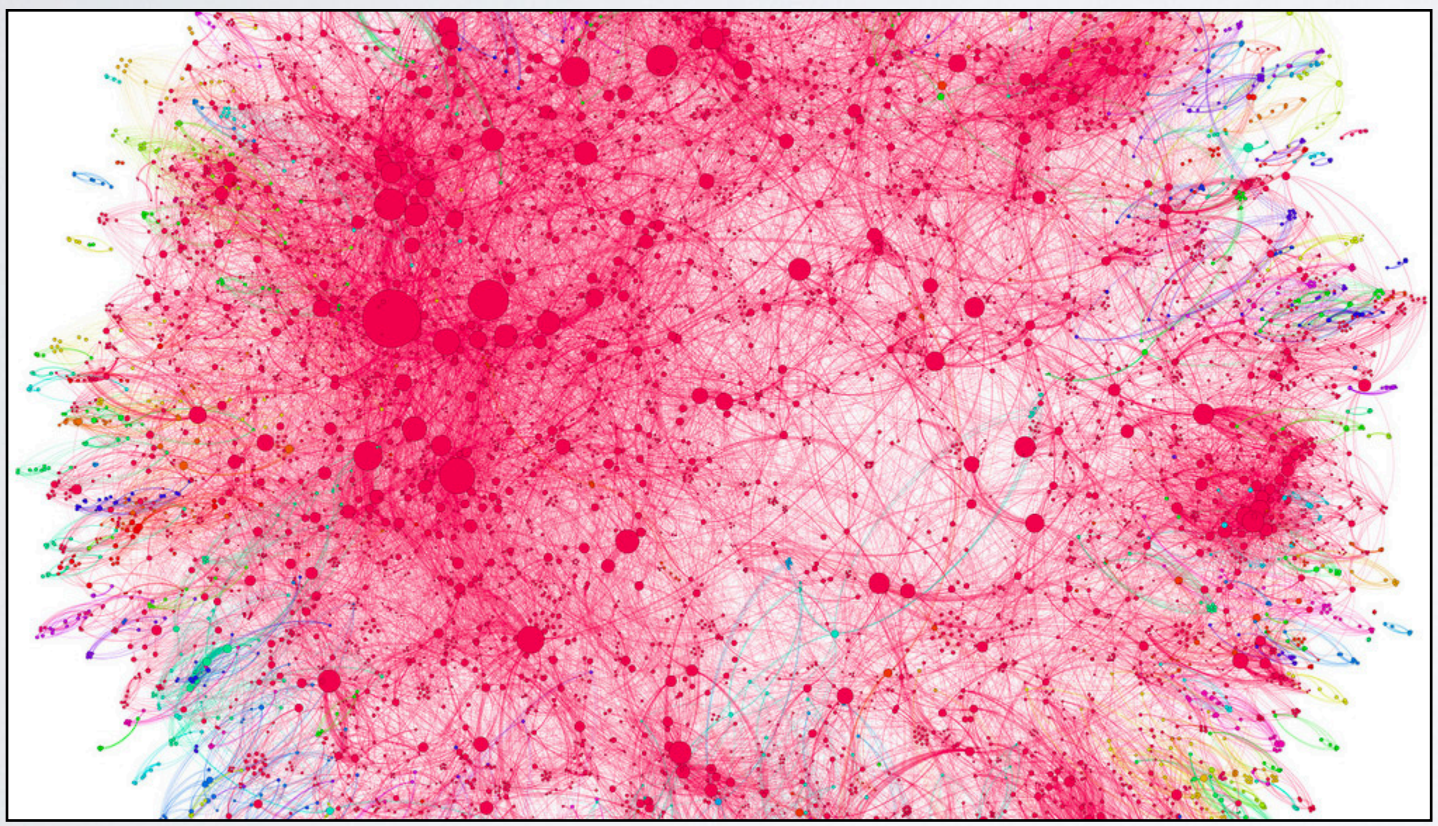
GOOGLE'S PAGERANK ALGORITHM

- Adjacency matrix A_{ij} - encodes network

example web graph



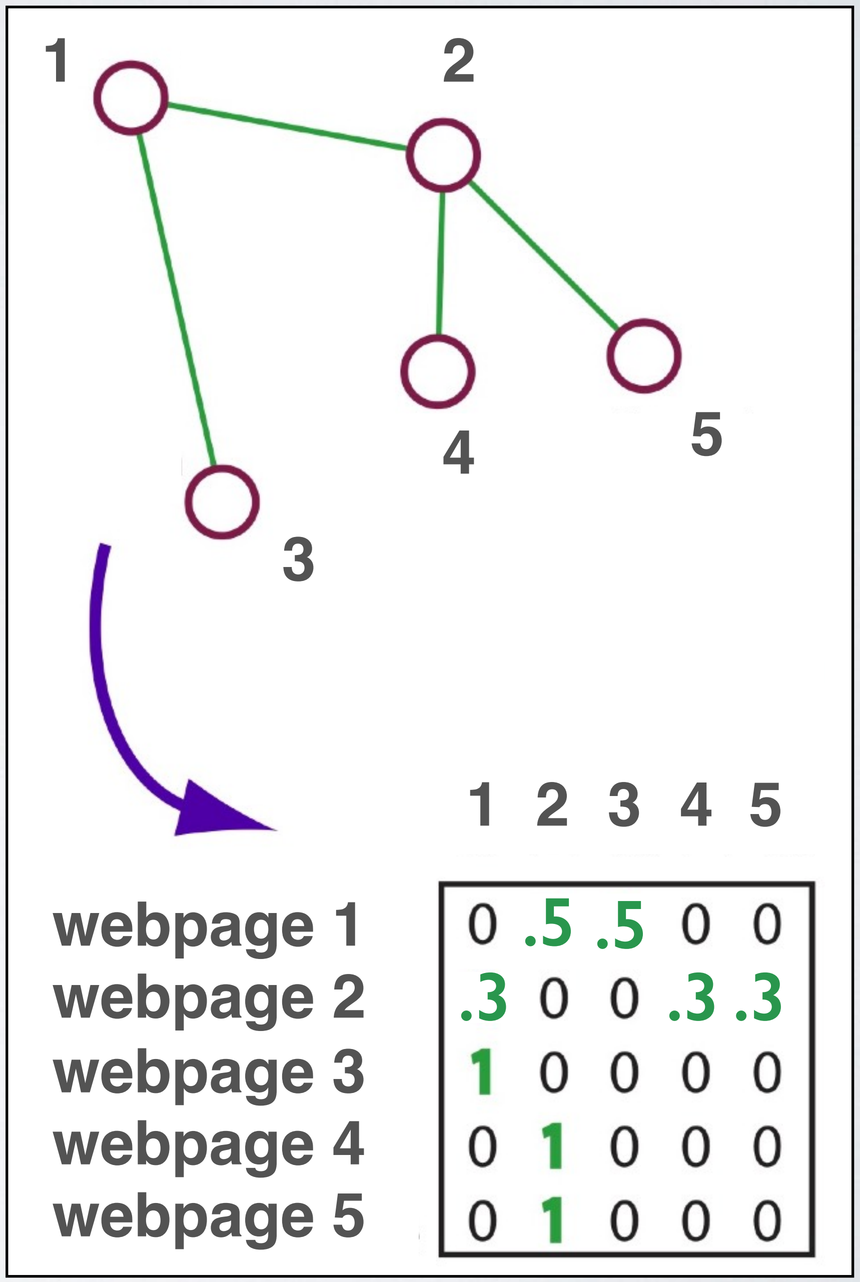
massive web graph



GOOGLE'S PAGERANK ALGORITHM

- Adjacency matrix A_{ij} - encodes network
- Node degree $d_i = \sum_j A_{ij}$
- Transition matrix $P_{ij} = A_{ij}/d_i$

example web graph



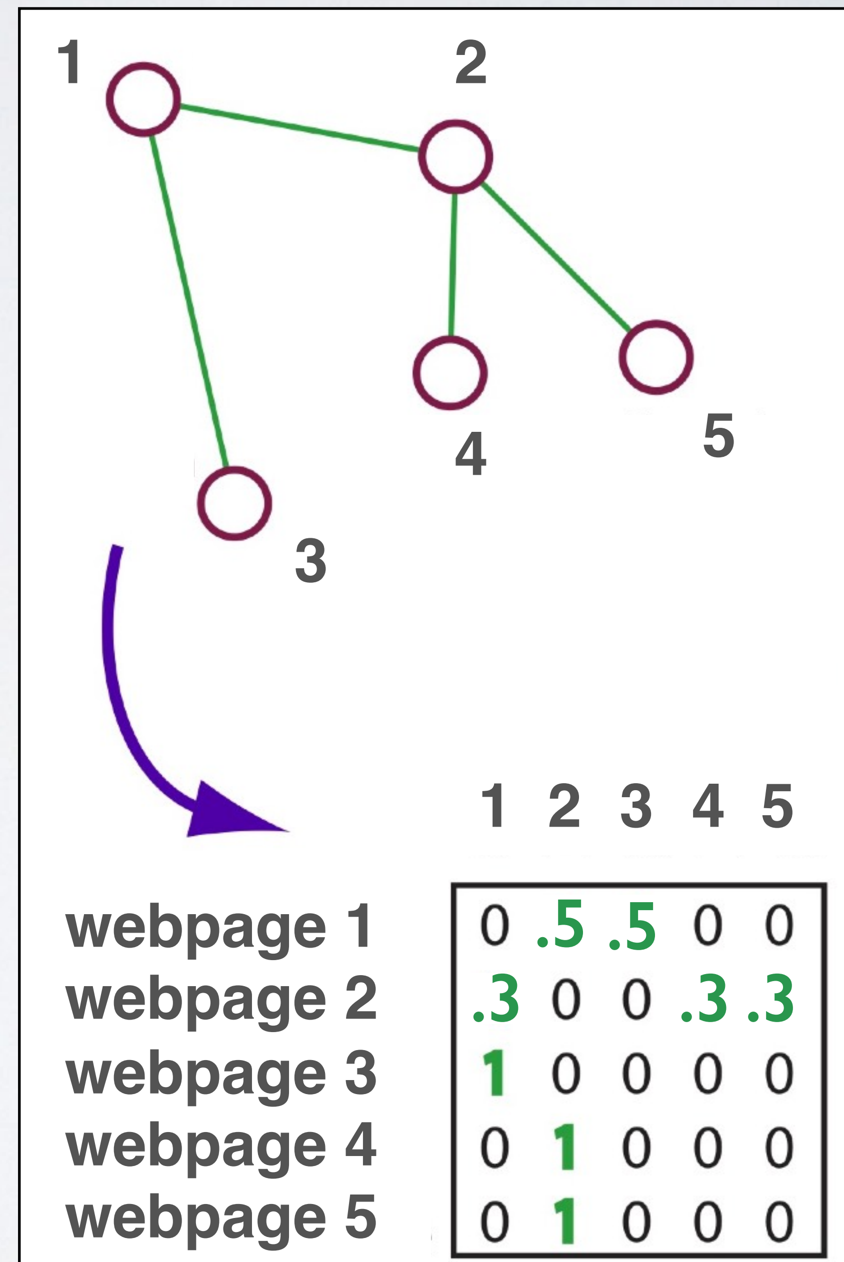
GOOGLE'S PAGERANK ALGORITHM

- Adjacency matrix A_{ij} - encodes network
- Node degree $d_i = \sum_j A_{ij}$
- Transition matrix $P_{ij} = A_{ij}/d_i$
- PageRank centrality matrix

$$C_{ij}^{(PR)} = (1-\alpha)P_{ij} + \alpha$$

$$\alpha = 0.15$$

example web graph



GOOGLE'S PAGERANK ALGORITHM

- Adjacency matrix A_{ij} - encodes network
- Node degree $d_i = \sum_j A_{ij}$
- Transition matrix $P_{ij} = A_{ij}/d_i$
- PageRank centrality matrix

$$C_{ij}^{(PR)} = (1-\alpha)P_{ij} + \alpha$$

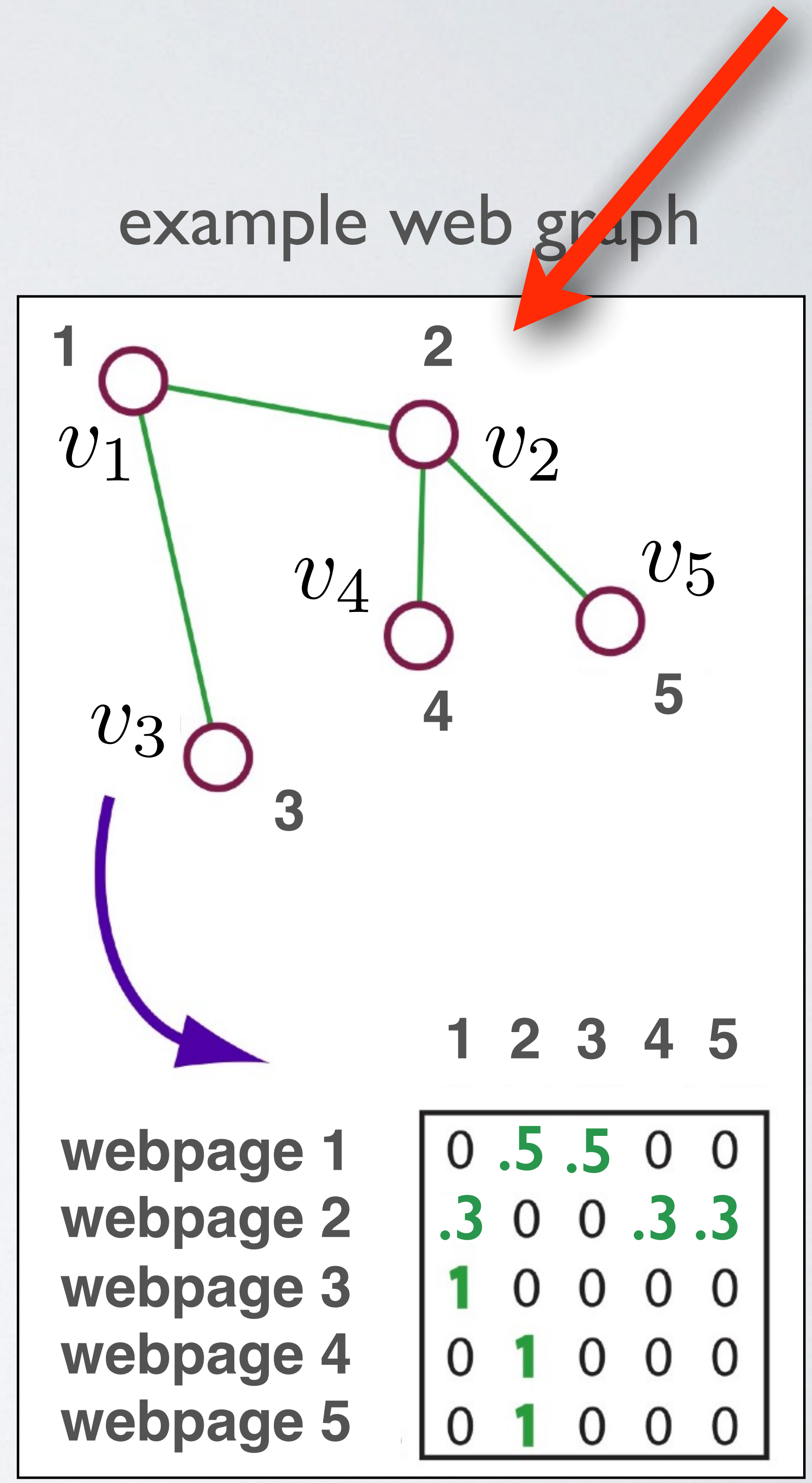
$$\alpha = 0.15$$

- PageRank vector $\mathbf{v}$ solves

$$C^{(PR)}\mathbf{v} = \lambda_{max}\mathbf{v}$$

- $\{v_i\}$ gives stationary distribution of random web surfers

example web graph



EIGENVECTOR-BASED CENTRALITIES

- Node rankings are indicated by a centrality score, which is computed as the dominant eigenvector of a "centrality matrix" C

$$C\mathbf{v} = \lambda_{max}\mathbf{v}$$

- Examples
  - Google's PageRank centrality -Brin & Page, 1998

$$C^{(PR)} = (1-\alpha)P + \alpha$$

  - Hub and Authority scores for directed networks -Kleinberg, 1999

$$C^{(hub)} = A^T A$$

$$C^{(auth)} = AA^T$$

  - Eigenvector centrality for undirected networks

$$C^{(evec)} = A$$

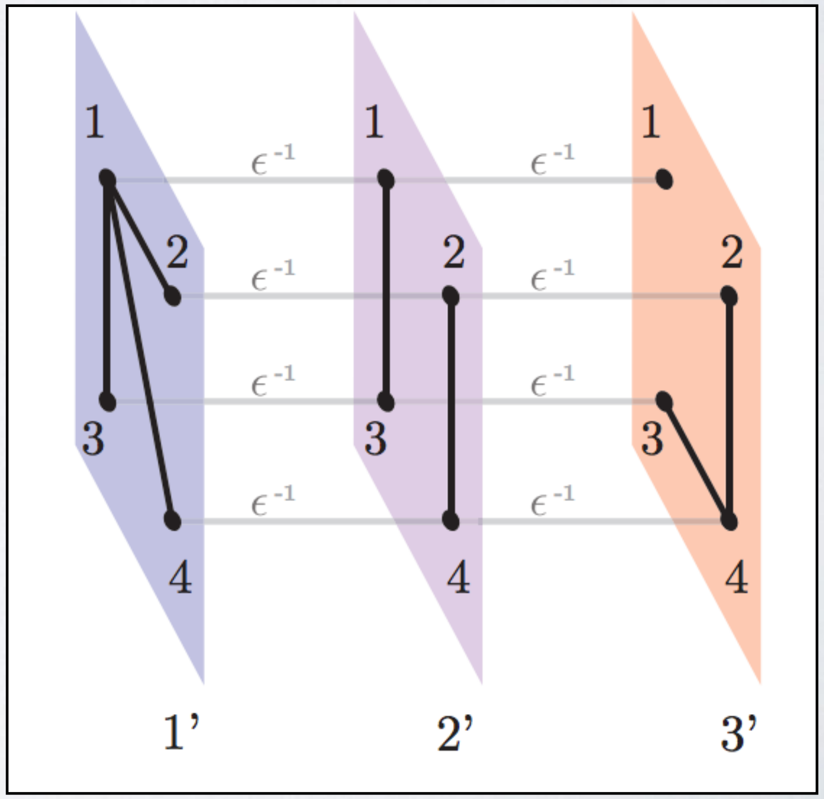
SUPRA-CENTRALITY MATRIX

- We introduce "supra-centrality" matrices as a temporal generalization of centrality
  - Consider T centrality matrices $\mathbf{C}^{(t)}$ for T time layers
  - We place them as diagonal blocks in a matrix and couple them together with inter-layer "identity" edges of weight ϵ^{-1}

$$\mathbb{C}(\epsilon) = \begin{bmatrix} \mathbf{C}^{(1)} & \epsilon^{-1}\mathbf{I} & 0 & \cdots \\ \epsilon^{-1}\mathbf{I} & \mathbf{C}^{(2)} & \epsilon^{-1}\mathbf{I} & \ddots \\ 0 & \epsilon^{-1}\mathbf{I} & \mathbf{C}^{(3)} & \ddots \\ \vdots & \ddots & \ddots & \ddots \end{bmatrix}$$

$$\mathbb{C}(\epsilon)\mathbf{v}(\epsilon) = \lambda_{\max}\mathbf{v}(\epsilon)$$

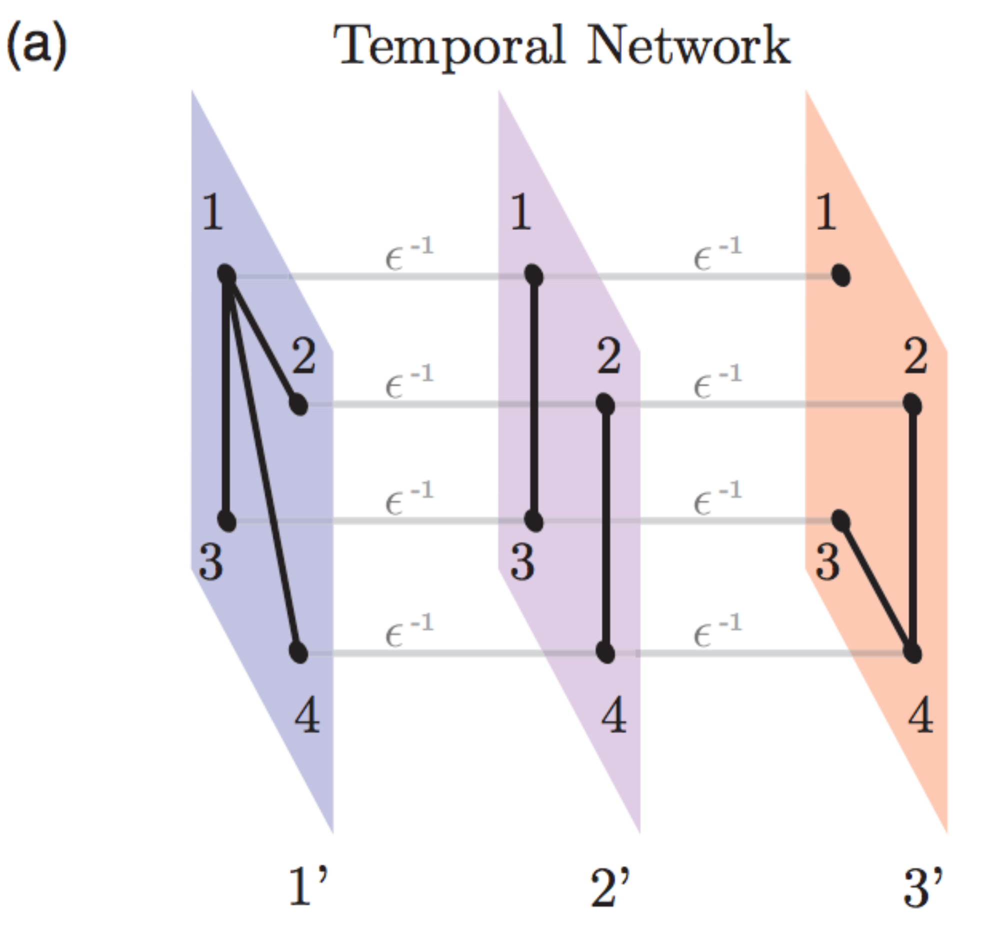
coupled time layers



JOINT, *MARGINAL* AND *CONDITIONAL* CENTRALITY

- We introduce a vocabulary for centrality motivated by statistics

$$\mathbb{C}(\epsilon)\mathbf{v}(\epsilon) = \lambda_{\max}\mathbf{v}(\epsilon)$$



(b) Centralities for $\epsilon = 0.5$

Time Layer

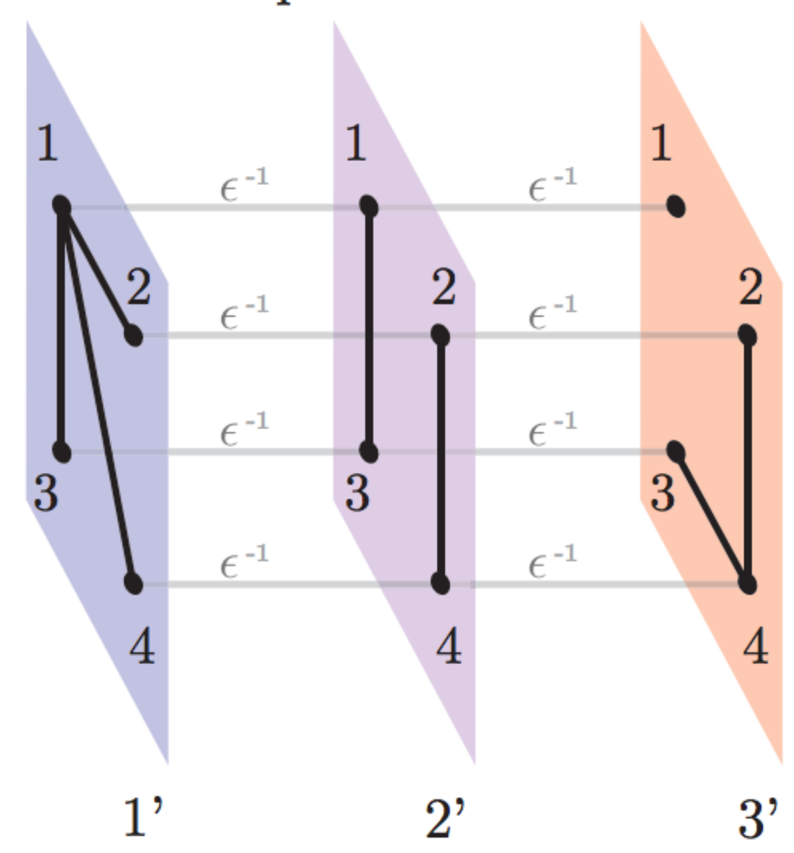
Node Index	1'	2'	3'	MNC
1	0.3570	0.3392	0.1679	0.8641
2	0.2516	0.3298	0.2357	0.8171
3	0.2471	0.3207	0.2312	0.7990
4	0.2655	0.3579	0.2927	0.9162
MLC	1.1213	1.3476	0.9275	

JOINT, *MARGINAL* AND *CONDITIONAL* CENTRALITY

- We introduce a vocabulary for centrality motivated by statistics

$$\mathbb{C}(\epsilon)\mathbf{v}(\epsilon) = \lambda_{\max}\,\mathbf{v}(\epsilon)$$

(a) Temporal Network



(b) Centralities for $\epsilon = 0.5$

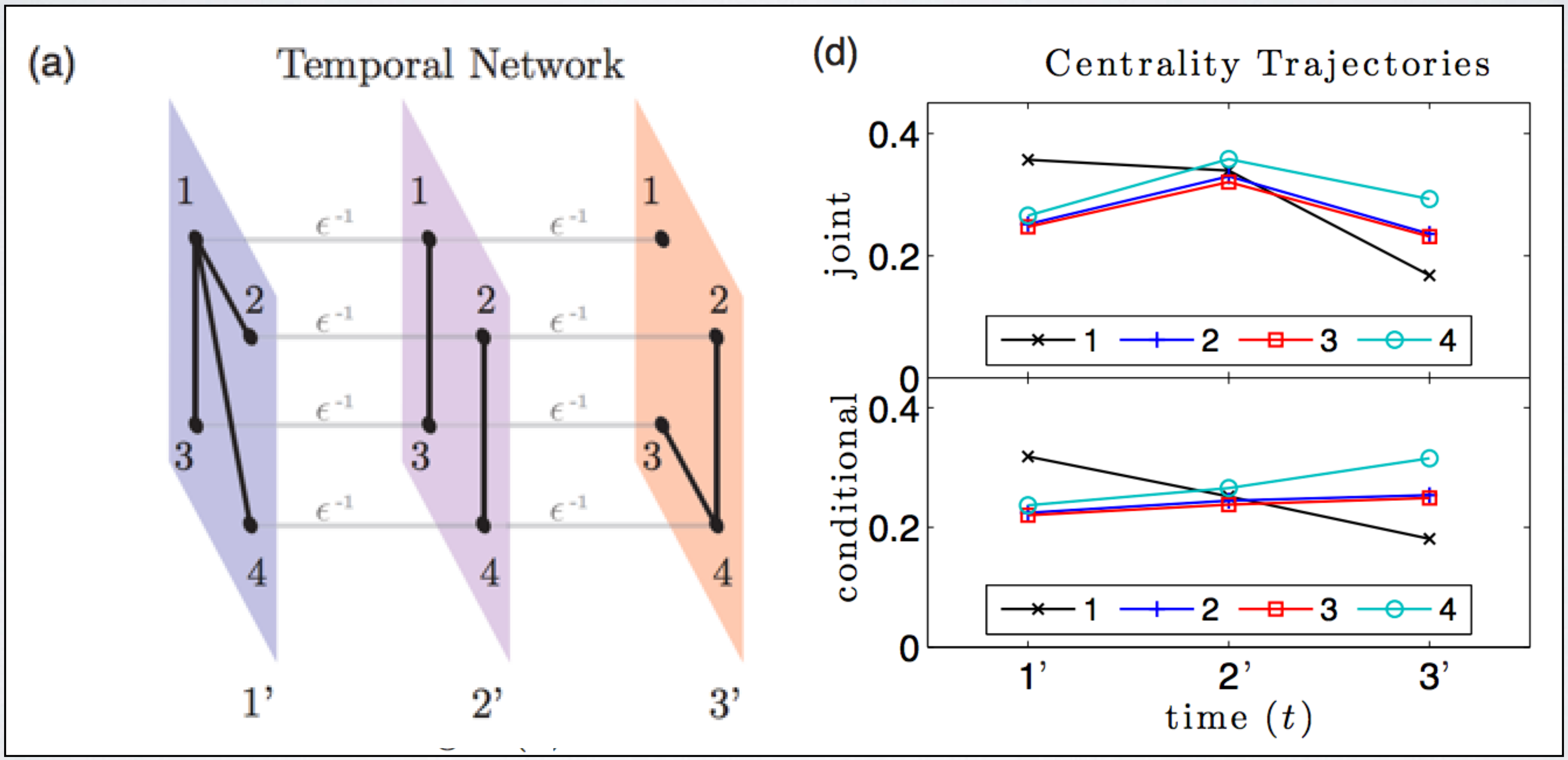
Time Layer

Node Index	1'	2'	3'	MNC
1	0.3570	0.3392	0.1679	0.8641
2	0.2516	0.3298	0.2357	0.8171
3	0.2471	0.3207	0.2312	0.7990
4	0.2655	0.3579	0.2927	0.9162
MLC	1.1213	1.3476	0.9275	

$$\frac{0.2516}{1.1213} = 0.2244$$

CENTRALITY TRAJECTORIES

- Conditional centrality reveals how the importances of nodes change with time



STRONG-COUPLING LIMIT: TIME-AVERAGED CENTRALITY

- We analyze the strong coupling limit $\epsilon \to 0^+$ which yields layer aggregation
- We conduct a singular perturbation analysis for $\mathbb{C}(\epsilon)\mathbf{v}(\epsilon) = \lambda_{\max}\mathbf{v}(\epsilon)$

$$\mathbf{v}(\epsilon) = \mathbf{v}_0 + \epsilon\mathbf{v}_1 + \epsilon^2\mathbf{v}_2 + \dots$$
$$\lambda_{\max} = \lambda_0 + \epsilon\lambda_1 + \epsilon^2\lambda_2 + \dots$$

STRONG-COUPLING LIMIT: TIME-AVERAGED CENTRALITY

- We analyze the strong coupling limit $\epsilon \to 0^+$ which yields layer aggregation
- We conduct a singular perturbation analysis for $\mathbb{C}(\epsilon)\mathbf{v}(\epsilon) = \lambda_{\max}\mathbf{v}(\epsilon)$

$$\left| \begin{aligned} \mathbf{v}(\epsilon) &= \mathbf{v}_0 + \epsilon\mathbf{v}_1 + \epsilon^2\mathbf{v}_2 + \dots \\ \lambda_{\max} &= \lambda_0 + \epsilon\lambda_1 + \epsilon^2\lambda_2 + \dots \end{aligned} \right.$$

- This yields two layer-aggregated centralities:
  - Time-averaged centrality $\{\alpha_i\}$, which describe the conditional centralities when they become constant across the layers
  - First-order mover scores $\{m_i\}$, which measure the extent of centrality change across the layers

CASE STUDY 1:

- We analyzed data from the Internet Movie Database (IMDb) to study the *golden age of hollywood,* 1920–1960,
- We consider 55 actors (26 male and 29 female) during the time period 1909–2009
- An edge from i to j in layer t indicates the number of movies in which i and j co-star and actor j is listed first in the movies *billing order*

CASE STUDY I:

Top Time-Averaged Centralities

Rank	Actor	α_i	m_i
1	Gable, Clark	0.3683	136.32
2	Marx, Groucho	0.3627	163.34
3	Marx, Harpo	0.2844	112.28
4	Garland, Judy	0.2820	100.28
5	Tracy, Spencer	0.2681	98.20
6	Stewart, James	0.2371	78.78
7	Crawford, Joan	0.2369	90.58
8	Astaire, Fred	0.2103	73.29
9	Marx, Chico	0.2055	86.39
10	Cagney, James	0.1779	69.00

Top Time-Averaged Centralities

	Actor	α_i	m_i
	Gable, Clark	0.3683	136.32
	Marx, Groucho	0.3627	163.34
	Marx, Harpo	0.2844	112.28
4	Garland, Judy	0.2820	100.28
5	Tracy, Spencer	0.2681	98.20
6	Stewart, James	0.2371	78.78
7	Crawford, Joan	0.2369	90.58
8	Astaire, Fred	0.2103	73.29
9	Marx, Chico	0.2055	86.39
10	Cagney, James	0.1779	69.00

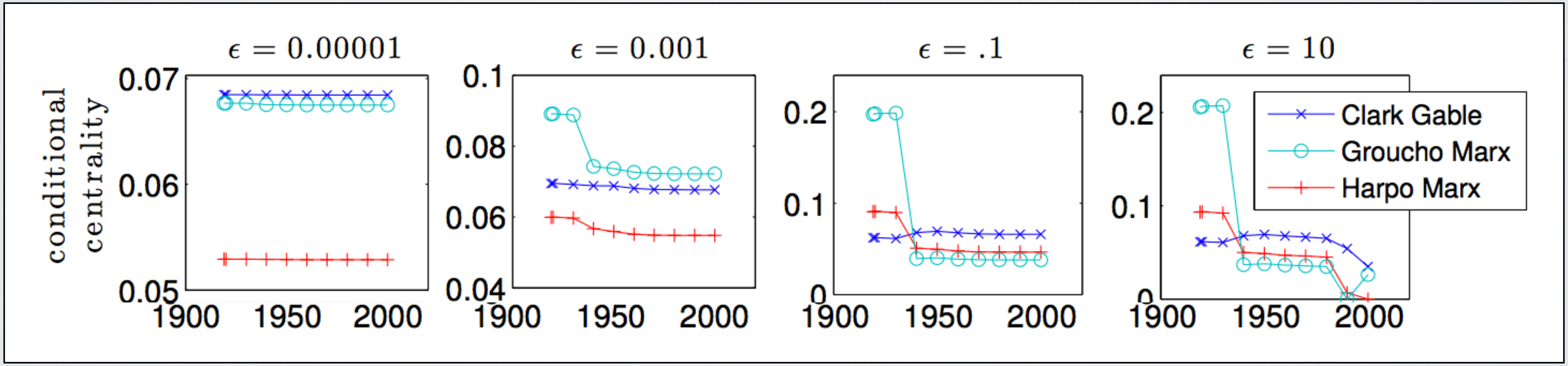
CASE STUDY I:

Top Time-Averaged Centralities

Rank	Actor	α_i	m_i
1	Gable, Clark	0.3683	136.32
2	Marx, Groucho	0.3627	163.34
3	Marx, Harpo	0.2844	112.28
4	Garland, Judy	0.2820	100.28
5	Tracy, Spencer	0.2681	98.20
6	Stewart, James	0.2371	78.78
7	Crawford, Joan	0.2369	90.58
8	Astaire, Fred	0.2103	73.29
9	Marx, Chico	0.2055	86.39
10	Cagney, James	0.1779	69.00

CASE STUDY 1:

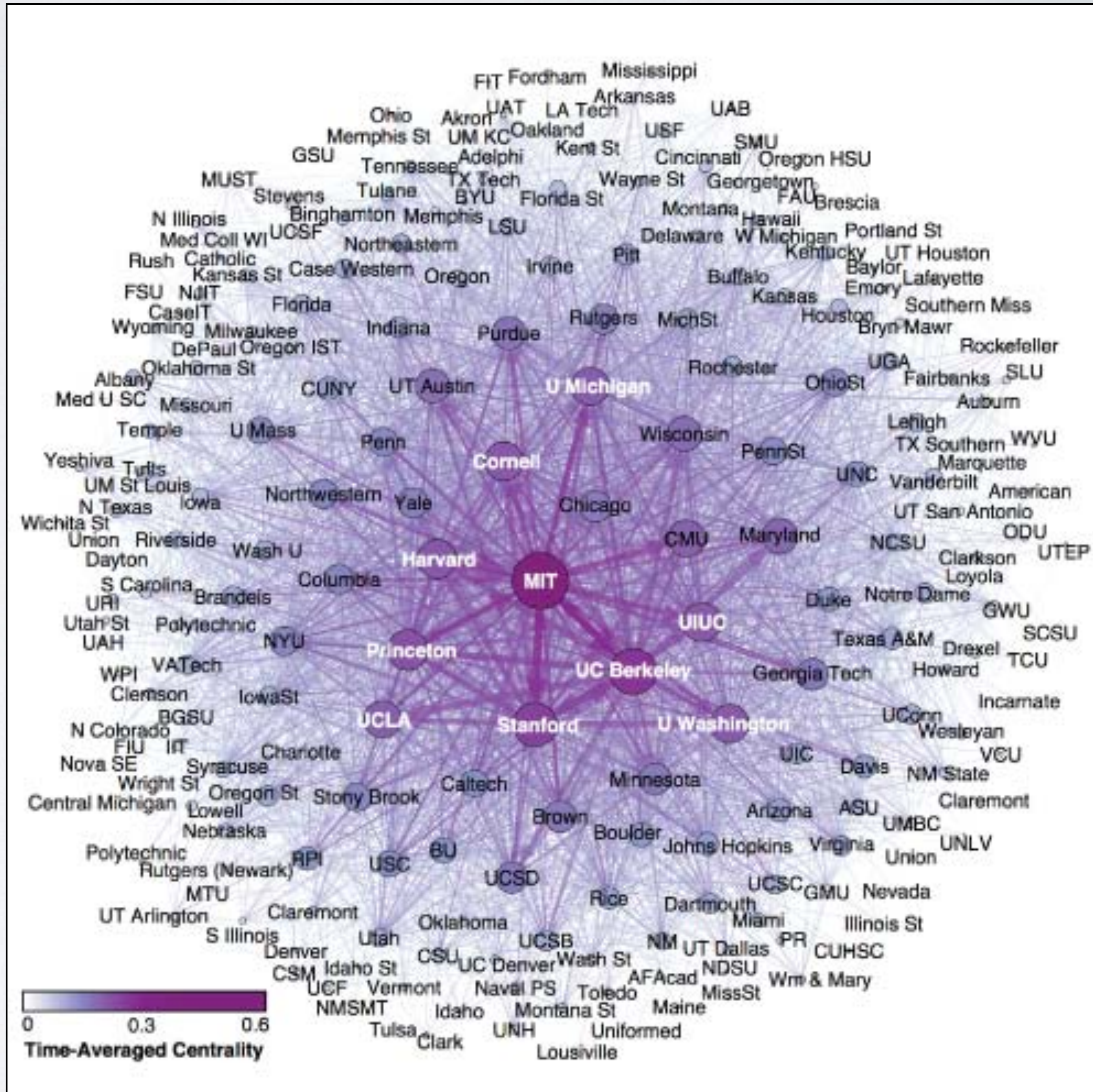
- The top actor, Clark Gable vs Groucho Marx, is too close to call using only time-average centrality
- We plot their conditional centrality trajectories for a few values of the layer coupling strength ϵ



- We find that Groucho was initially top ranked, but Clark had a longer career

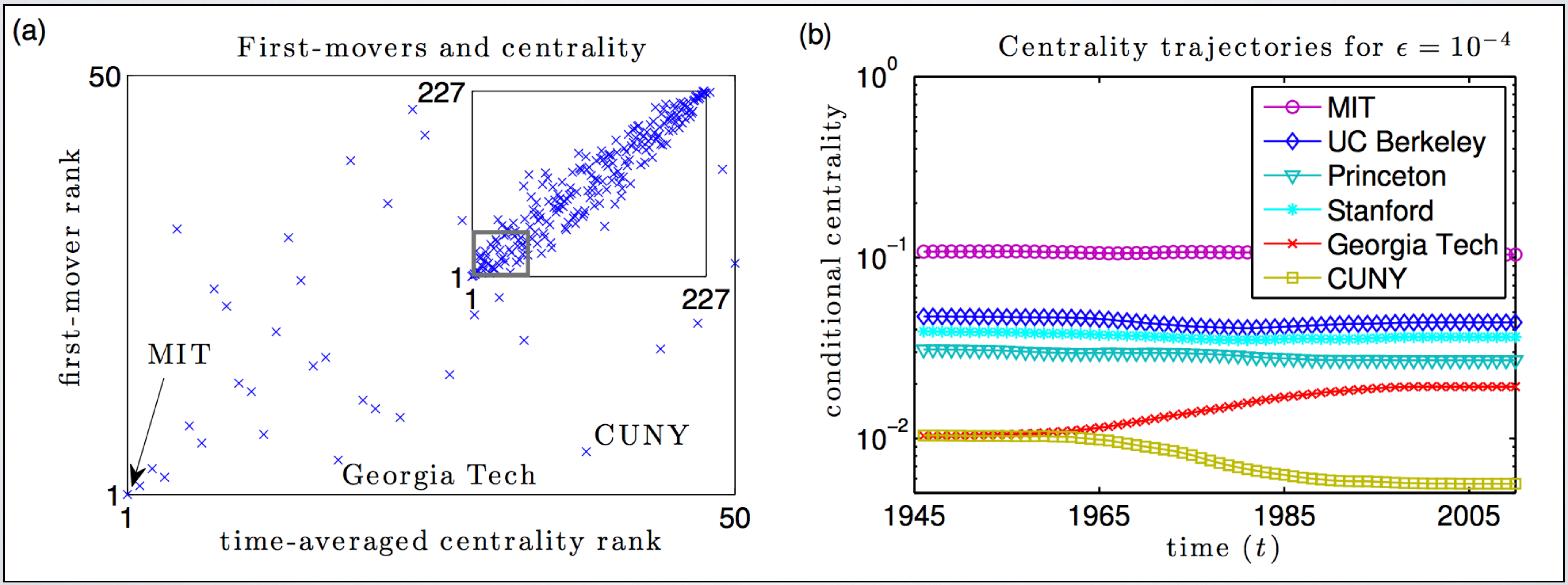
CASE STUDY 2: FLOW OF MATH PHDS

- Network encodes graduation and hiring of math PhDs using data from the Mathematical Genealogy Project
- Edges reflect the number of PhD students that graduate from university i and then later teach at university j
- Color indicates time-averaged centrality



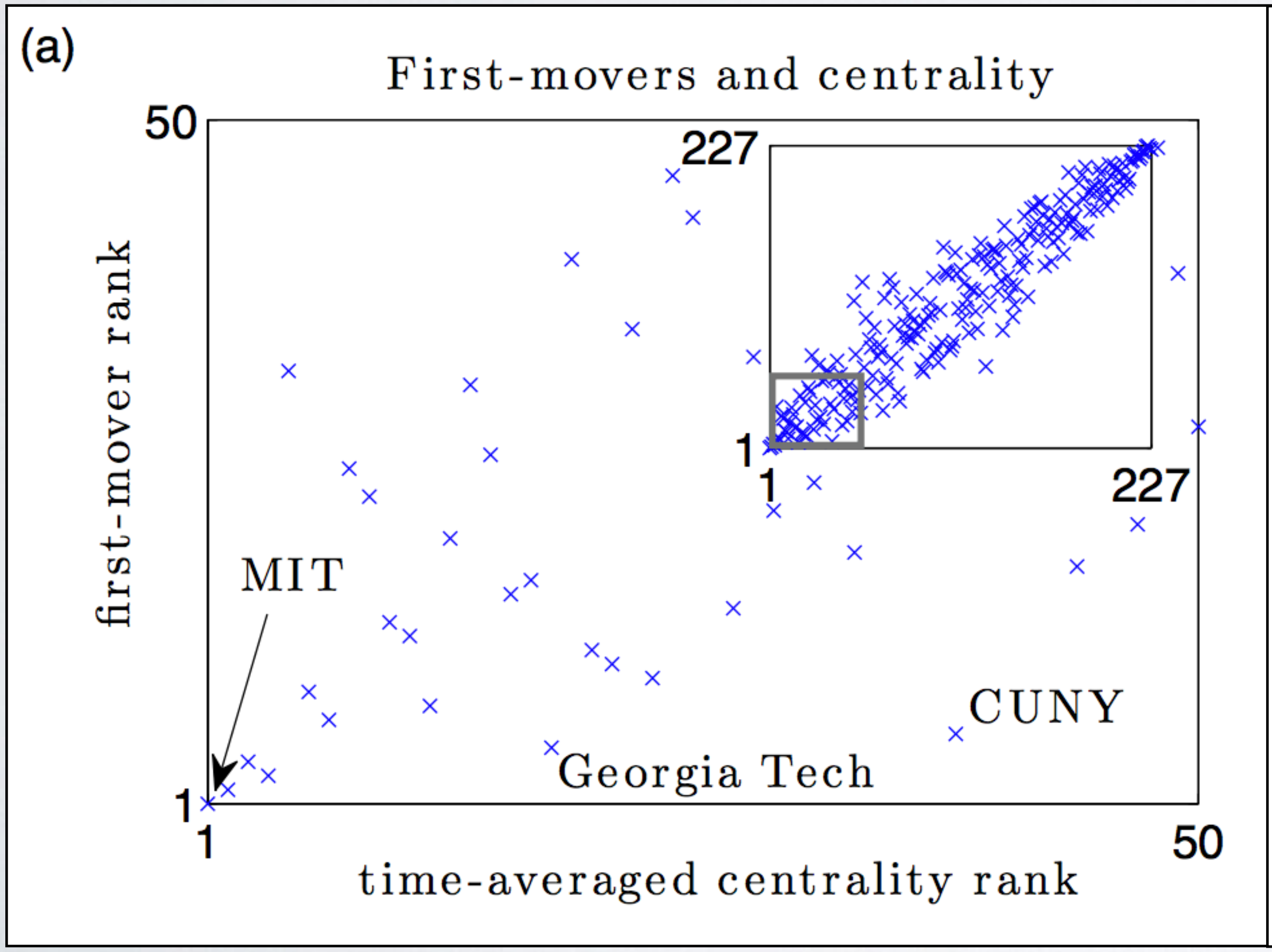
OUTLIER MATH DEPARTMENTS

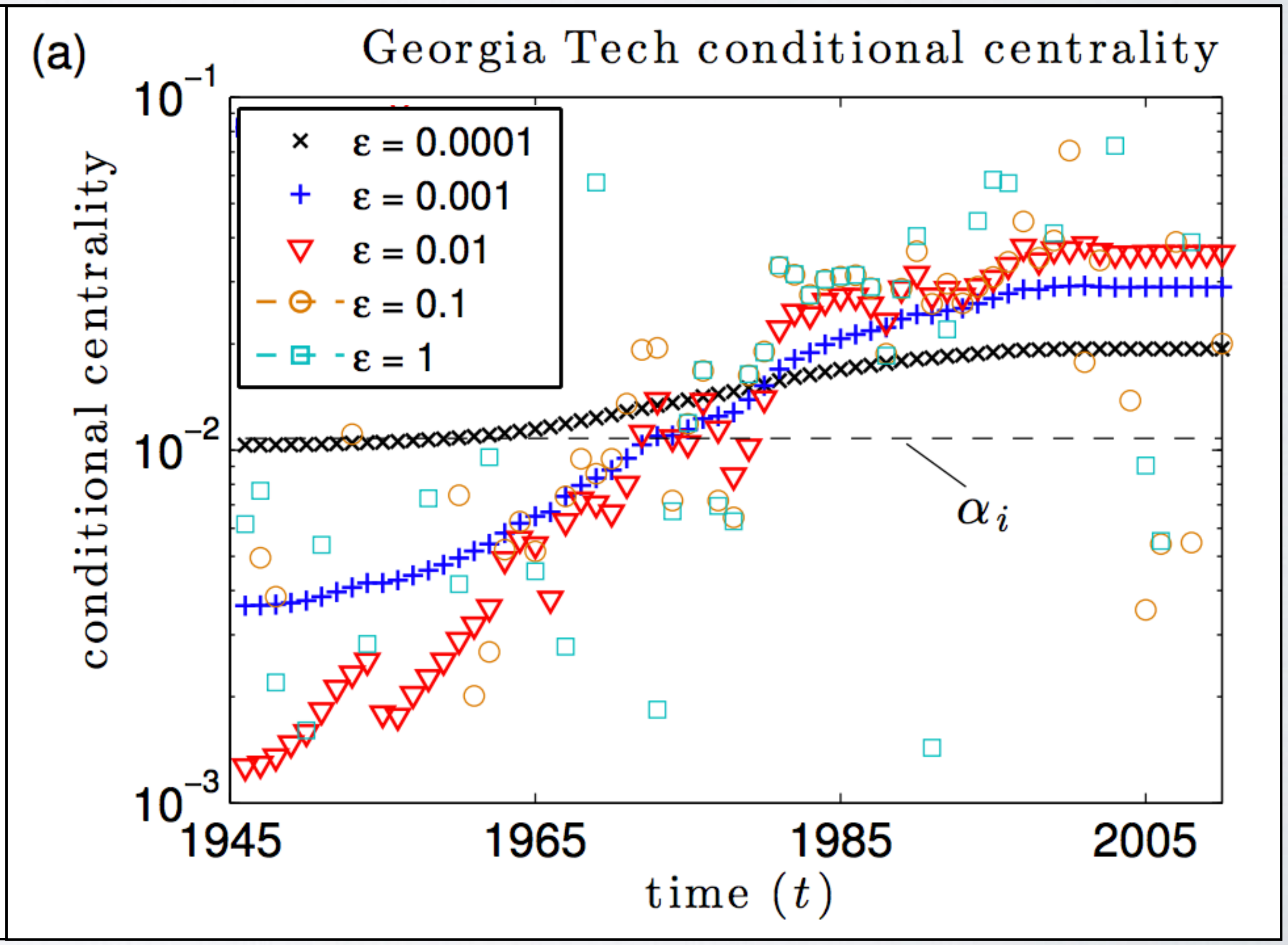
- Outliers have unusually large *first-order-mover scores* versus their *time-averaged centrality* (Georgia Tech and CUNY)



OUTLIER MATH DEPARTMENTS

- Outliers have unusually large *first-order-mover scores* versus their *time-averaged centrality* (Georgia Tech and CUNY)



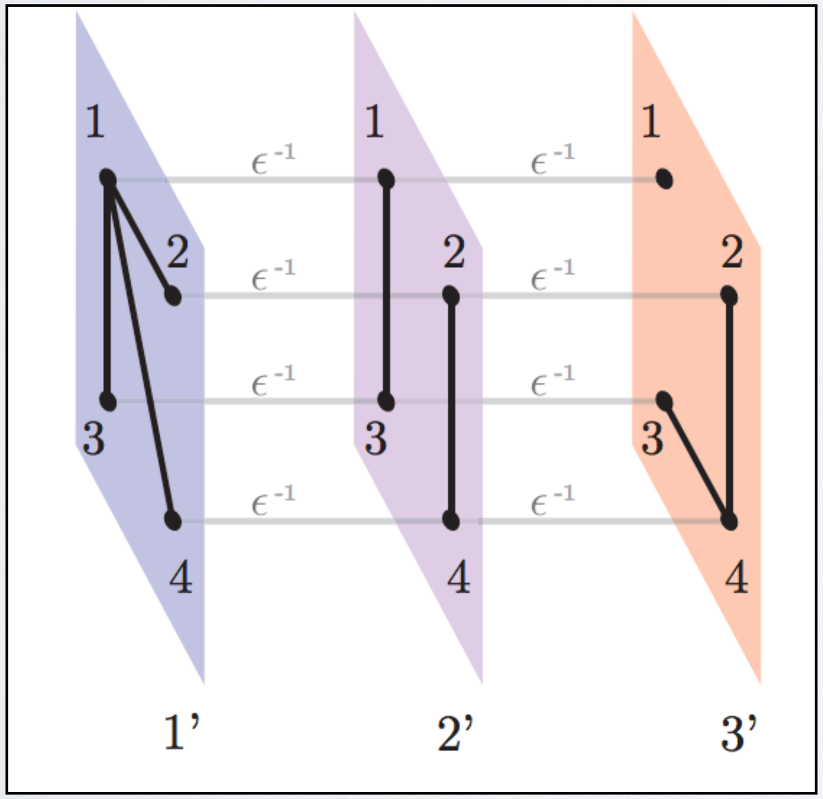


LAYER AGGREGATION

- Our singular perturbation analysis for $\mathbb{C}(\epsilon)\mathbf{v}(\epsilon) = \lambda_{\max}\mathbf{v}(\epsilon)$ reveals that

$$\begin{aligned} \mathbf{v}(\epsilon) &= \mathbf{v}_0 + \epsilon\mathbf{v}_1 + \epsilon^2\mathbf{v}_2 + \dots \\ \lambda_{\max} &= \lambda_0 + \epsilon\lambda_1 + \epsilon^2\lambda_2 + \dots \end{aligned}$$

strong coupling implements a layer aggregation: $\sum_t \mathbf{C}^{(t)} \sin^2\left(\frac{\pi t}{T+1}\right)$



$$\mathbb{C}(\epsilon) = \begin{bmatrix} \mathbf{C}^{(1)} & \epsilon^{-1}\mathbf{I} & 0 & \cdots \\ \epsilon^{-1}\mathbf{I} & \mathbf{C}^{(2)} & \epsilon^{-1}\mathbf{I} & \ddots \\ 0 & \epsilon^{-1}\mathbf{I} & \mathbf{C}^{(3)} & \ddots \\ \vdots & \ddots & \ddots & \ddots \end{bmatrix}$$

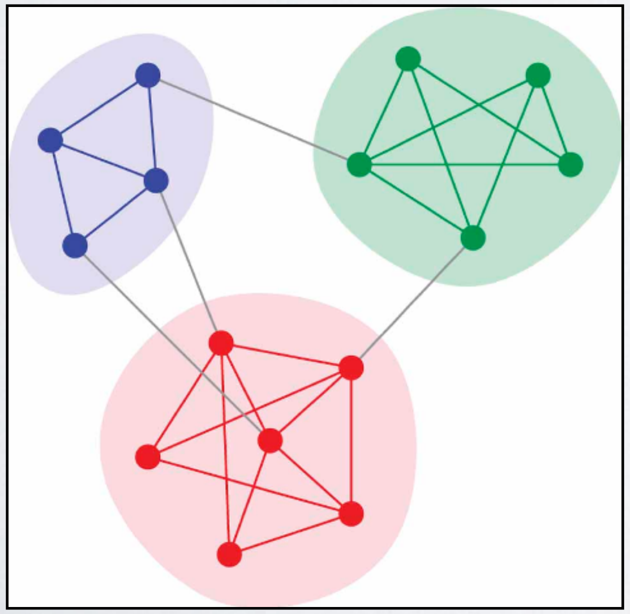
LAYER AGGREGATION

- Our singular perturbation analysis for $\mathbb{C}(\epsilon)\mathbf{v}(\epsilon) = \lambda_{\max}\mathbf{v}(\epsilon)$ reveals that

$$\left|\begin{array}{l} \mathbf{v}(\epsilon) = \mathbf{v}_0 + \epsilon\mathbf{v}_1 + \epsilon^2\mathbf{v}_2 + \ldots \\ \lambda_{\max} = \lambda_0 + \epsilon\lambda_1 + \epsilon^2\lambda_2 + \ldots \end{array}\right.$$

strong coupling implements a layer aggregation: $\sum_t \mathbf{C}^{(t)} \sin^2\left(\frac{\pi t}{T+1}\right)$

- We now consider layer aggregation and its affects on community detection

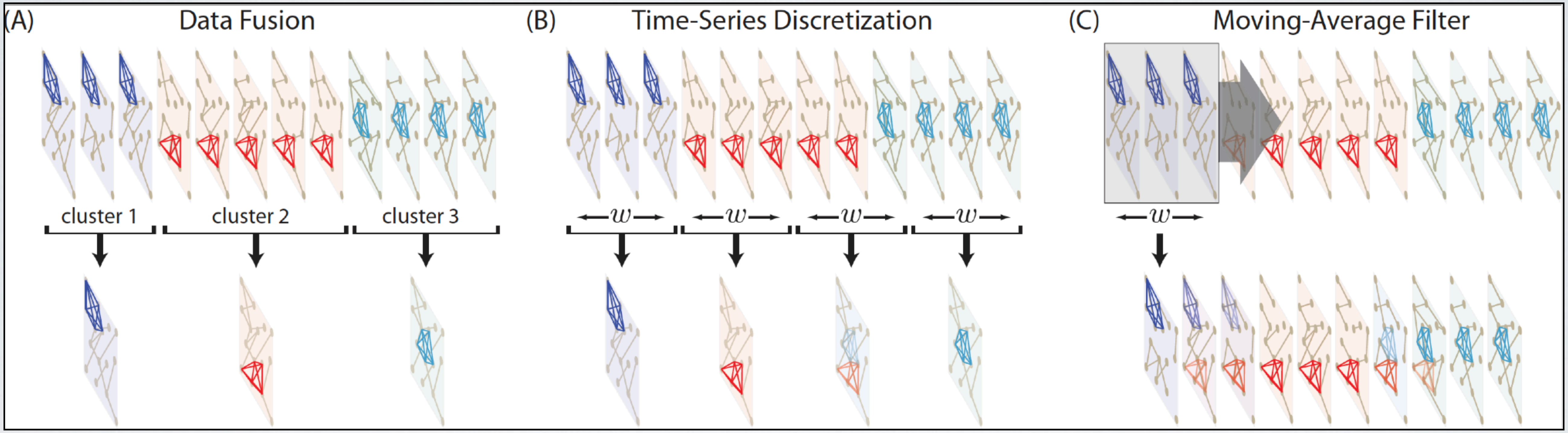


- Community detection is akin to data clustering and aims to identify groups of well-connected nodes

- **Survey of applications:**
S Shai, N Stanley, C Granell, D Taylor & PJ Mucha (2017) arXiv:1705.02305

DETECTABILITY OF COMMUNITIES IN PREPROCESSED TEMPORAL NETWORKS

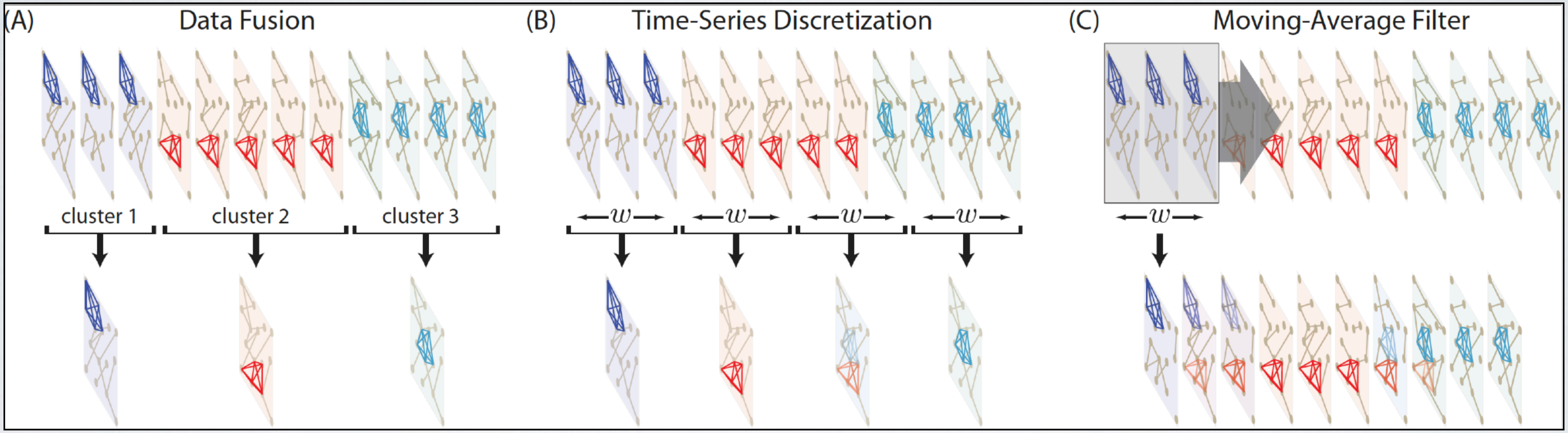
- What are the effects of <u>layer-aggregation</u> on the detectability of communities?



- Can we optimally preprocess network data to maximally enhance community detection?

DETECTABILITY OF COMMUNITIES IN PREPROCESSED TEMPORAL NETWORKS

- What are the effects of layer-aggregation on the detectability of communities?



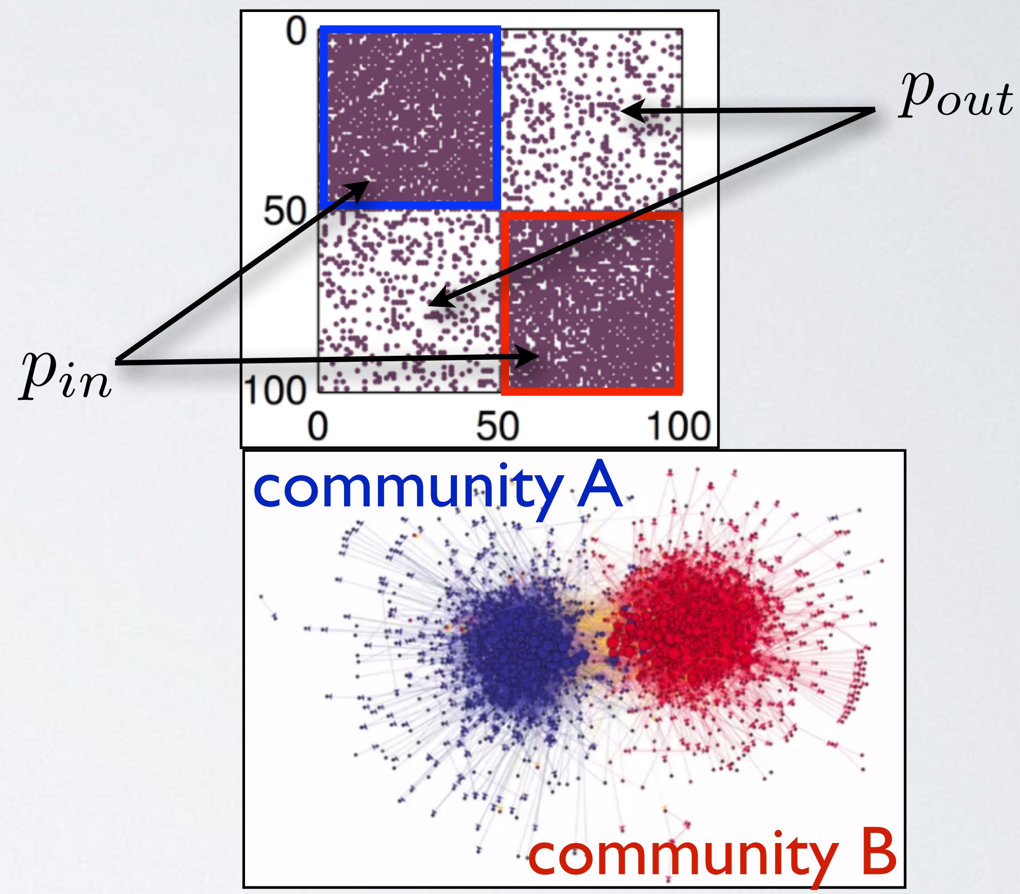
- Can we optimally preprocess network data to maximally enhance community detection?
  - We approach these questions using random-matrix theory to analyze fundamental limits for community detection.

2-COMMUNITY NETWORK MODEL

- *L* network layers are drawn from one stochastic block model (SBM)
  - Create edge (i, j) with probability

$$P_{ij} = \begin{cases} p_{in} & \text{if } c_i = c_j \\ p_{out} & \text{if } c_i \neq c_j \end{cases}$$

$$p_{in} > p_{out}$$

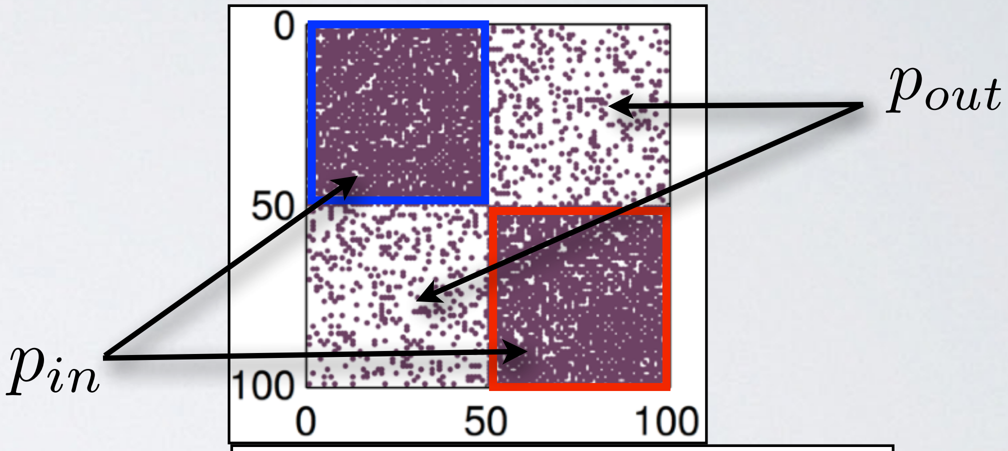


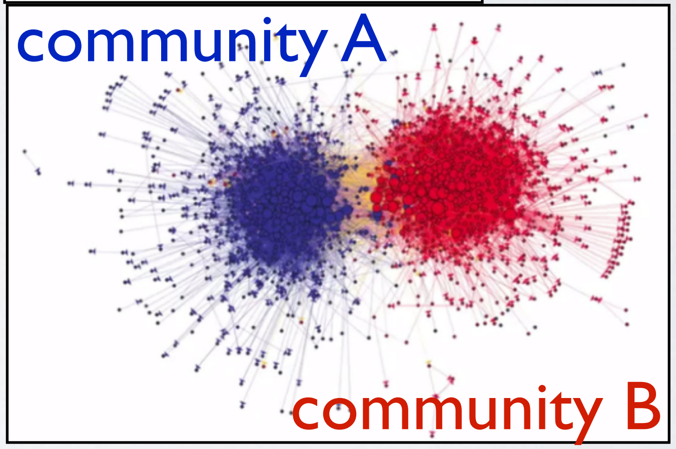
2-COMMUNITY NETWORK MODEL

- *L* network layers are drawn from one stochastic block model (SBM)
  - Create edge (i, j) with probability

$$P_{ij} = \begin{cases} p_{in} & \text{if } c_i = c_j \\ p_{out} & \text{if } c_i \neq c_j \end{cases}$$

$$p_{in} > p_{out}$$





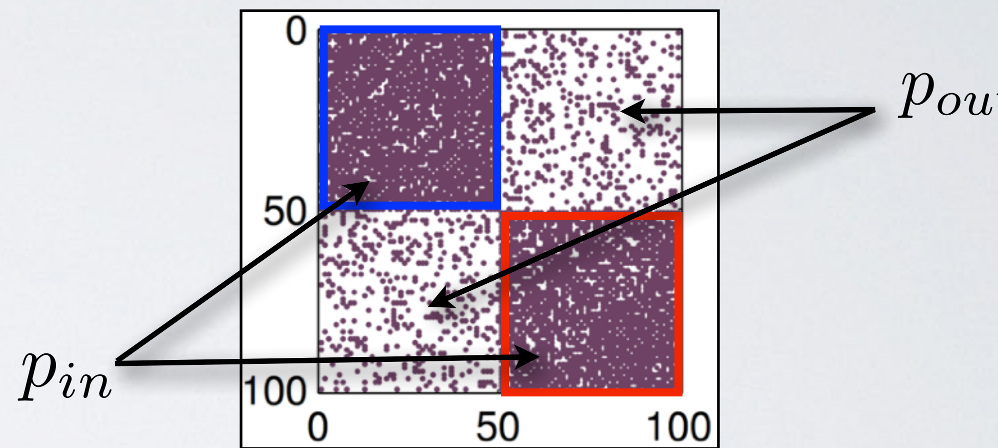
- It is helpful to study the variables
  - mean edge probability: $\rho = (p_{in} + p_{out})/2$
  - probability difference: $\Delta = p_{in} - p_{out}$

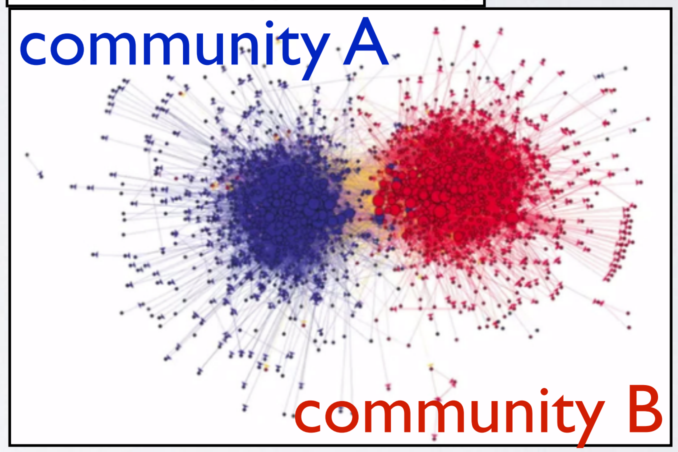
2-COMMUNITY NETWORK MODEL

- **L network layers are drawn from one stochastic block model (SBM)**
  - Create edge (i, j) with probability

$$P_{ij} = \begin{cases} p_{in} \text{ if } c_i = c_j \\ p_{out} \text{ if } c_i \neq c_j \end{cases}$$

$$p_{in} > p_{out}$$



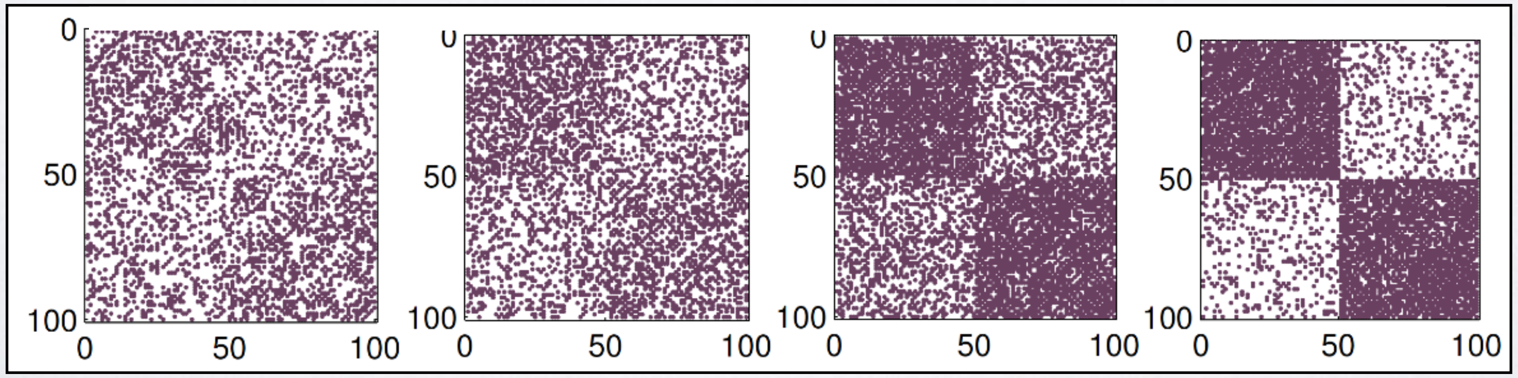


- **It is helpful to study the variables**
  - mean edge probability: $\rho = (p_{in} + p_{out})/2$
  - probability difference: $\Delta = p_{in} - p_{out}$

- **Detectability phase transition at $\Delta^* > 0$**

undetectable

small Δ



detectable

large Δ

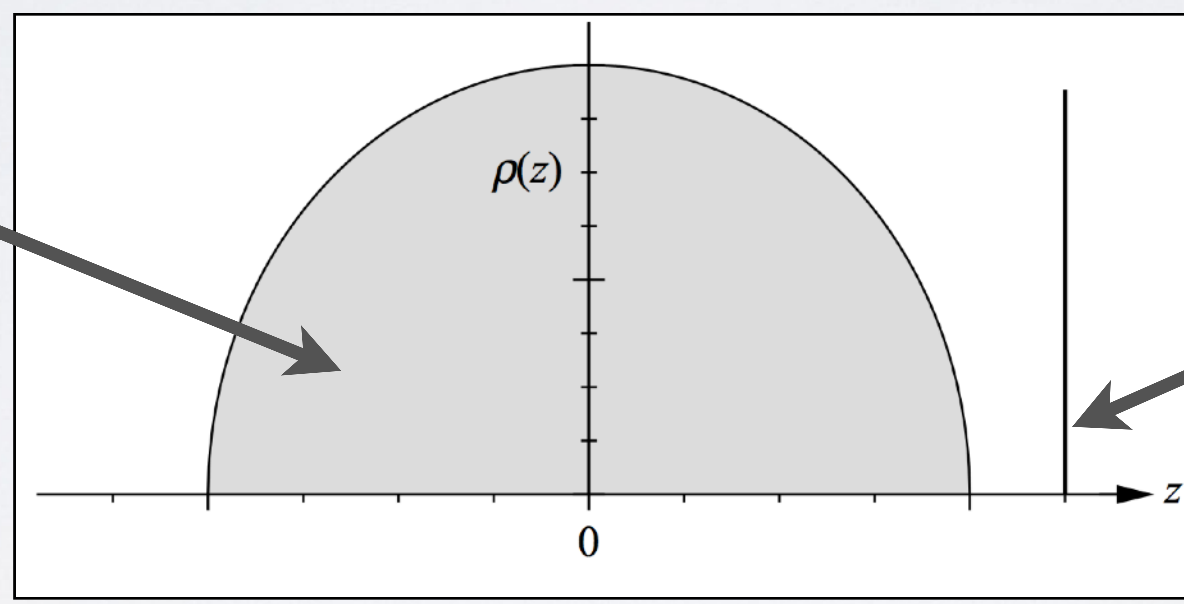
ANALYSIS WITH RANDOM-MATRIX THEORY

- **We develop random matrix theory for the modularity matrix based on** Nadakuditi & Newman, PRL 2012 **and** T Peixoto, PRL 2013
- **We analyze the distribution of eigenvalues for the modularity matrix in the large *N* limit**

ANALYSIS WITH RANDOM-MATRIX THEORY

- **We develop <u>random matrix theory</u> for the <u>modularity matrix</u> based on** Nadakuditi & Newman, PRL 2012 **and** T Peixoto, PRL 2013
- **We analyze the <u>distribution of eigenvalues</u> for the modularity matrix in the large *N* limit**

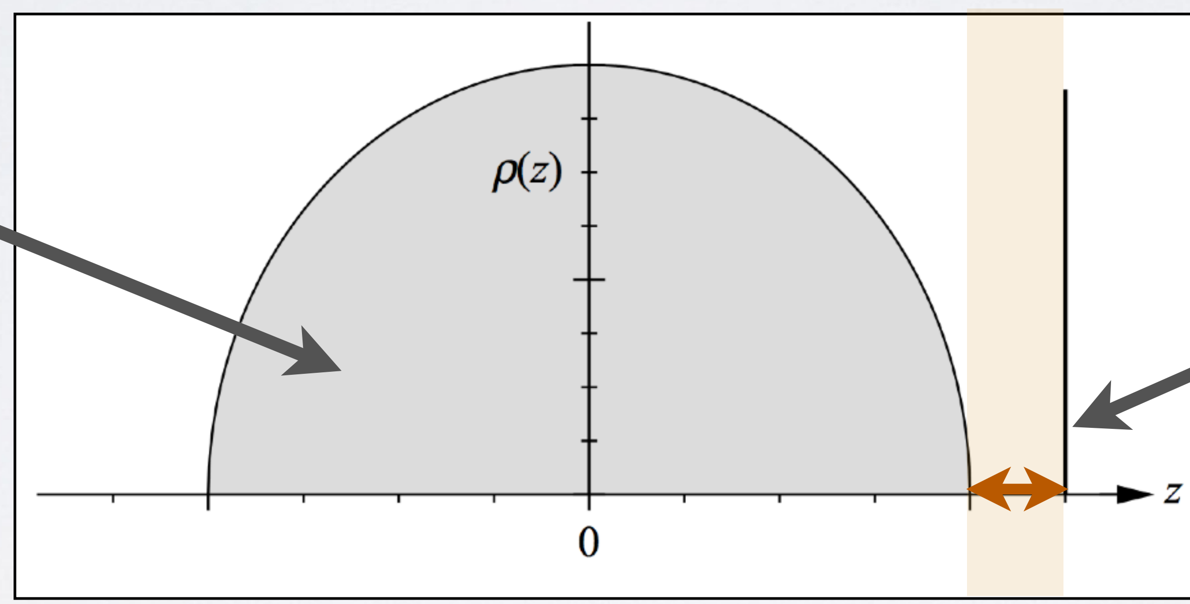
bulk eigenvalues



isolated eigenvalue whose eigenvector encodes community structure

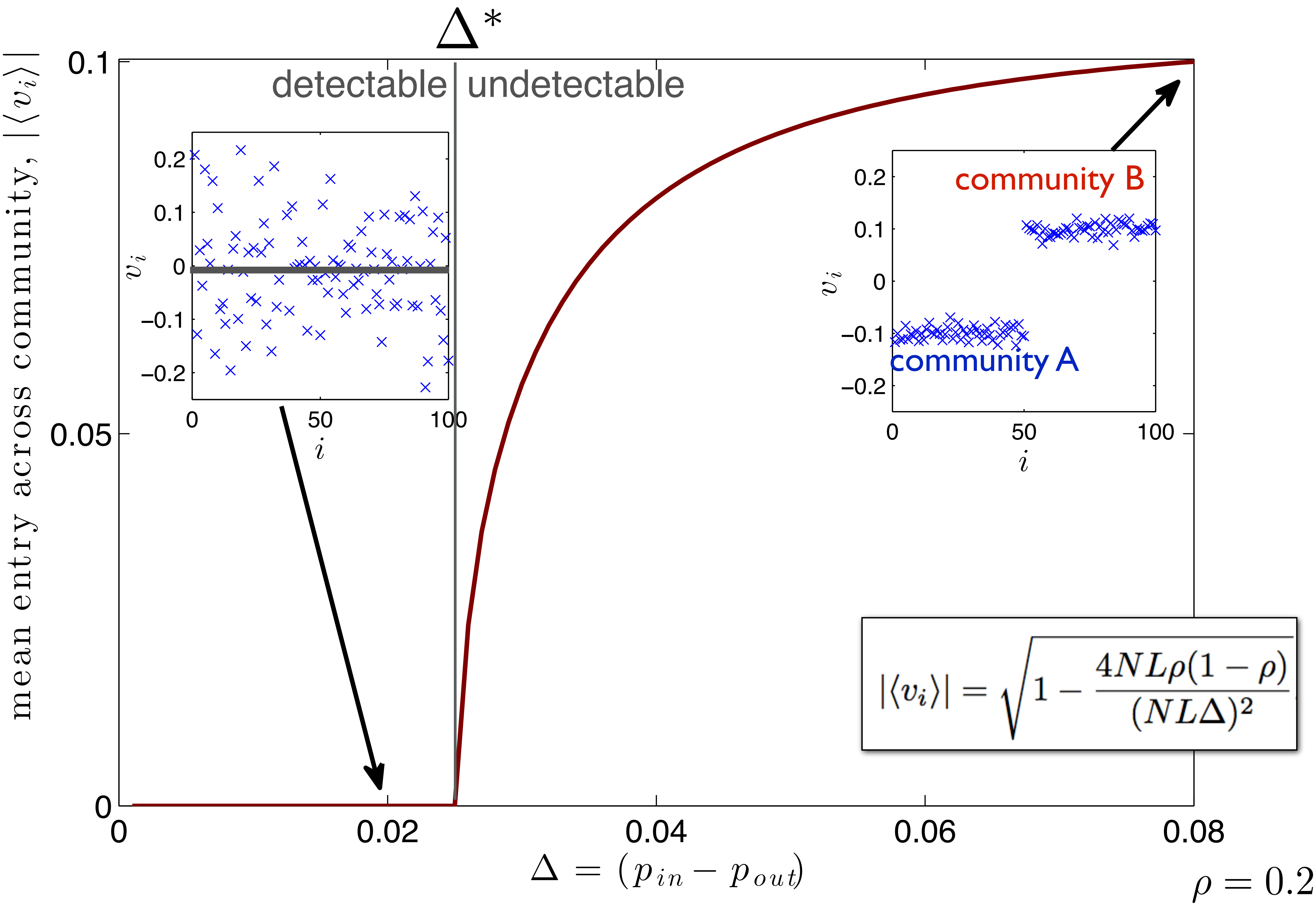
ANALYSIS WITH RANDOM-MATRIX THEORY

- We develop random matrix theory for the modularity matrix based on Nadakuditi & Newman, PRL 2012 **and** T Peixoto, PRL 2013
- We analyze the distribution of eigenvalues for the modularity matrix in the large *N* limit

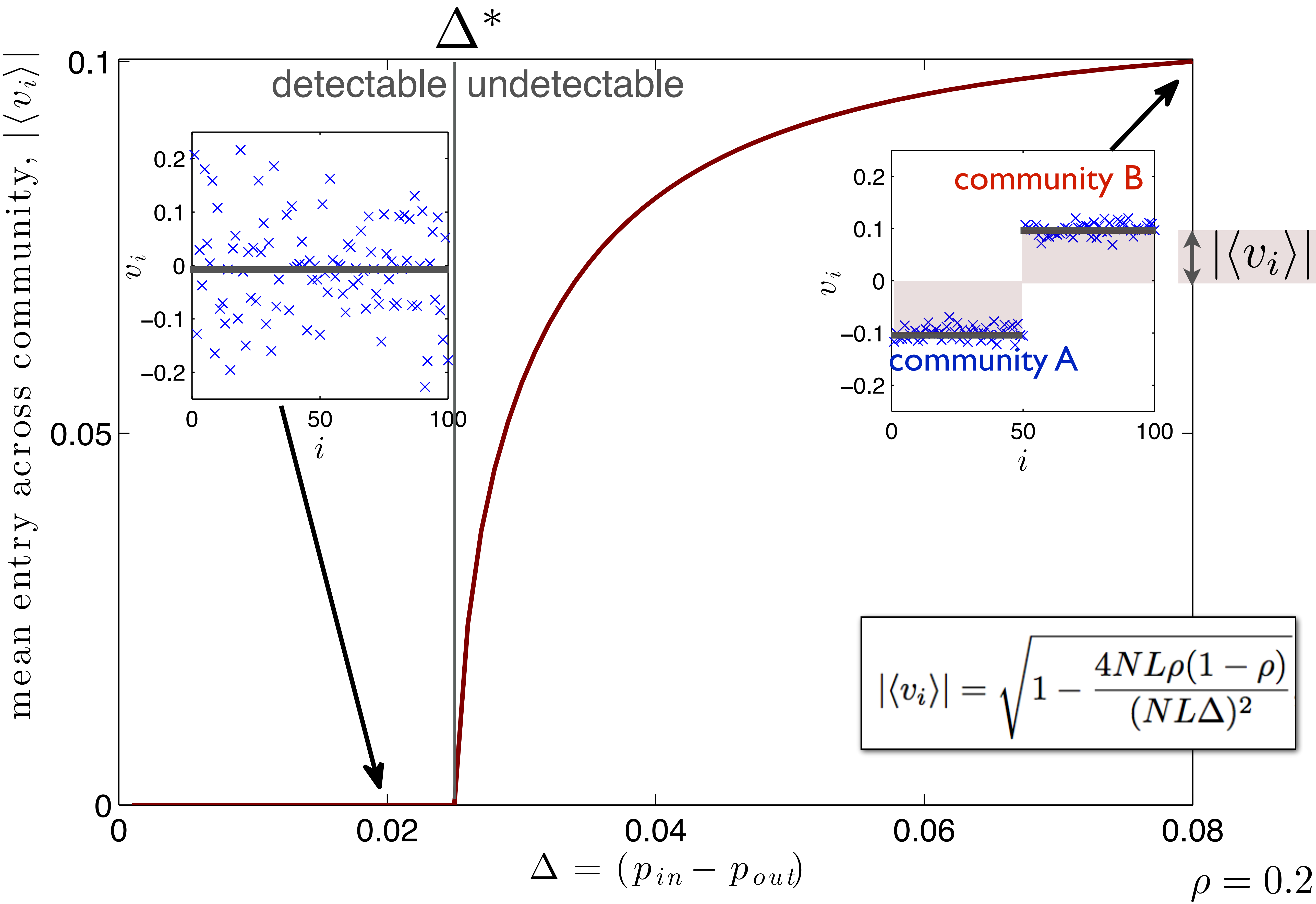


- As the gap disappears, the dominant eigenvector v becomes a random vector

PHASE TRANSITION FOR THE DOMINANT EIGENVECTOR

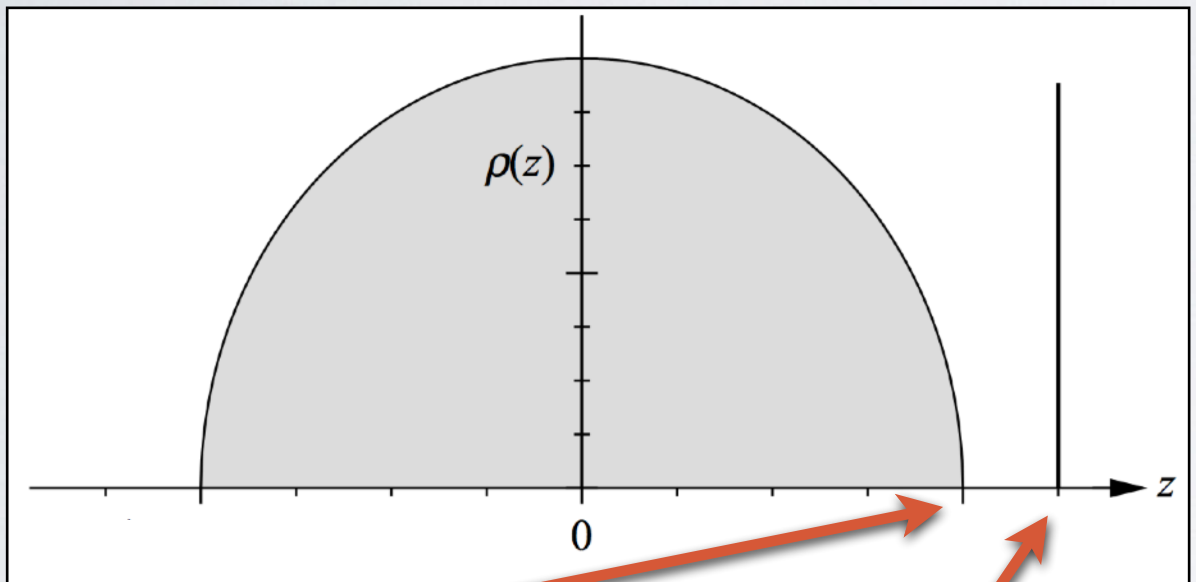


PHASE TRANSITION FOR THE DOMINANT EIGENVECTOR



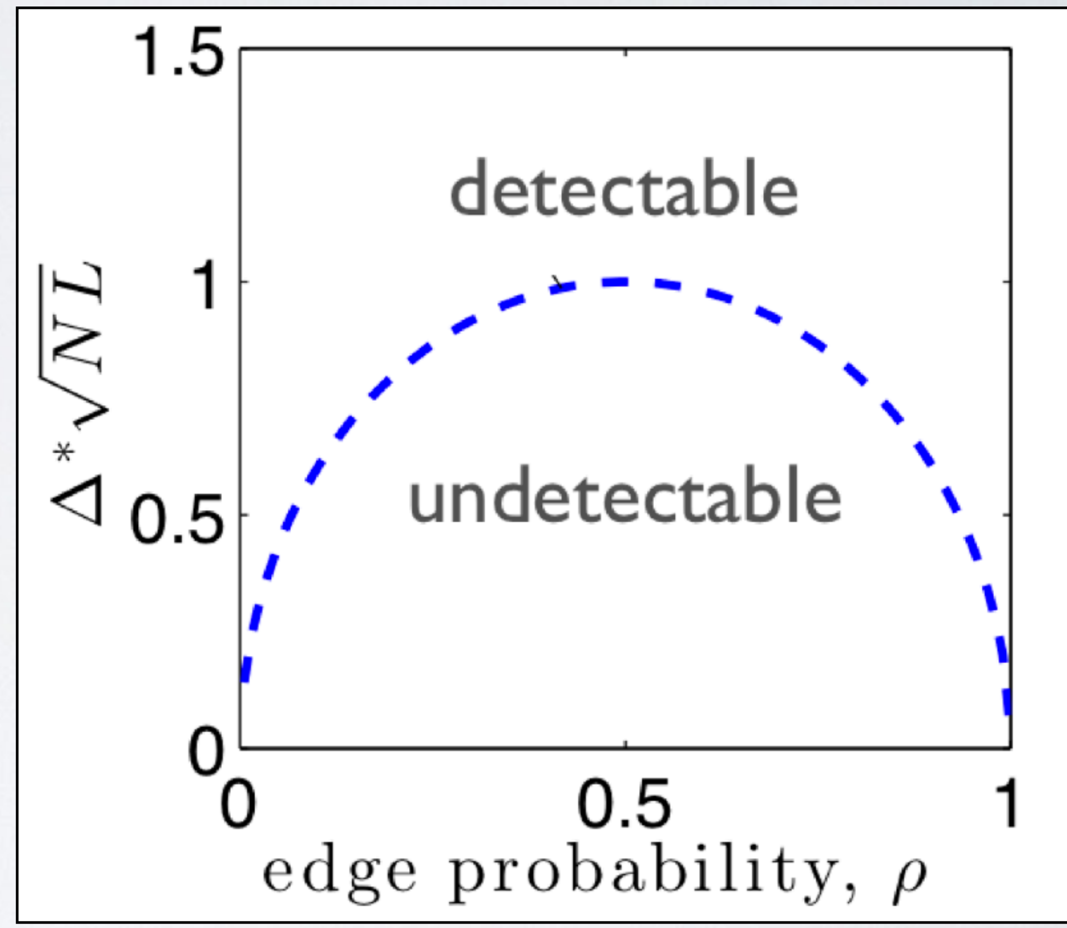
DETECTABILITY LIMIT VANISHES WITH INCREASING NUMBER OF LAYERS

Detectability equation: $\Delta^* = \sqrt{\rho(1-\rho)/NL}$



$$\lambda_2 = \sqrt{4NL\rho(1-\rho)}$$

$$\lambda_1 = NL\Delta/2 + 2\rho(1-\rho)/\Delta$$



Detectability limit vanishes as $\mathcal{O}(1/\sqrt{L})$

CONCLUSION

- **Part 1: layer-coupling and centrality**
  - While myriad real-world networks are multilayer, network analyses have traditionally been developed for single-layer networks
  - We extended eigenvector centrality to temporal networks by coupling together centrality matrices
  - We introduced concepts of joint, marginal and conditional centrality, allowing us to study centrality "trajectories" across time
  - The strong coupling limit recovers layer aggregation as a weighted sum of the layers adjacency matrices

CONCLUSION

- **Part 1: layer-coupling and centrality**
  - While myriad real-world networks are multilayer, network analyses have traditionally been developed for single-layer networks
  - We extended eigenvector centrality to temporal networks by coupling together centrality matrices
  - We introduced concepts of joint, marginal and conditional centrality, allowing us to study centrality "trajectories" across time
  - The strong coupling limit recovers layer aggregation as a weighted sum of the layers adjacency matrices

- **Part 2: layer aggregation and community detection**
  - Aggregating together similar layers can enhance the detection of structural patterns such as communities
  - We are developing random-matrix theory to study this phenomenon, which includes studying fundamental limits for community detection
  - We obtain $\mathcal{O}(1/\sqrt{L})$ as the scaling behavior for how the detectability limit vanishes with increasing number of layers L

ACKNOWLEDGMENTS

- Papers
  - D Taylor, SA Meyers, A Clauset, MA Porter & PJ Mucha (2017) Eigenvector-based centrality measures for temporal networks. Multiscale Modeling and Simulation, 15(1), 537-574.
  - D Taylor, RS Caceres & PJ Mucha (2017) Super-resolution community detection in layer-aggregated multilayer networks. Physical Review X:1609.04376.
  - D Taylor, S Shai, N Stanley & PJ Mucha (2016) Enhanced detectability of community structure in multilayer networks through layer aggregation. Physical Review Letters 116, 228301.
  - N Stanley, S Shai, D Taylor & PJ Mucha (2016) Clustering network layers with the strata multilayer stochastic block model. IEEE Transactions on Network Science 3, 95-105.

- Funding
  - NIH Award Number RO1HD075712
  - NSF Grant Number DMS1127914
  - James S. McDonnell Foundation Award 220020315