Funded by NSF

Visual Analytics for discovering group structures in networks

Takashi Nishikawa[1,2]

Adilson E. Motter[2,3]

[1]Department of Mathematics, Clarkson University

[2]Department of Physics & Astronomy, Northwestern University

[3]Northwestern Institute on Complex Systems (NICO), Northwestern University

LANL seminar March 16, 2010

Exploratory Analysis in Networks

Most network analyses try to detect a given type of structure, such as

- Small-world
  Watts & Strogatz, Nature **393**, 440 (1998), ...
- Scale-free
  Barabási & Albert, Science **286**, 509 (1999), ...
- Community structure
  Newman & Girvan, PRE **69**, 026113 (2004),
  Porter, Onnela, & Mucha, Notices Amer Math Soc **56**, 1082 (2009),
  ....

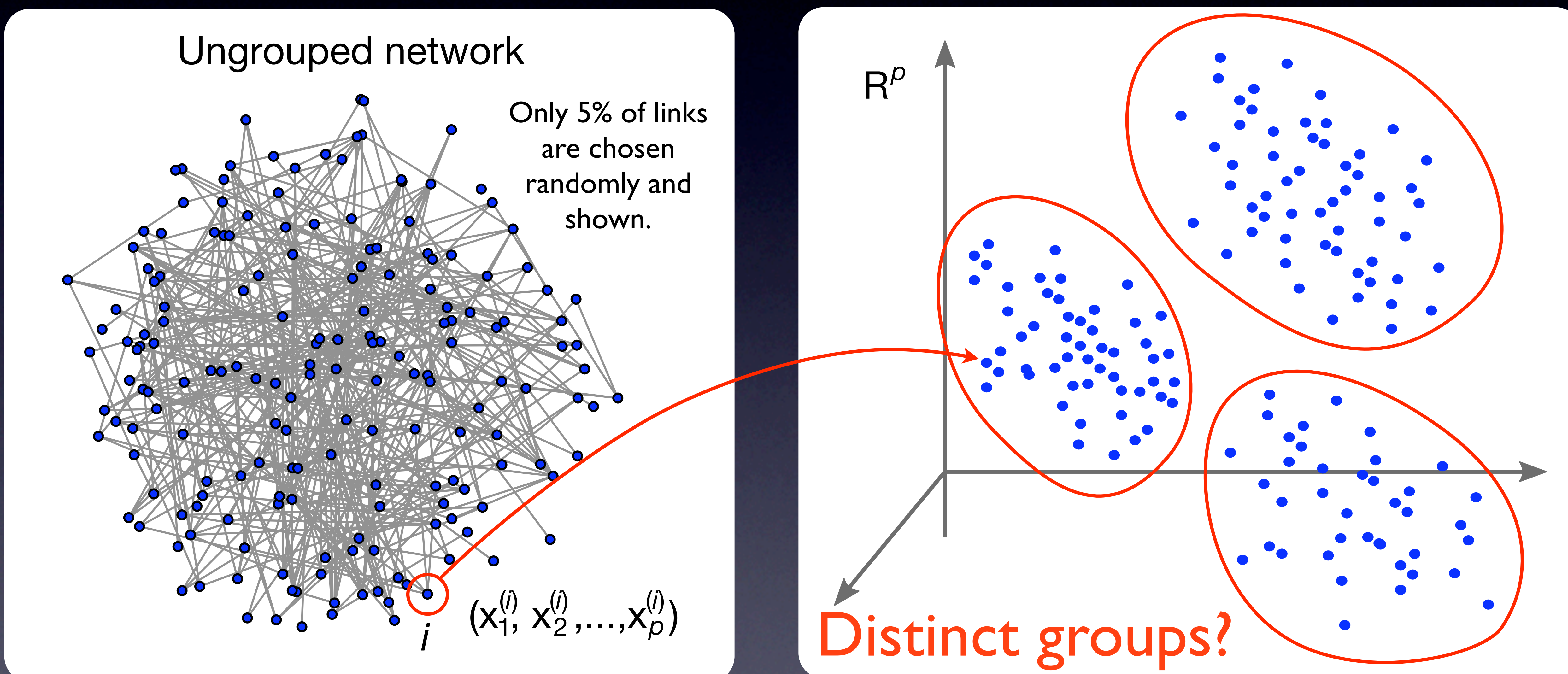
Exploratory Analysis in Networks

- An exploratory analysis tries to discover an unknown structure.
  Newman & Leicht, PNAS **104**, 9564 (2007)
- We take visual analytics approach for this problem.

Visual Analytics

- Visual interaction between the human user and the analytical tools for analyzing and interpreting high-dimensional complex data
- We take this approach for discovering an unknown group structure among nodes in a network data set.

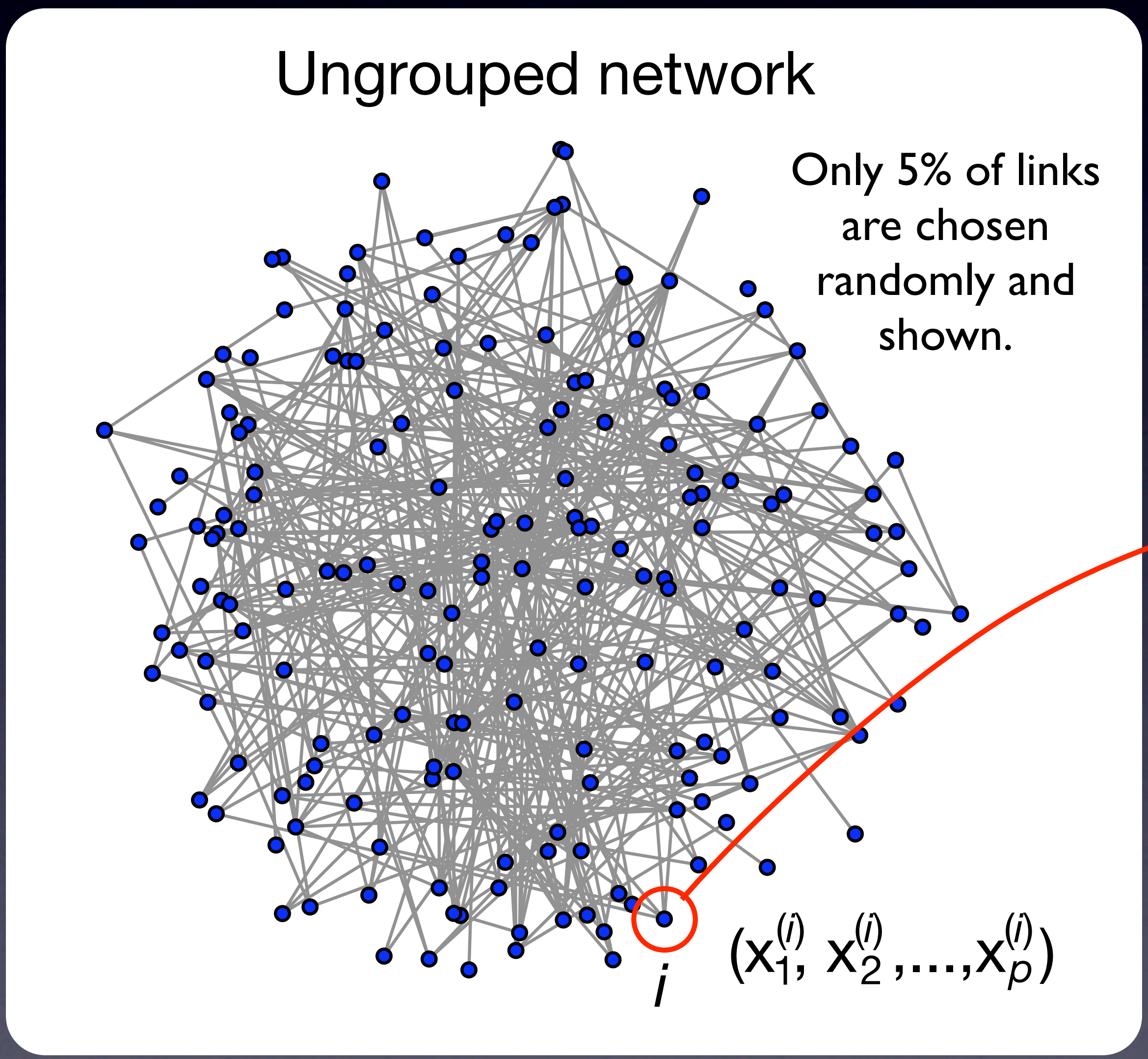
Discovering Unknown Group Structures

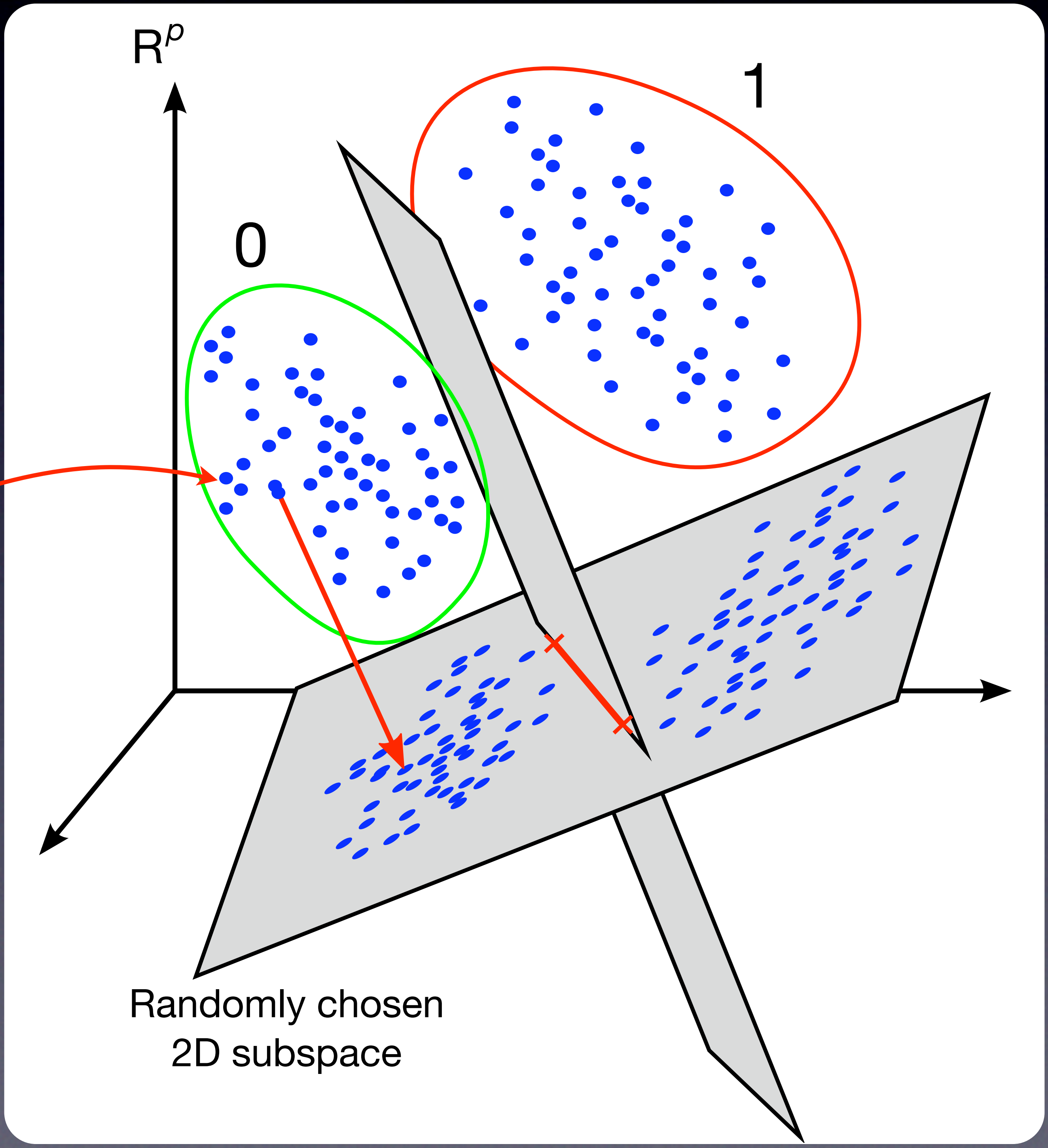


Degree
Neighbor's mean degree
2nd neighbor's mean degree
Clustering coefficient

Betweenness centrality
Laplacian eigenvectors
Normalized Laplacian eigenvectors
Mean shortest path

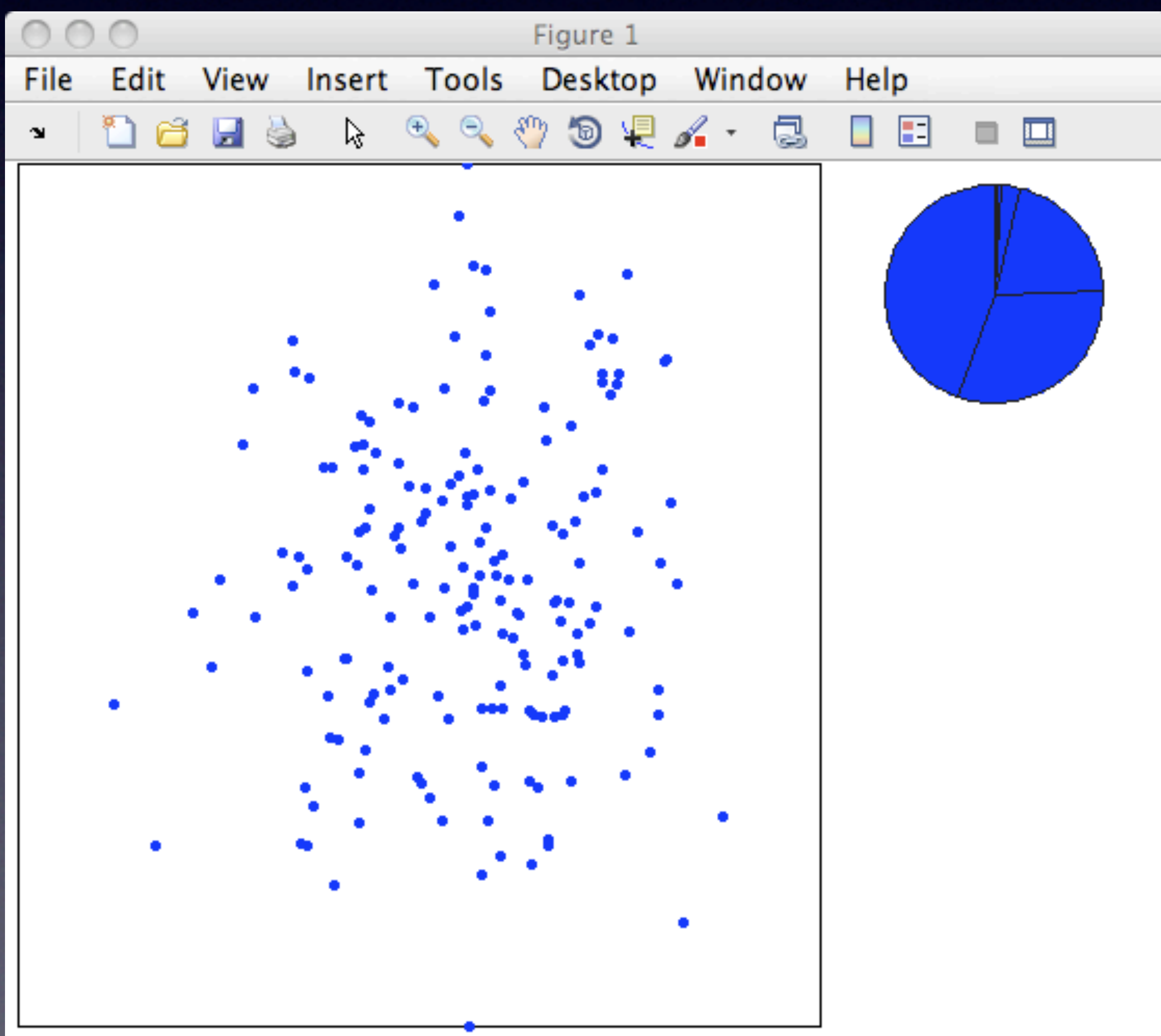
How it works: Random 2D Projection



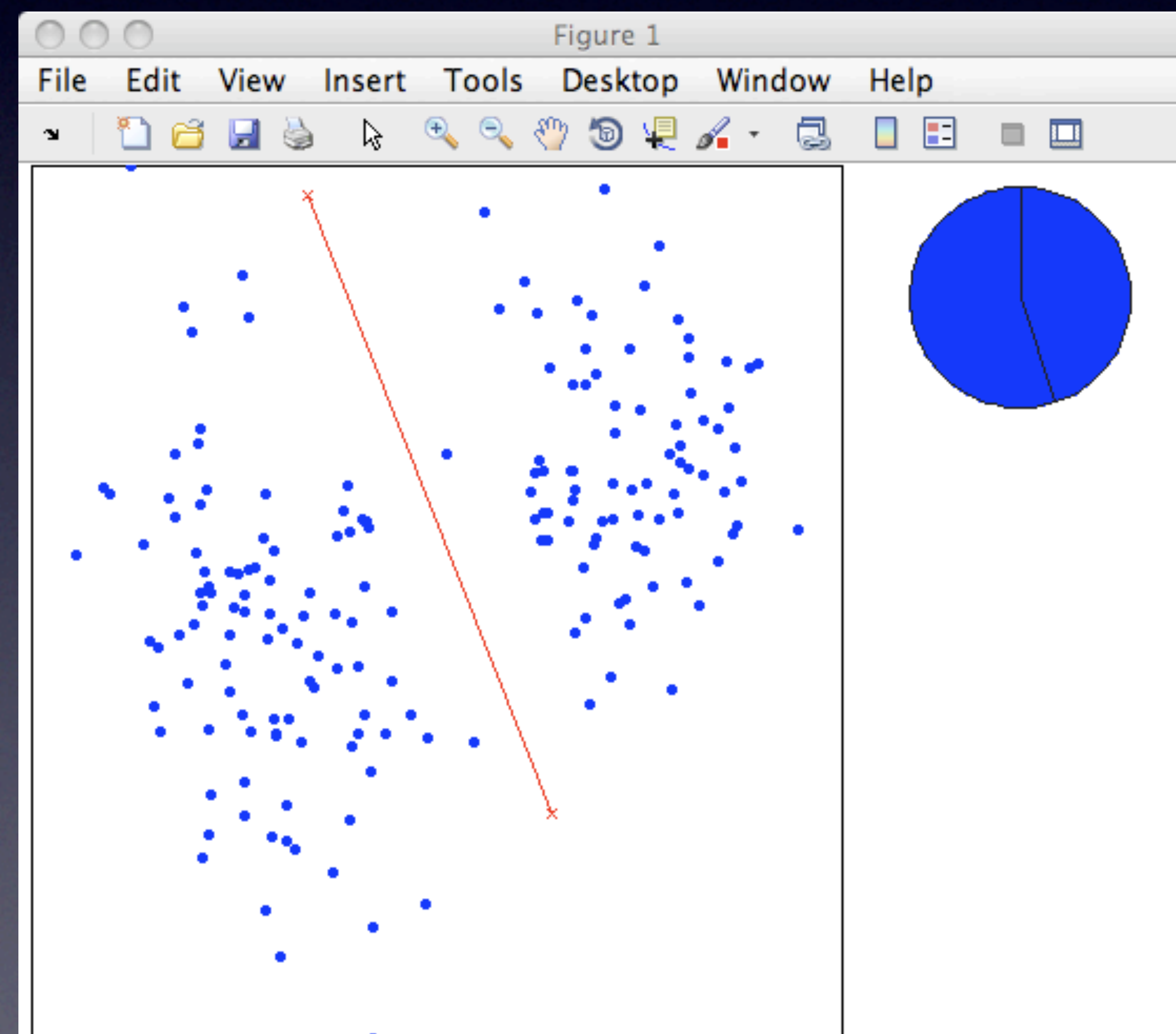


How it works: Visual Interactions

Reject

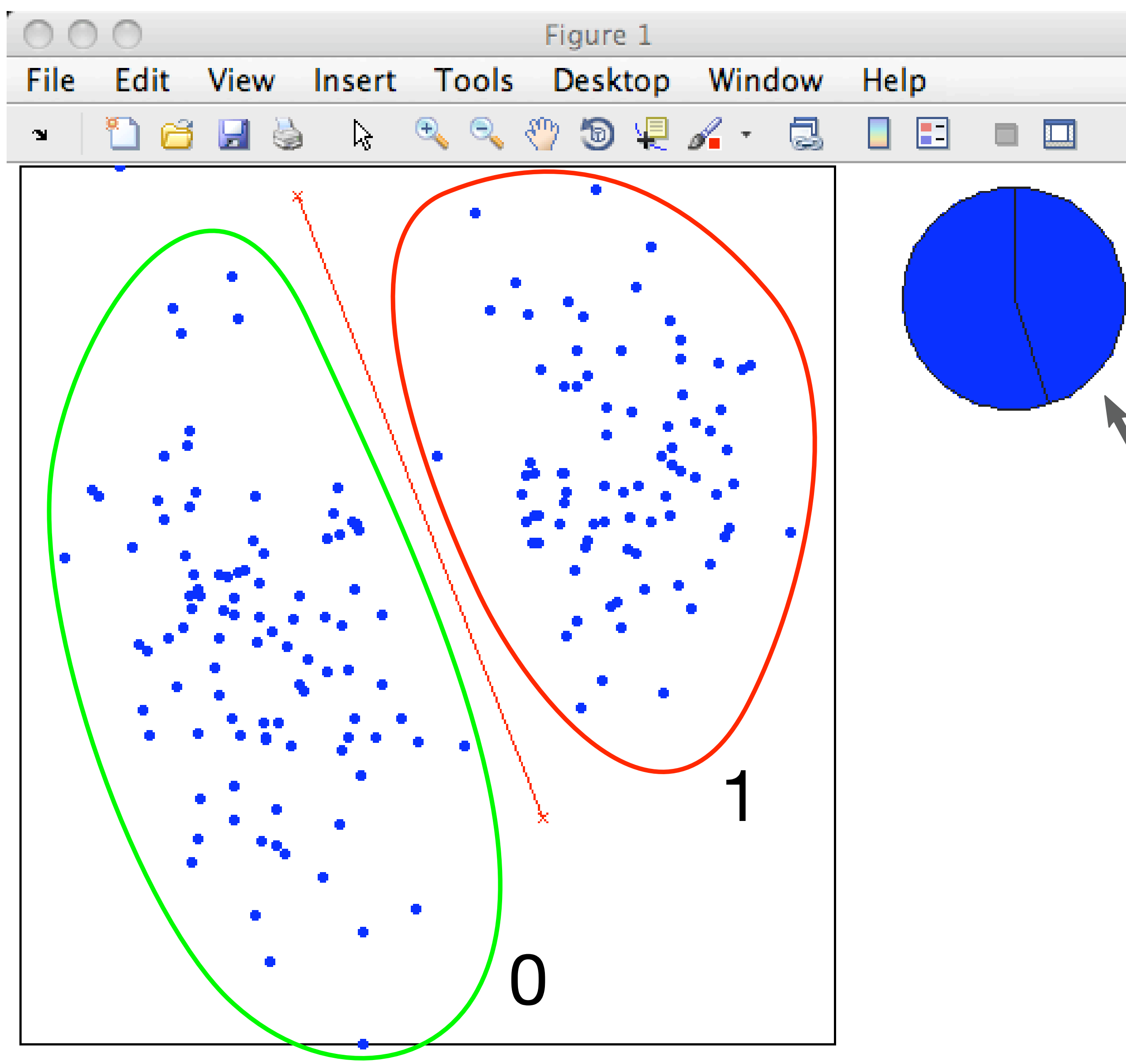


Divide



How it works: Patterns of User Inputs

Divide

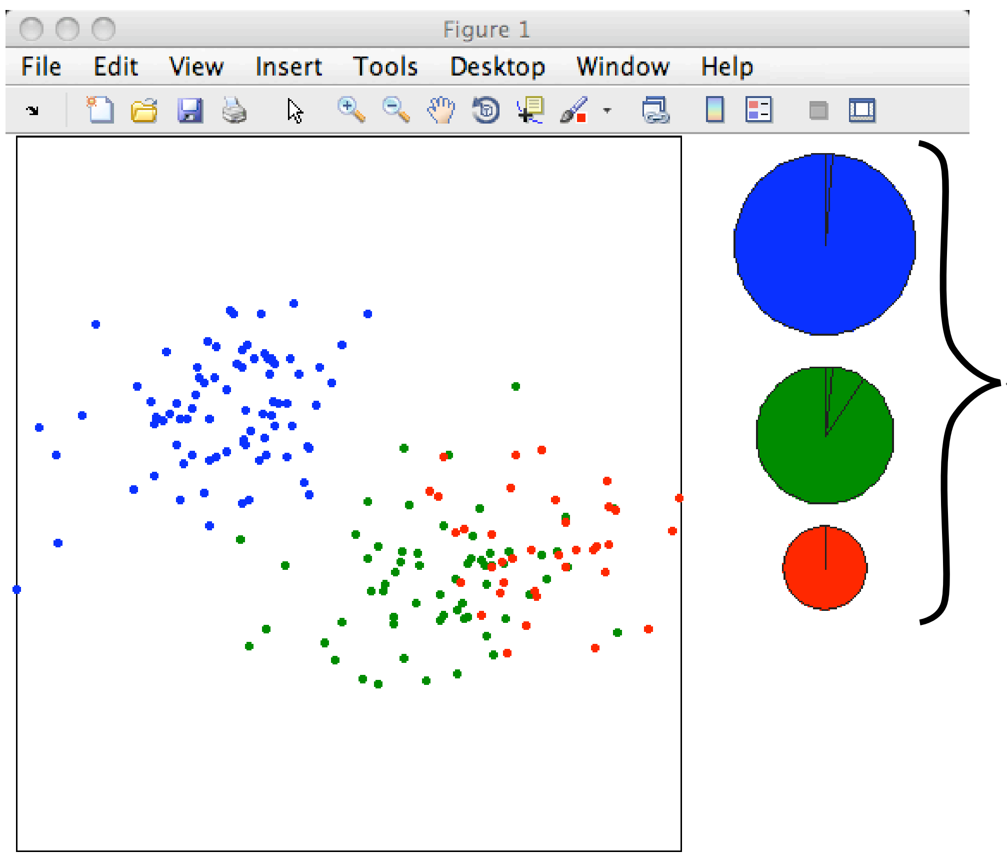


Binary patterns of divisions

Data point	Projection #1	#2	#3	#4
#1	0	0	0	0
#2	0	1	0	0
#3	0	1	0	1
#4	0	0	0	0
#5	1	1	1	1
#6	1	1	0	1
#7	1	1	0	0
#8	1	1	1	1
#9	1	1	0	1
⋮	⋮	⋮	⋮	⋮

How it works: Grouping Binary Patterns

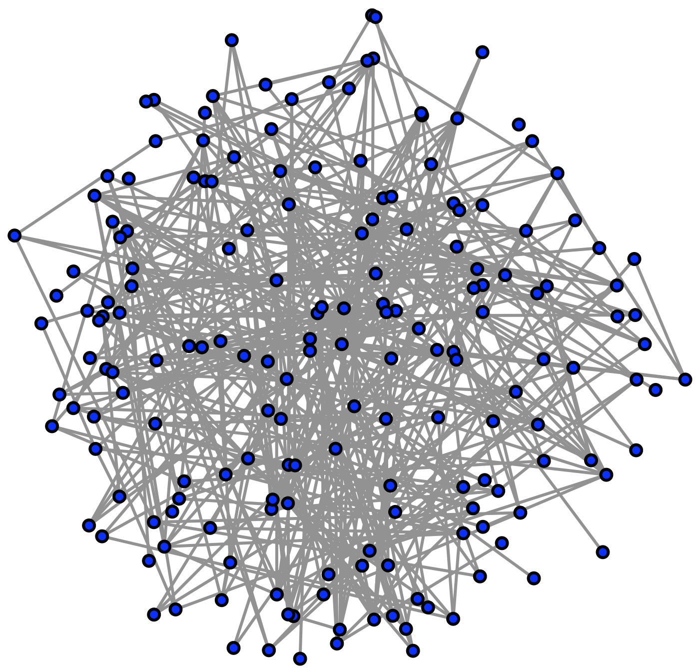
Choose number of groups



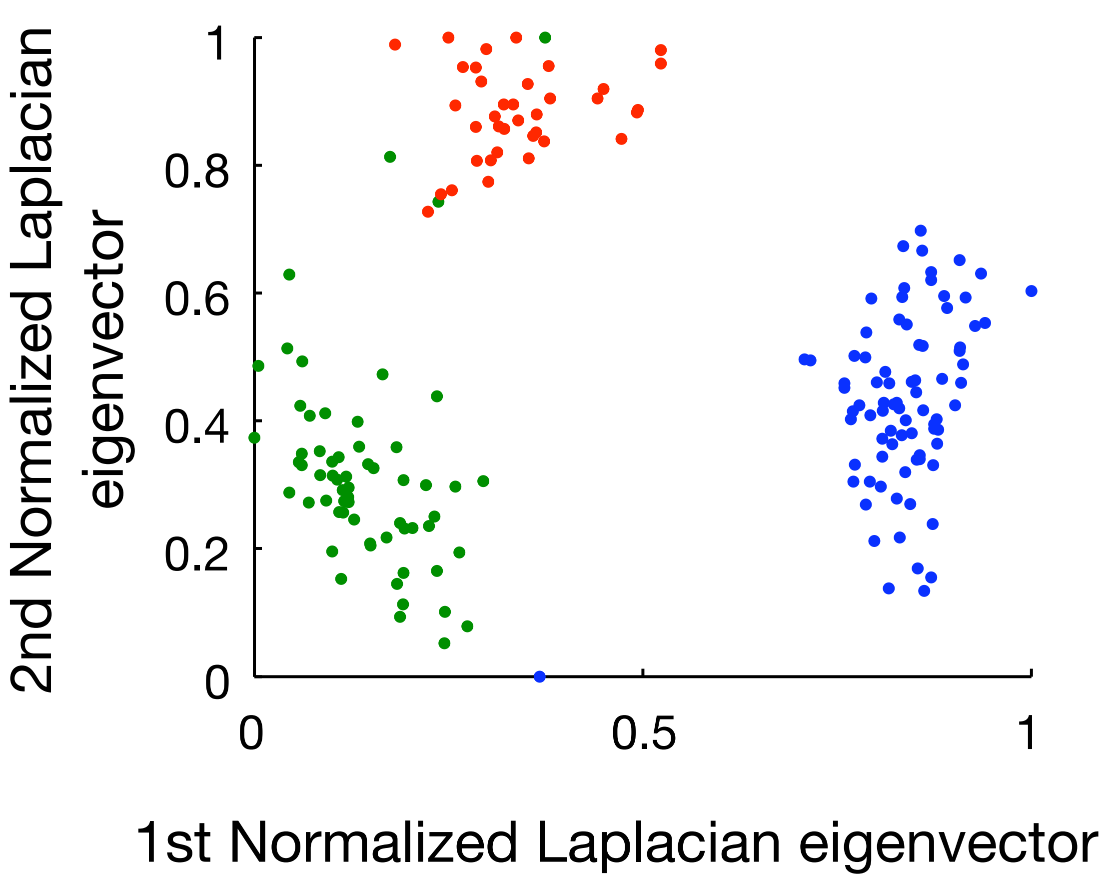
Grouping of binary pattern

Group	Principal pattern	Other patterns
#1	0000	0100
#2	1111	N/A
#3	1101	1100, 0101

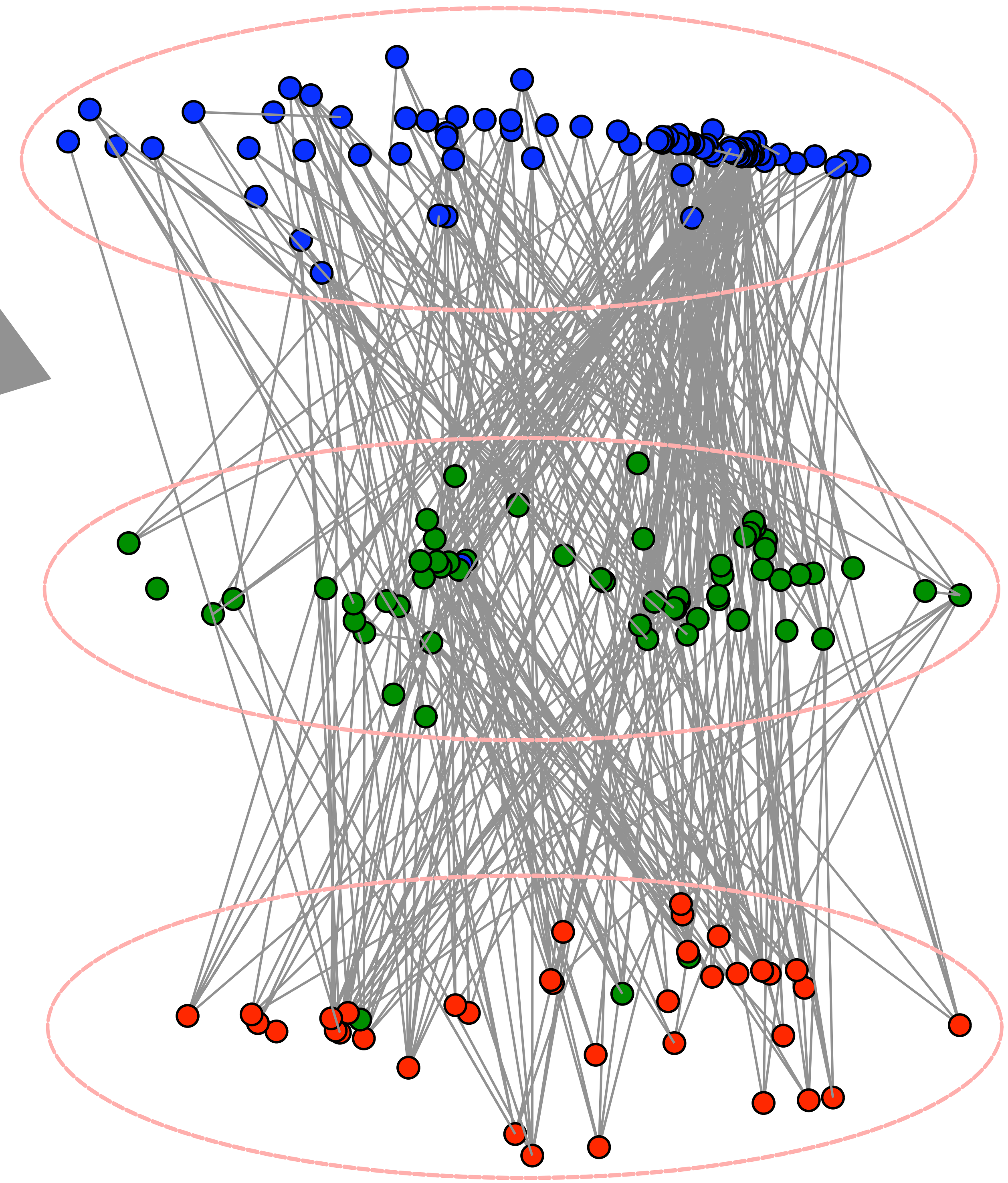
Ungrouped network



Discriminating node properties



Grouped network



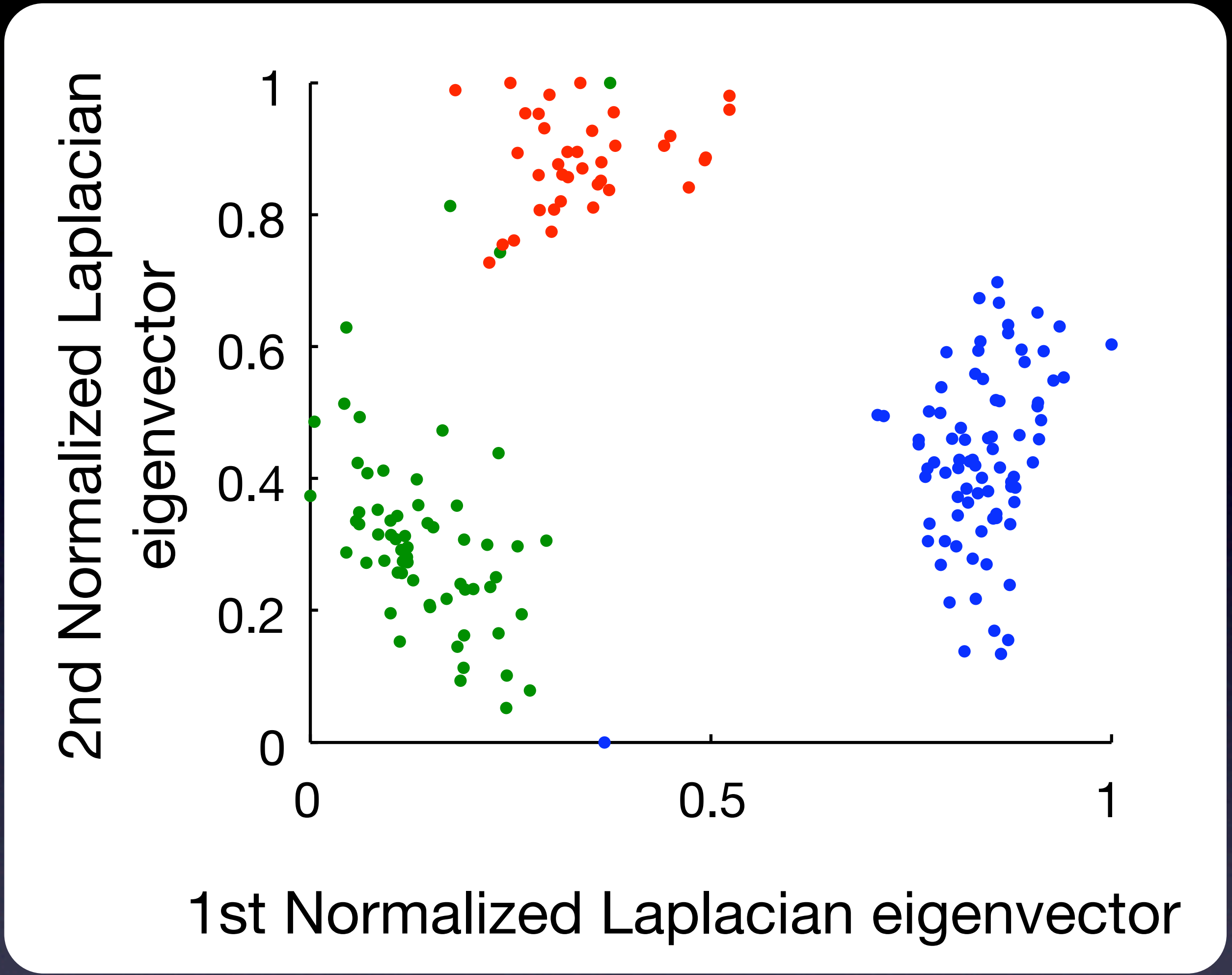
Colors - discovered assignment
Ellipses - prescribed assignment

Properties Discriminating Nodes

For node property j, define

$$D_j = \sum_{k=1}^{K} \sum_{k'=1}^{K} |\bar{x}_j^{(k)} - \bar{x}_j^{(k')}|$$

$$\bar{x}_j^{(k)} = \frac{1}{|G_k|} \sum_{i \in G_k} x_j^{(i)}$$

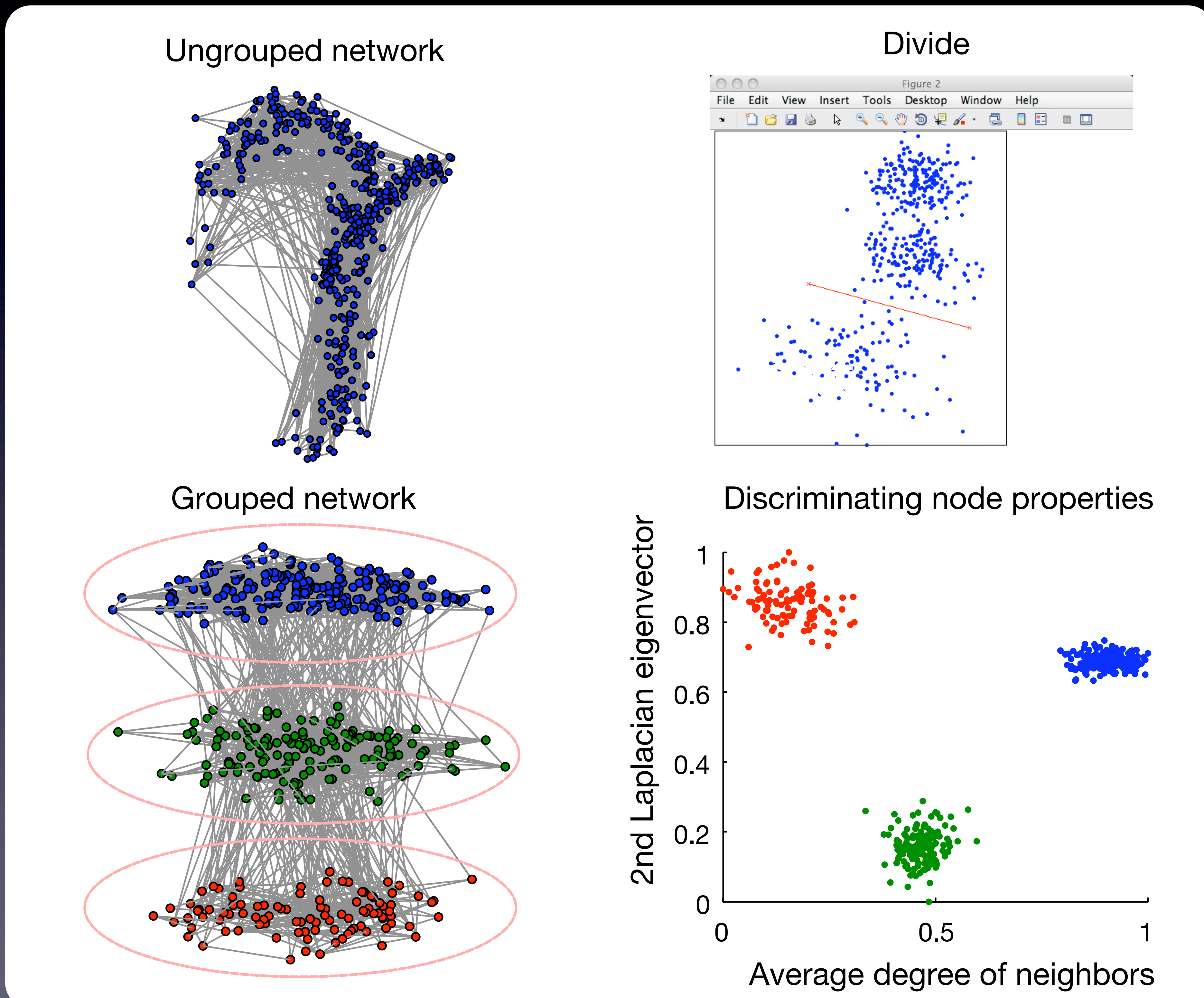


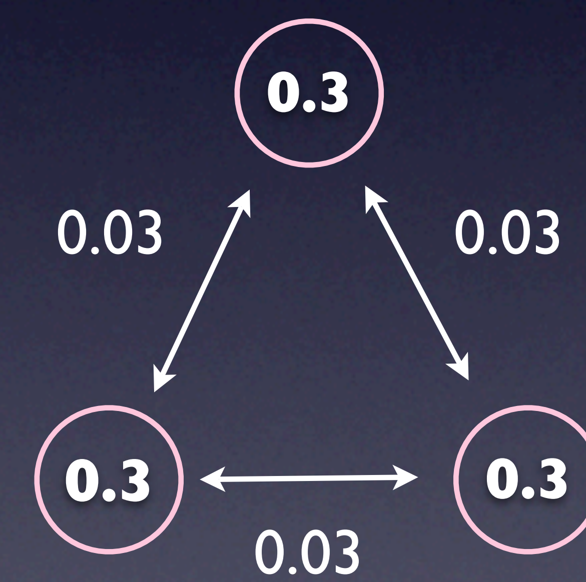
G_k: set of indices for nodes in group k

$|G_k|$ = the number of nodes in group k

We choose node properties with largest D_j, and project points onto the corresponding subspace.

Example 2
Community Structure



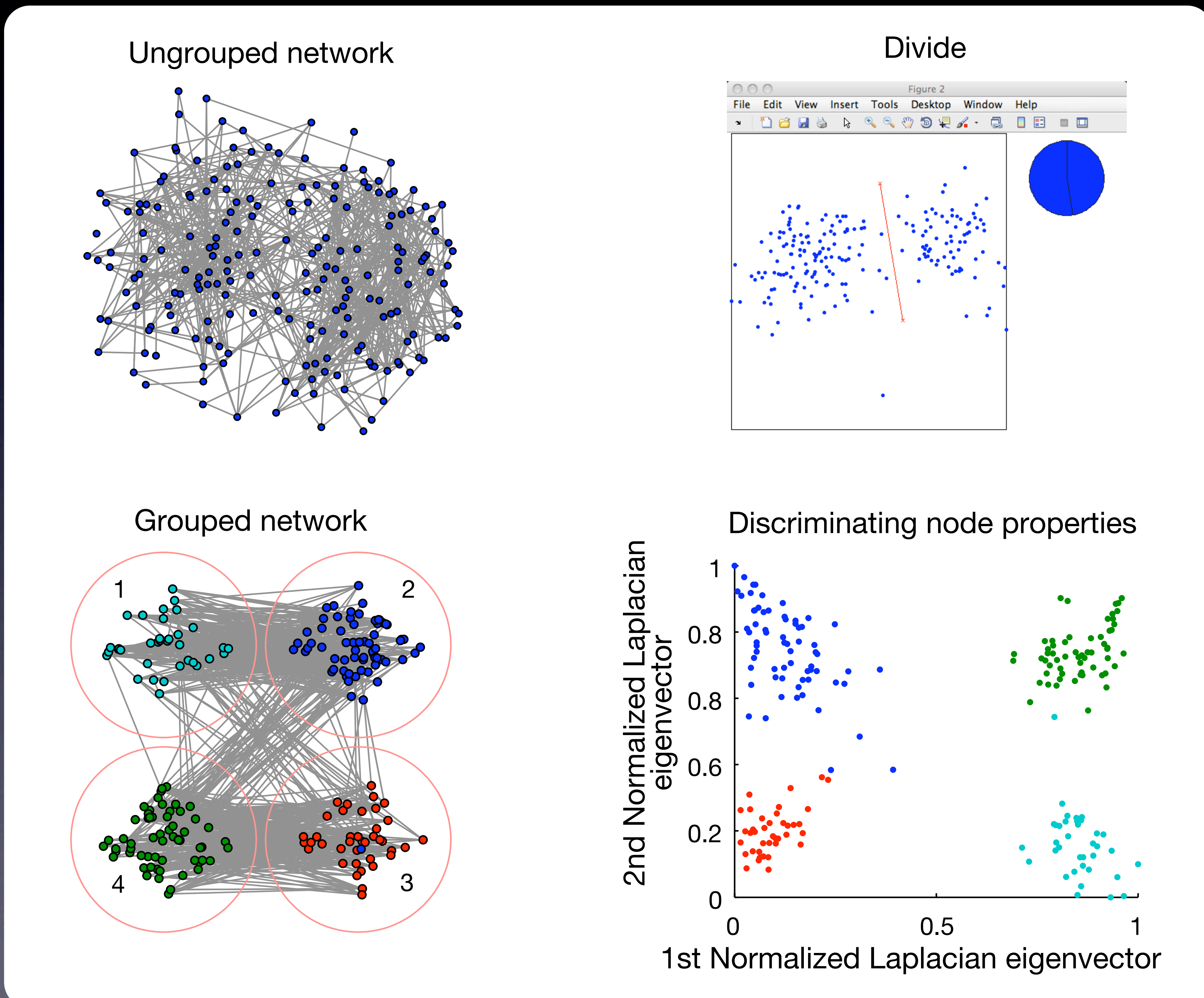


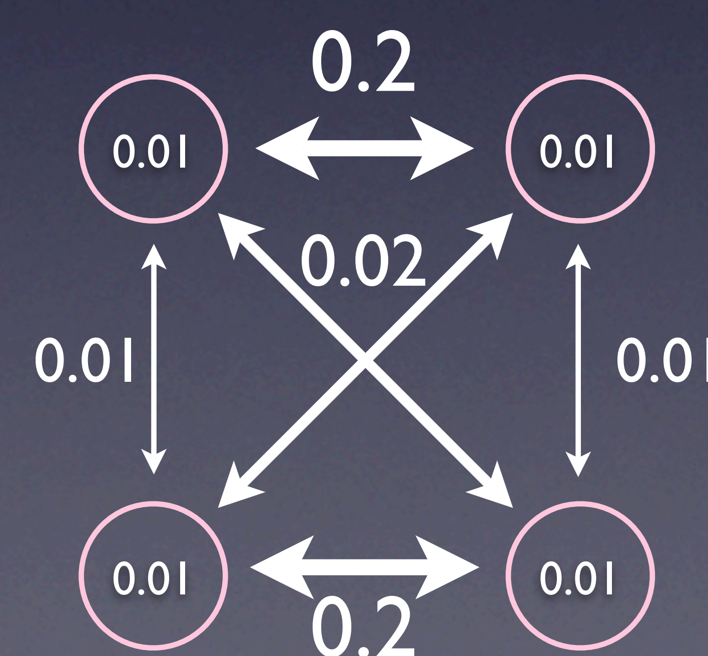
$$p_{\text{within}} = 0.3$$

$$p_{\text{across}} = 0.03$$

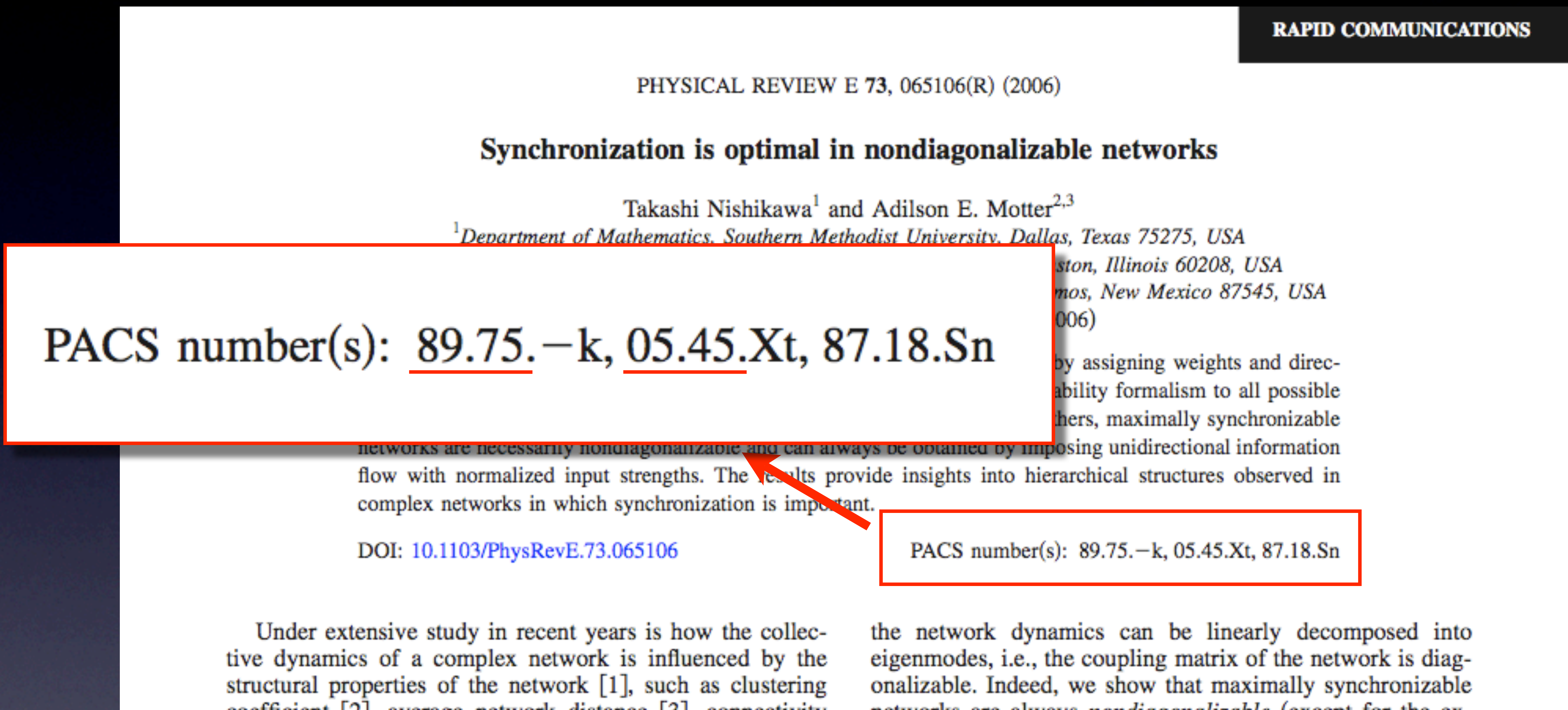
Example 3

Mixed Group Structure





Example 4: PACS Network



RAPID COMMUNICATIONS

PHYSICAL REVIEW E **73**, 065106(R) (2006)

Synchronization is optimal in nondiagonalizable networks

Takashi Nishikawa[1] and Adilson E. Motter[2,3]

[1]*Department of Mathematics, Southern Methodist University, Dallas, Texas 75275, USA*

...ston, Illinois 60208, USA

...mos, New Mexico 87545, USA

...006)

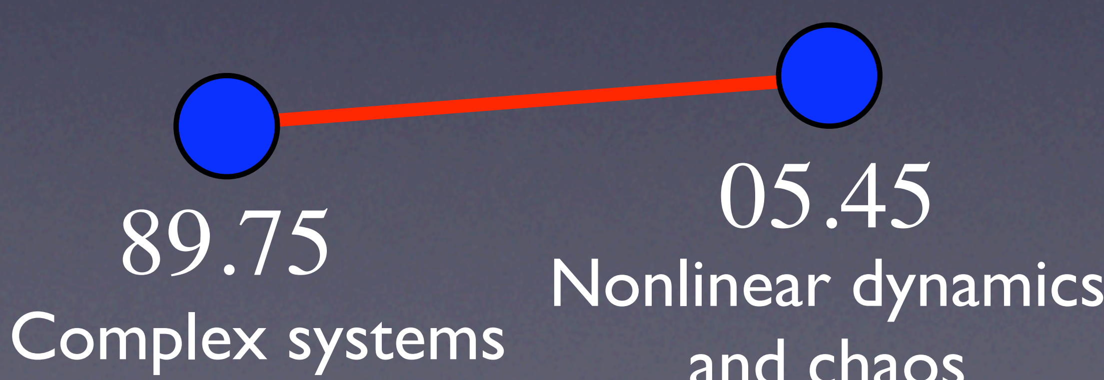
...by assigning weights and direc-
...ability formalism to all possible
...thers, maximally synchronizable
networks are necessarily nondiagonalizable and can always be obtained by imposing unidirectional information flow with normalized input strengths. The results provide insights into hierarchical structures observed in complex networks in which synchronization is important.

DOI: 10.1103/PhysRevE.73.065106

PACS number(s): 89.75.−k, 05.45.Xt, 87.18.Sn

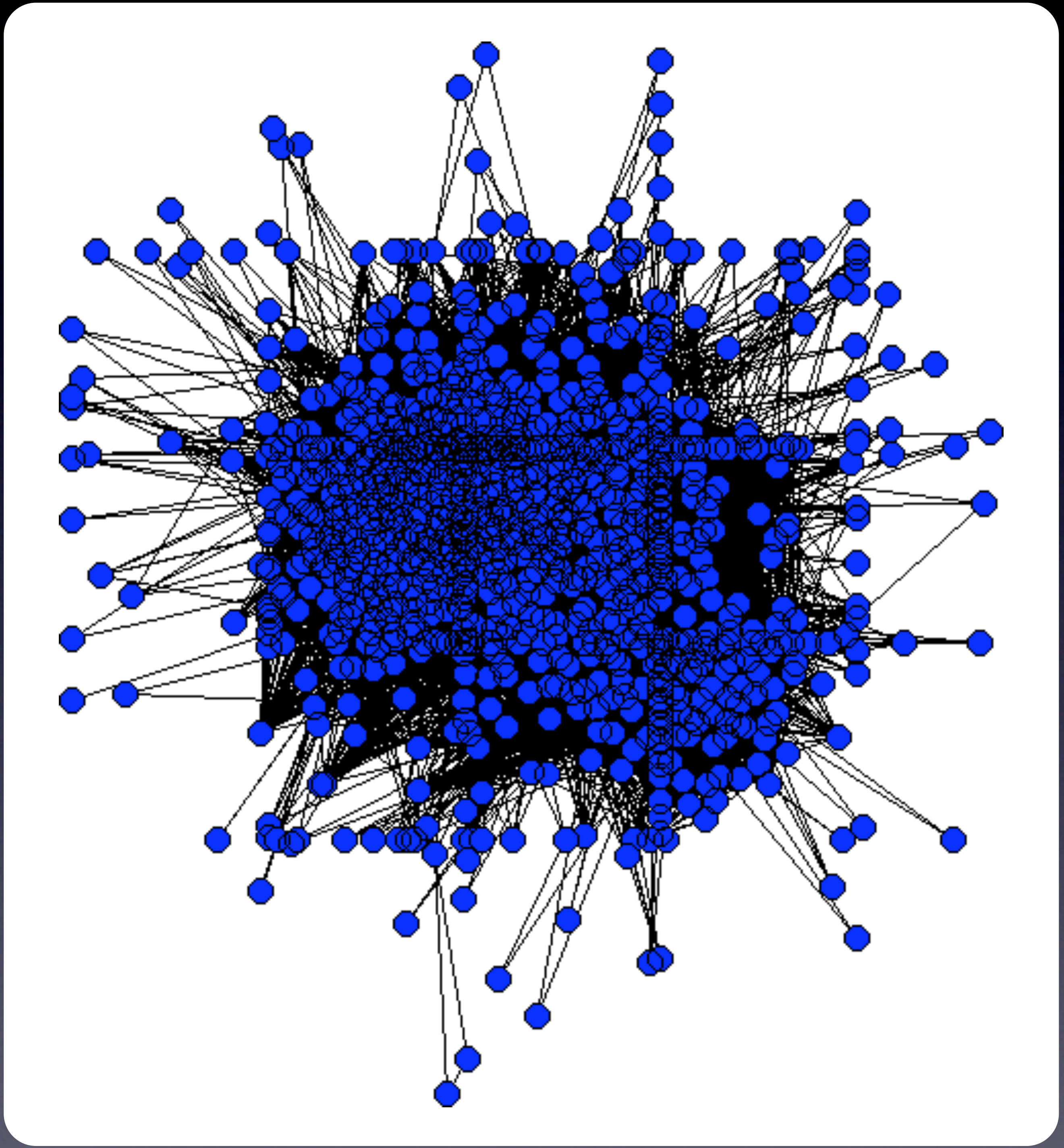
Under extensive study in recent years is how the collective dynamics of a complex network is influenced by the structural properties of the network [1], such as clustering coefficient [2], average network distance [3], connectivity ... the network dynamics can be linearly decomposed into eigenmodes, i.e., the coupling matrix of the network is diagonalizable. Indeed, we show that maximally synchronizable networks are always *nondiagonalizable* (except for the ex-...

PACS number(s): 89.75.−k, 05.45.Xt, 87.18.Sn



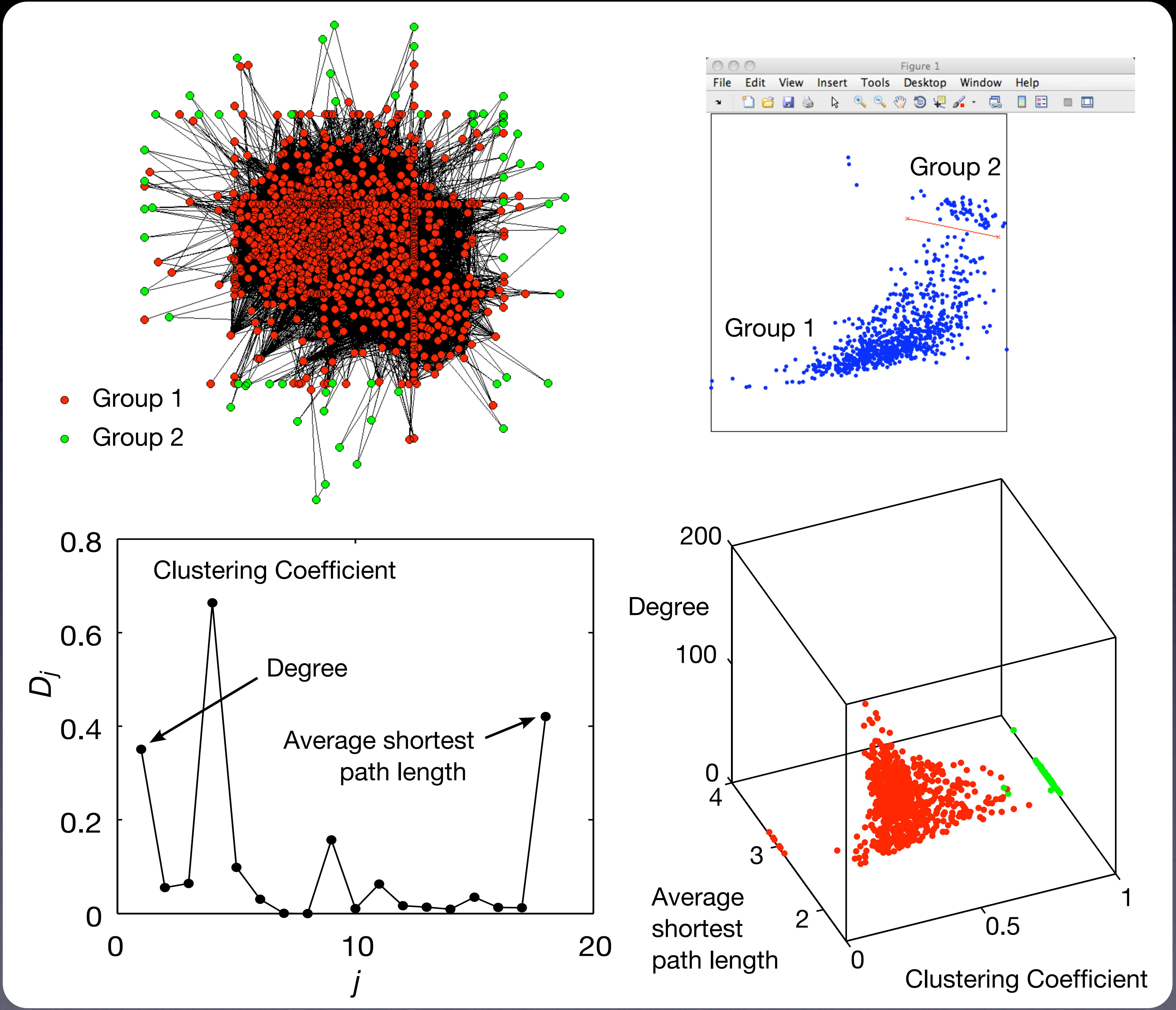
Nodes are connected if there are more papers than expected by random & independent occurrence

Example 4: PACS Network

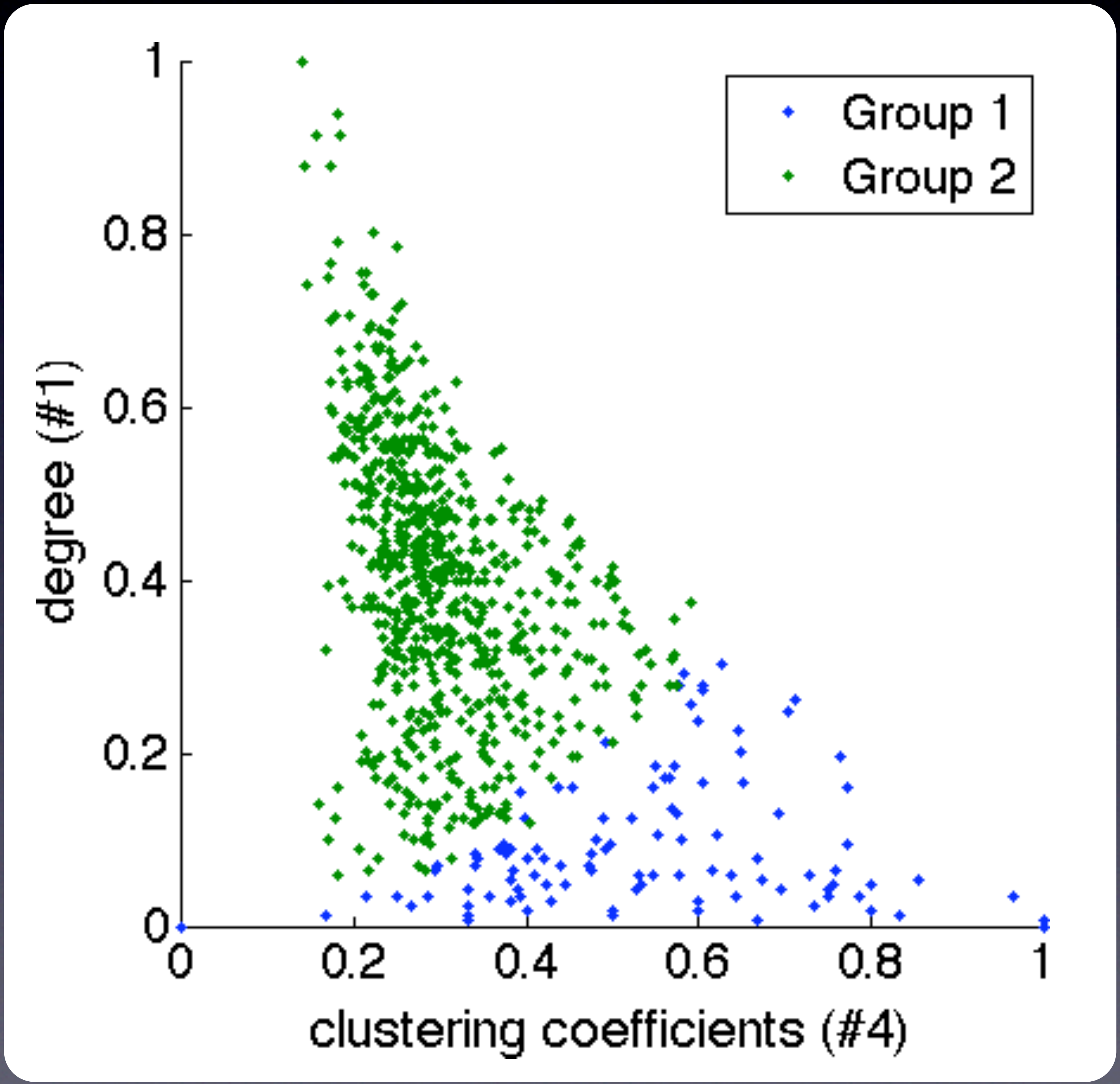


Data source: http://www.atsweb.neu.edu/ngulbahce/pacsdata.html
Reference: M. Herrera, D.C. Roberts and N. Gulbahce, arxiv:0904.1234

Example 4: PACS Network



A problem with k-Means Clustering



Future Work

Comprehensive benchmarking

- As a network group detection method: against community detection algorithms, expectation-maximization method, ...
- As a high-dimensional clustering method: against k-means clustering, unsupervised SVM, ...

Future Work

- Can we substitute human eye with 2-means clustering or unsupervised two-class SVM?
- Exploring a high-dimensional network parameter space, given only coarse & non-uniform sample points: How does dynamical properties (cascading, synchronization, etc.) depend on network structural parameters?