Lecture 9 Graph Partitioning – Application of Laplacian

n_1 $\quad n_1+n_2=n \quad$ n_2

Goal: Minimizing # of links between two groups – R

$$R = \frac{1}{2}\sum_{\substack{i,j \text{ in different} \\ \text{groups}}} A_{ij} \quad \text{— "cut size"}$$

Given n_1, n_2 $\qquad$ N-P hard

Result: $\quad \mathbb{L}$: $\quad 0 = \lambda_1 \le \lambda_2 \le \cdots \le \lambda_n$

$$\mathbb{L}\cdot \underset{\sim}{v}_2 = \lambda_2 \underset{\sim}{v}_2 \qquad \underset{\sim}{v}_2 = \begin{pmatrix} \updownarrow \end{pmatrix}_{n\times 1} \rightarrow \begin{pmatrix} n_1 \\ \hline n_2 \end{pmatrix} \text{ or } \begin{pmatrix} n_2 \\ \hline n_1 \end{pmatrix}$$

Sort elements of $\underset{\sim}{v}_2$ $\qquad$ Choose one with smaller cut size

<u>Proof</u> $\quad$ Introducing

$$s_i \equiv \begin{cases} +1, & i \in \text{Group 1} \\ -1, & i \in \text{Group 2} \end{cases}$$

$$\Rightarrow \quad \frac{1}{2}(1 - s_i s_j) = \begin{cases} 1, & i\,\&\,j \text{ in different groups} \\ 0, & i\,\&\,j \text{ in the same group} \end{cases}$$

$$\Rightarrow \quad R = \frac{1}{4}\sum_{i,j} A_{ij}(1 - s_i s_j)$$

$$\boxed{\sum_{ij} A_{ij} = \sum_i k_i = \sum_i k_i s_i^2 = \sum_{ij} k_i \delta_{ij} s_i s_j}$$

$$= \frac{1}{4}\sum_{i,j}(k_i \delta_{ij} - A_{ij}) s_i s_j$$

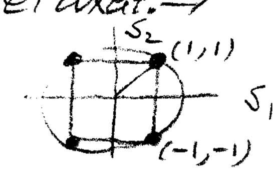
$$= \frac{1}{4}\sum_{i,j} L_{ij} s_i s_j$$

$$= \frac{1}{4}\underset{\sim}{s}^T \cdot \mathbb{L} \cdot \underset{\sim}{s}, \qquad \underset{\sim}{s} \equiv \begin{pmatrix} s_1 \\ s_n \end{pmatrix} \text{ — defines division}$$

Goal: For given $\mathbb{L}$, find $\underset{\sim}{s}$ that minimizes R

Difficulty: s_i takes on ± 1 values only

Relaxat. ⇒ $\quad s_i$ takes on any real value but with constraints (two)



(1) $\quad$ n – dim. hypercube $\qquad |s| = \sqrt{n}$

$\qquad$ or $\quad \sum_i s_i^2 = n^2$ – exclude trivial solution $|\underset{\sim}{s}| = 0$

(2) $\quad \sum_i s_i = n_1 - n_2 \quad$ or $\quad \underset{\sim}{1}^T \cdot \underset{\sim}{s} = n_1 - n_2$

Optimization: Lagrange multipliers: λ, 2μ

$$\frac{\partial}{\partial s_j}\left[\underset{\sim}{s}^T \cdot \mathbb{L} \cdot \underset{\sim}{s} + \lambda\left(n - \sum_j s_j^2\right) + 2\mu\left((n_1 - n_2) - \sum_j s_j\right)\right] = 0$$

$$\Rightarrow \quad \sum_j L_{ij} s_j = \lambda s_i + \mu$$

$$\Rightarrow \quad \boxed{\mathbb{L}\cdot \underset{\sim}{s} = \lambda \underset{\sim}{s} + \mu \underset{\sim}{1}}$$

$\mathbb{L}\cdot \underset{\sim}{1} = 0$ $\qquad$ $\underset{\sim}{1}^T \cdot \mathbb{L} \cdot \underset{\sim}{s} = \lambda\, \underset{\sim}{1}^T \cdot \underset{\sim}{s} + \mu\, \underset{\sim}{1}^T \cdot \underset{\sim}{1}$

$\underset{\sim}{1}^T \cdot \mathbb{L} = 0$ $\qquad$ $0 = \lambda(n_1 - n_2) + \mu n$

$$\Rightarrow \quad \mu = -[(n_1 - n_2)/n]\,\lambda$$

Define $\quad \underset{\sim}{x} \equiv \underset{\sim}{s} + \frac{\mu}{\lambda}\underset{\sim}{1} = \underset{\sim}{s} - \frac{n_1 - n_2}{n}\underset{\sim}{1}$

Try to find eigenvalue & eigenvectors of $\mathbb{L}$

$$\mathbb{L}\cdot\underset{\sim}{x} = \mathbb{L}\cdot\left(\underset{\sim}{s} + \frac{\mu}{\lambda}\underset{\sim}{1}\right) = \mathbb{L}\cdot\underset{\sim}{s}$$

but $\mathbb{L}\cdot\underset{\sim}{s} = \lambda\underset{\sim}{s} + \mu\underset{\sim}{1} = \lambda\underset{\sim}{x}$ $\Big\}$ $\mathbb{L}\cdot\underset{\sim}{x} = \lambda\underset{\sim}{x}$ (eigenvalue λ, eigenvector $\underset{\sim}{x}$)

Optimization problem: choose an eigenvector of $\mathbb{L}$ so that R is minimum

(choosing $\underset{\sim}{s}$ is like choosing $\underset{\sim}{x}$)

How about $\underset{\sim}{1}$?

$$\underset{\sim}{1}^T\cdot\underset{\sim}{x} = \underset{\sim}{1}^T\cdot\underset{\sim}{s} - \frac{\mu}{\lambda}\underset{\sim}{1}^T\cdot\underset{\sim}{1} = (n_1-n_2) - \frac{n_1-n_2}{n}\cdot n = 0$$

$$\Rightarrow \underset{\sim}{x}\perp\underset{\sim}{1} \Rightarrow \underset{\sim}{1} \text{ is not a choice}$$

which eigenvector to choose?

$$R = \frac{1}{4}\underset{\sim}{s}^T\cdot\mathbb{L}\cdot\underset{\sim}{s} = \frac{1}{4}\underset{\sim}{x}^T\cdot\mathbb{L}\cdot\underset{\sim}{x} = \frac{\lambda}{4}\underset{\sim}{x}^T\cdot\underset{\sim}{x}$$
$$= \frac{\lambda}{4}\left(\underset{\sim}{s}^T\cdot\underset{\sim}{s} + \frac{\mu}{\lambda}\left(\underset{\sim}{s}^T\cdot\underset{\sim}{1} + \underset{\sim}{1}^T\cdot\underset{\sim}{s}\right) + \frac{\mu^2}{\lambda^2}\underset{\sim}{1}^T\cdot\underset{\sim}{1}\right)$$
$$= \frac{\lambda}{4}\left(n - 2\,\frac{n_1-n_2}{n}(n_1-n_2) + \frac{(n_1-n_2)^2}{n^2}\,n\right)$$
$$= (\lambda/4)\cdot 4n_1n_2/n = \frac{n_1n_2}{n}\lambda$$

Cut size $\sim \lambda$

Prescription: Choose $\underset{\sim}{x}$ to be the eigenvector a.w. the smallest nontrivial eigenvalue of $\mathbb{L}$

$$\Rightarrow \underset{\sim}{v_2} \;\&\; \lambda_2 \qquad \mathbb{L}\cdot\underset{\sim}{v_2} = \lambda_2\underset{\sim}{v_2}$$

$$\underset{\sim}{s} = \underset{\sim}{v_2} + \frac{n_1-n_2}{n}\underset{\sim}{1}$$

$$\Rightarrow \underset{\Downarrow\,\pm 1}{s_i} = \underbrace{v_{2i} + (n_1-n_2)/n}_{\text{can be any value}} \text{ — optimal value of } s_i$$

Choose $\underset{\sim}{s}$ as close to the optimal value as possible

$$\Rightarrow \text{Making } \underset{\sim}{s}^T\cdot\left(\underset{\sim}{v_2} + \frac{n_1-n_2}{n}\underset{\sim}{1}\right) \text{ as large as possible}$$

$$\Rightarrow \text{Maximize}\left\{\sum s_i\left(v_{2i} + \frac{n_1-n_2}{n}\right)\right\}$$

$$\underset{\sim}{v_2} = \begin{pmatrix} v_{21} \\ v_{22} \\ \vdots \\ v_{2n}\end{pmatrix}\begin{matrix}\text{– node 1} \\ \text{– node 2} \\ \\ \text{– node } n\end{matrix} \;\xrightarrow{\text{Sorting}}\; \begin{pmatrix} v_{21}' \\ v_{22}' \\ \vdots \\ \hline \vdots \\ v_{2n}'\end{pmatrix}\begin{matrix}\Big\}\, n_1 \\ \\ \}\, n_2\end{matrix} \begin{matrix}\text{- node 1'} \\ \text{- node 2'} \\ \\ \text{node } n'\end{matrix}$$

$$v_{21}' \geq v_{22}' \geq \cdots \geq v_{2n}'$$

Still two possibilities — compare values of R

(Definition of n_1 & n_2 — arbitrary)

$$\begin{pmatrix} v_{21}' \\ v_{22}' \\ v_{23}' \\ \hline v_{24}' \\ v_{25}'\end{pmatrix}\begin{matrix}\Big\}\, n_1 = 3 \\ \\ \}\, n_2 = 2\end{matrix} \quad\text{or}\quad \begin{pmatrix} v_{21}' \\ v_{22}' \\ \hline v_{23}' \\ v_{24}' \\ v_{25}'\end{pmatrix}\begin{matrix}\}\, n_1 = 2 \\ \\ \Big\}\, n_2 = 3\end{matrix}$$

Network Community Structure

Community structure in social and biological networks

M. Girvan*†‡ and M. E. J. Newman*§

*Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501; †Department of Physics, Cornell Univ
§Department of Physics, University of Michigan, Ann Arbor, MI 48109-1120

Edited by Lawrence A. Shepp, Rutgers, State University of New Jersey–New Brunswick, Piscataway, N
December 6, 2001)

A number of recent studies have focused on the statistical properties of networked systems such as social networks and the Worldwide Web. Researchers have concentrated particularly on a few properties that seem to be common to many networks: the small-world property, power-law degree distributions, and network transitivity. In this article, we highlight another property that is found in many networks, the property of community structure, in which network nodes are joined together in tightly knit groups, between which there are only looser connections. We propose a method for detecting such communities, built around the idea of using centrality indices to find community boundaries. We test our method on computer-generated and real-world graphs whose community structure is already known and find that the method detects this known structure with high sensitivity and reliability. We also apply the method to two networks whose community structure is not well known—a collaboration network and a food web—and find that it detects significant and informative community divisions in both cases.

In this article,
show, appears to
community stru
clustering, but w
the other mean
preceding parag
networks—netw
tween individua
networks seem t
within which ve
which connecti
network with su
(Certainly it is
join together to
communities ar
hierarchical fasl
section.) The ab
could clearly hav

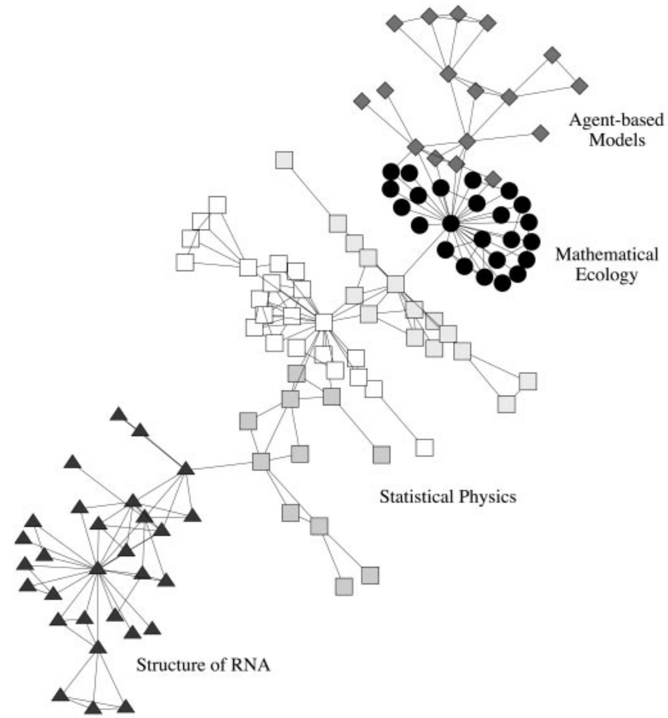


Fig. 6. The largest component of the Santa Fe Institute collaboration network, with the primary divisions detected by our algorithm indicated by different vertex shapes.

PNAS 99, 7821-7826 (2002)

Estimating the Number of Communities

PRL **117,** 078301 (2016) PHYSICAL REVIEW LETTERS week ending 12 AUGUST 2016

Estimating the Number of Communities in a Network

M. E. J. Newman[1,2] and Gesine Reinert[3]

[1]*Department of Physics, University of Michigan, Ann Arbor, Michigan 48109, USA*

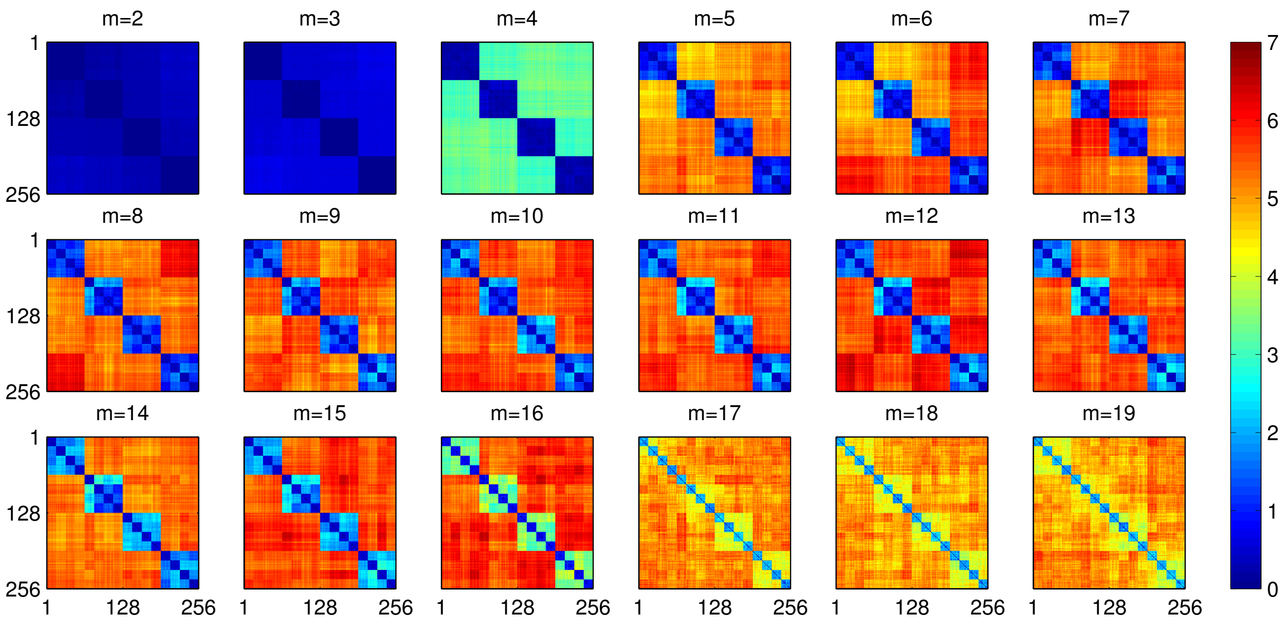
[2]*Rudolph Peierls Centre for Theoretical Physics, University of Oxford, 1 Keble Road, Oxford OX1 3NP, United Kingdom*

[3]*Department of Statistics, University of Oxford, 24-29 St. Giles, Oxford OX1 3LB, United Kingdom*

(Received 16 May 2016; published 11 August 2016)

Community detection, the division of a network into dense subnetworks with only sparse connections between them, has been a topic of vigorous study in recent years. However, while there exist a range of effective methods for dividing a network into a specified number of communities, it is an open question how to determine exactly how many communities one should use. Here we describe a mathematically principled approach for finding the number of communities in a network by maximizing the integrated likelihood of the observed network structure under an appropriate generative model. We demonstrate the approach on a range of benchmark networks, both real and computer generated.

Dynamics Based Approach – Paradigm Shift?



Z. Zhuo, S.-M. Cai, M. Tang, and Y.-C. Lai, "Accurate detection of hierarchical communities in complex networks based on nonlinear dynamical evolution," preprint (2018).

FIG. 3. **Emergence of hierarchical community structure through synchronization**. The benchmark network has the structure of (14,3), with its eigenvalue spectrum shown in Fig. 1. The nodal dynamical system is that of a chaotic Rössler oscillator. The coupling parameter is continuously decreased so that, starting from $m = 2$, the nontrivial eigenmodes lose their transverse stability one after another. Shown are the synchronization error matrix constructed from all the pairwise distances between the nodal dynamical variables, where the distances are color coded. For each panel, the integer value of m corresponds to the case where the $(m-1)$ nontrivial eigenmodes (from two to m) are transversely unstable. For $m \leq 4$, the synchronization states reflect correctly the four large communities at second hierarchical level. As m is increased from 4 to 16 (corresponding to continuous decrease in the actual value of the coupling parameter), the degree of inter-community synchronization at the first hierarchical level of 16 communities is gradually weakened, revealing the community structure at the smaller scale. Insofar as $m \leq 16$, there is local synchronization within each of the 16 communities. For $m \geq 17$, synchronization at the small scale begins to deteriorate, revealing more refined structures within each such community.

Dynamics Based Approach – A Real Example

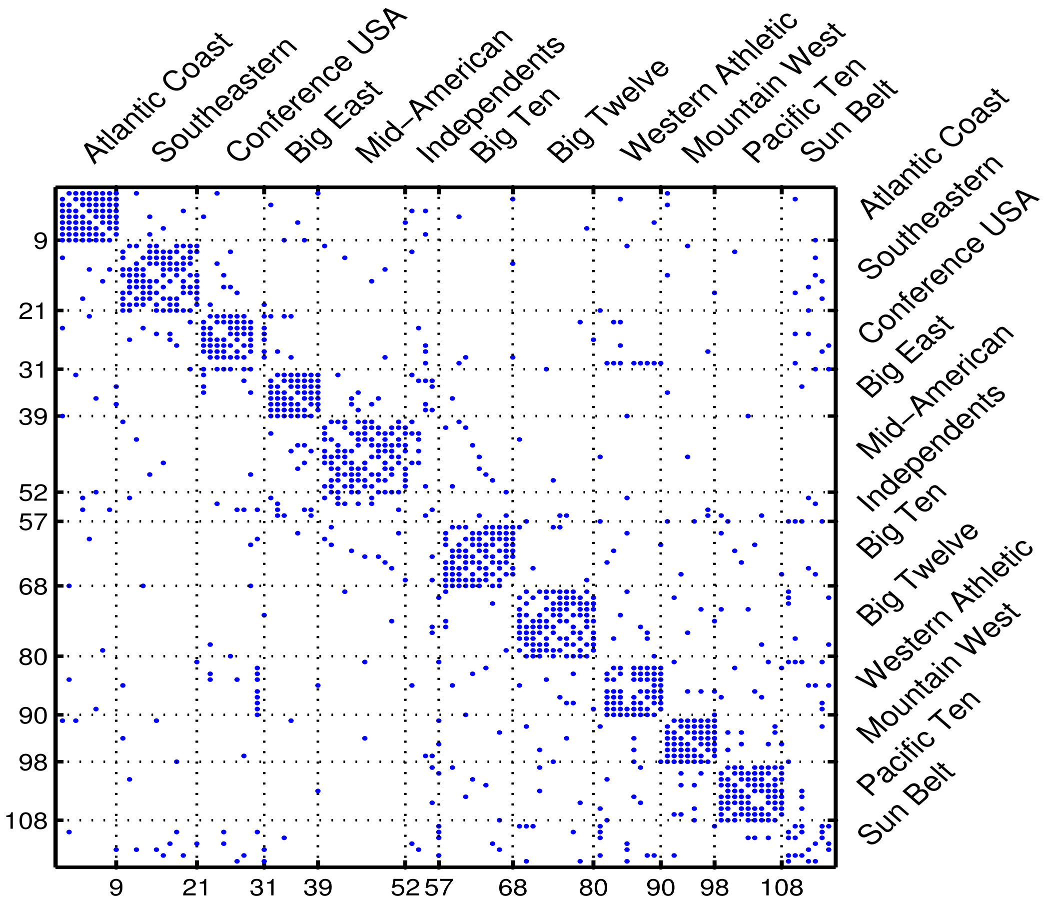


FIG. 6. **Structure of the American college football game network as represented by the adjacency matrix**. There are altogether 115 teams (115 nodes in the network), which are divided into 12 conferences - separated by lines. The names of the conferences are noted. Intra-conference games are more frequent than inter-conference ones, giving rise to a community structure. Two anomalies are the "Independents" and "Sun Belt" conferences, which have fewer intra-conference games. The conferences are organized into a hierarchical structure because inter-conference teams that are geographically close to one another are more likely to play in a game.