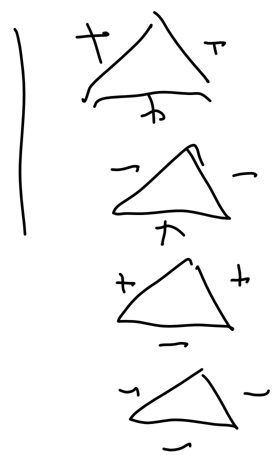
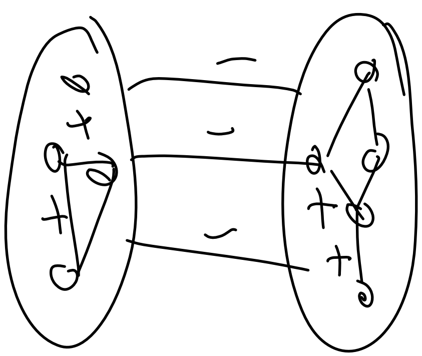
Spectral graph partitioning & Community structure of networks

CS 322: (Social and Information) Network Analysis
Jure Leskovec
Stanford University

Announcements

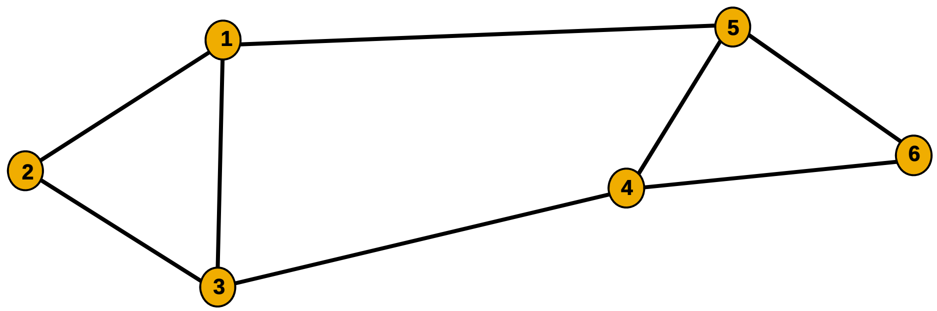
- Task:
  - Find coalitions in signed networks
- Awards:
  - Extra credit: 10%, 6%, 4%
  - **European chocolates!**



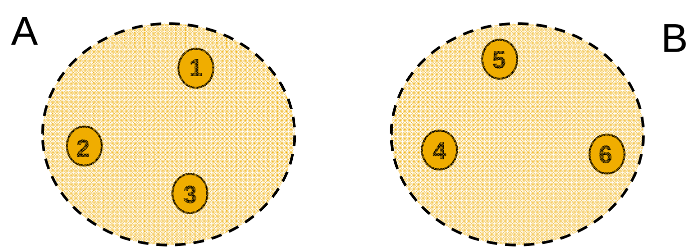


Graph partitioning

- Graph G(V,E):



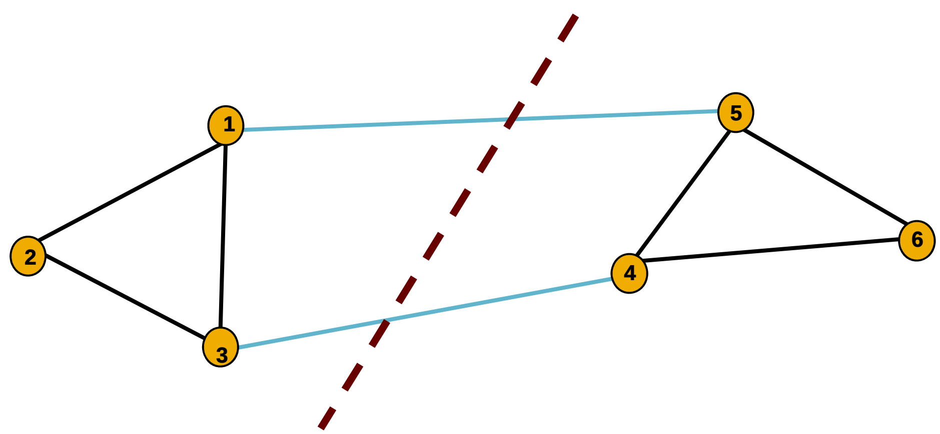
- Bi-partitioning task:
  - Divide vertices into two disjoint groups (A,B)



- Questions:
  - How can we define a "good" partition of the graph?
  - How can we efficiently identify such a partition?

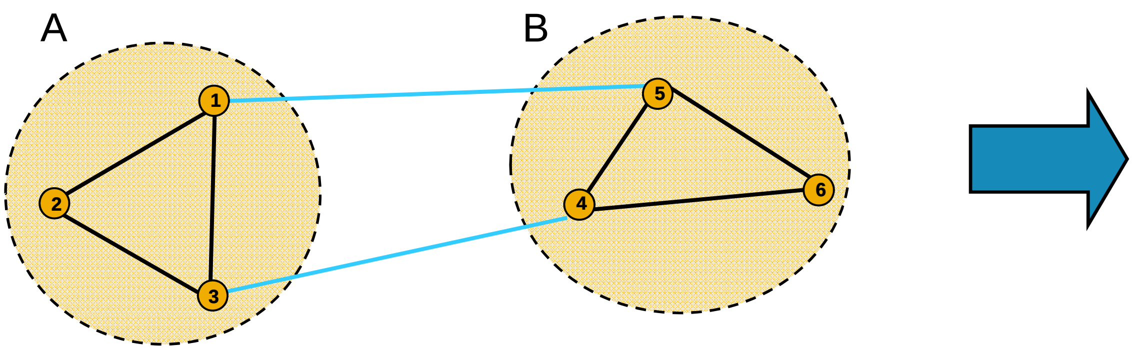
Graph partitioning

- Maximize the number of within-group connections
- Minimize the number of between-group connections



Graph Cuts

- Express partitioning objectives as a function of the "edge cut" of the partition.
- Cut: Set of edges with only one vertex in a group. $cut(A,B) = \sum_{i \in A, j \in B} w_{ij}$



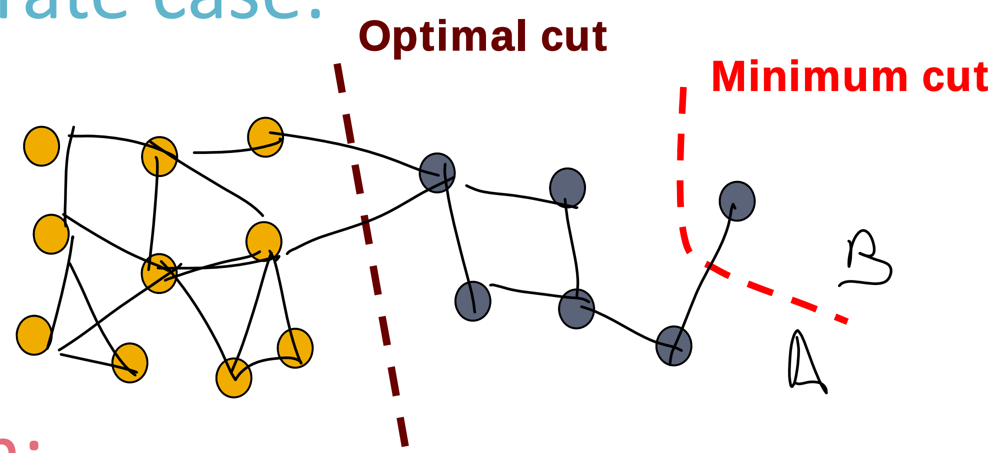
$cut(A,B) = 2$

Graph Cut Criteria

- Criterion: Minimum-cut
  - Minimise weight of connections between groups

$$\min \; cut(A,B)$$

- Degenerate case:



- Problem:
  - Only considers external cluster connections
  - Does not consider internal cluster density

Graph Cut Criteria

- Criterion: Normalised-cut [Shi-Malik, '97]
  - Consider the connectivity between groups relative to the density of each group

$$\min Ncut(A,B) = \frac{cut(A,B)}{vol(A)} + \frac{cut(A,B)}{vol(B)}$$

  - Normalize the association between groups by volume.
    - Vol(A): The total weight of the edges originating from group A.
- Why use this criterion?
  - Minimizing the normalized cut is equivalent to maximizing normalized association.
  - Produces more balanced partitions

Question

- How do we efficiently find a good partition?
- Problem:
  - Computing optimal cut is NP-hard

Spectral Graph Partitioning

- A = adjacency matrix of G
  - A_{ij} =1 if (i,j) is an edge
    0 else
- x is a vector in $R^n = (x_1, \ldots, x_n)$
  - (just a labeling of the vertices of G)

- What is the meaning of Ax?

$$\begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \lambda \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

What is the meaning of Ax?

$$\begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \lambda \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

- Each entry x_j is a sum of labels x_i of all neighbors
- j^{th} coordinate of Ax:
  - Sum of the x-values at all neighbors of node j. Make this a new value at node j
- Spectral Graph Theory:
  - Analyse the "spectrum" of matrix representing a graph.
  - Spectrum : The eigenvectors of a graph, ordered by the magnitude(strength) of their corresponding eigenvalues. $\Lambda = \{\lambda_1, \lambda_2, \ldots, \lambda_n\}$

Example

- Suppose all nodes in G have degree d (G is d-regular) and G is connected
- What are some eigenvalues/vectors of G?
- $Ax = \lambda x$

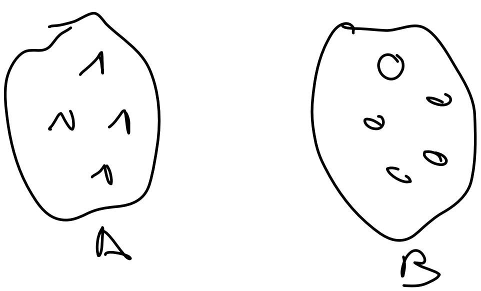
if $x = (1, 1, 1, \ldots, 1)$

$Ax = d\,x$

$\lambda = d$

Example

- What if G is not connected?
  - Say G has 2 components
- What are some eigenvectors?
  - All 1 on A and 0 on B or vice versa

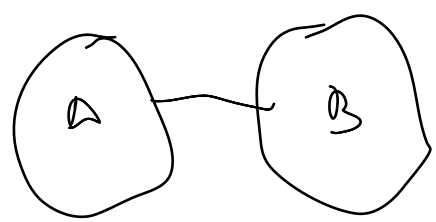


$(1\ 1\ 1,\ 0\ 0\ 0)$

$(0\ 0\ 0,\ 1\ 1\ 1)$

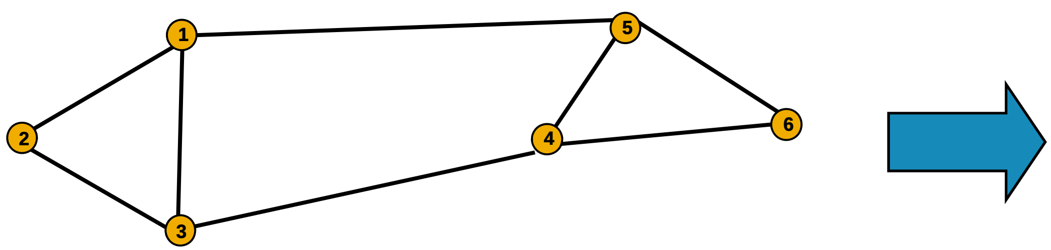
$Ax = \lambda x$

$\lambda = d$



Matrix Representations

- Adjacency matrix (A):
  - n x n matrix
  - $A=[a_{ij}]$, $a_{ij}=1$ if edge between node i and j

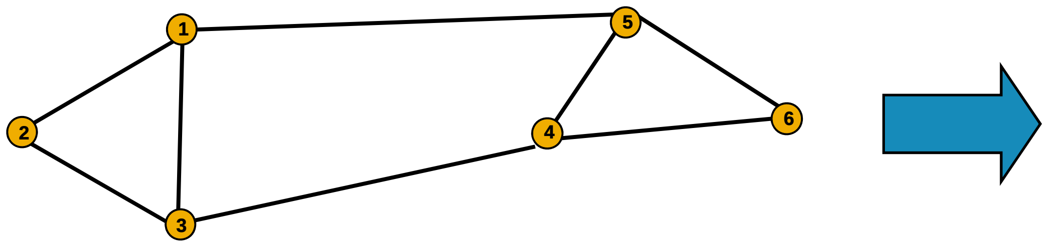


	1	2	3	4	5	6
1	0	1	1	0	1	0
2	1	0	1	0	0	0
3	1	1	0	1	0	0
4	0	0	1	0	1	1
5	1	0	0	1	0	1
6	0	0	0	1	1	0

- Important properties:
  - Symmetric matrix
  - Eigenvectors are real and orthogonal

Matrix Representations (continued)

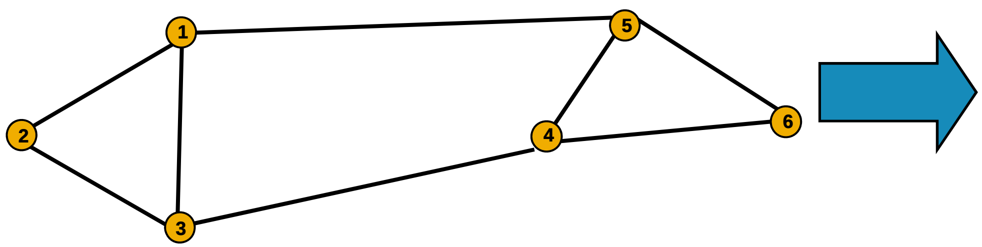
- Degree matrix (D)
  - n x n diagonal matrix
  - $D=[d_{ii}]$, d_{ii} = degree of node i



	1	2	3	4	5	6
1	3	0	0	0	0	0
2	0	2	0	0	0	0
3	0	0	3	0	0	0
4	0	0	0	3	0	0
5	0	0	0	0	3	0
6	0	0	0	0	0	2

Matrix Representations

- Laplacian matrix (L):
  - n x n symmetric matrix



$L = D - A$

	1	2	3	4	5	6
1	3	-1	-1	0	-1	0
2	-1	2	-1	0	0	0
3	-1	-1	3	-1	0	0
4	0	0	-1	3	-1	-1
5	-1	0	0	-1	3	-1
6	0	0	0	-1	-1	2

- What is trivial eigenvector?

$x = (1, 1, 1, 1, 1 \ldots 1)$

$\lambda_1 = 0$

- Important properties:
  - Eigenvalues are non-negative real numbers
  - Eigenvectors are real and orthogonal

λ_2 as optimization problem

- For symmetric matrix M:

$$\lambda_2 = \min \frac{x^T M x}{x^T x} = x^T M x$$

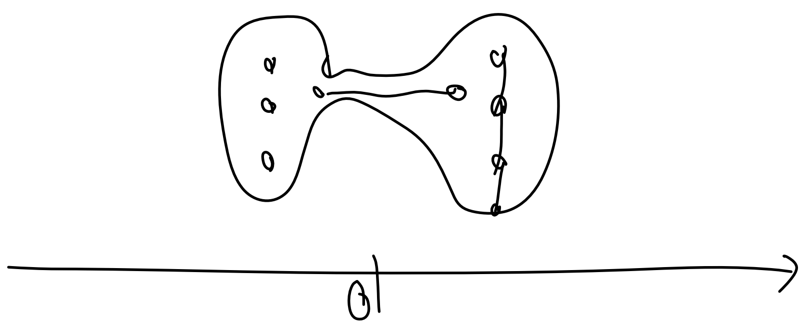
- What is the meaning of min $x^T L x$ on G?

$$x^T L x = \sum_{(i,j)\in E} (x_i - x_j)^2$$

λ_2 as optimization problem

- What else do we know about x?
  - *x* is unit vector $\sum x_i^2 = 1$
  - x is orthogonal to first eigenvector (1,….,1) thus: $\sum x_i = 0$
- Then:

$$\lambda_2 = \min_{\substack{\text{All labelings of} \\ \text{nodes so that} \\ \text{sum}(x_i)=0}} \frac{\sum (x_i - x_j)^2}{\sum x_i^2}$$



Find Optimal Cut (Hall'70, Fiedler'73)

- Express partition (A,B) as a vector

$$x_i = \begin{cases} +1 \text{ if } \mathrm{i} \in \mathrm{A} \\ -1 \text{ if } j \in \mathrm{B} \end{cases}$$

- We can minimize the cut of the partition by finding a non-trivial vector *x* that minimizes:

$$f(x) = \sum_{(i,j) \in E} (x_i - x_j)^2$$

Rayleigh Theorem

$$f(x) = \sum_{(i,j)\in E} (x_i - x_j)^2 = x^T L\, x$$

- The minimum value is given by the 2nd smallest eigenvalue λ_2 of the Laplacian matrix *L*

- The optimal solution for *x* is given by the corresponding eigenvector λ_2, referred as the Fiedler Vector

So far…

- How can we define a "good" partition of a graph?
  - Minimise a given graph cut criterion.
- How can we efficiently identify such a partition?
  - Approximate using information provided by the eigenvalues and eigenvectors of a graph.

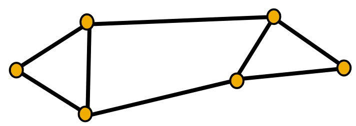
- Spectral Clustering (Simon et. al,'90)

Spectral Clustering Algorithms

- Three basic stages:
  1. Pre-processing
     - Construct a matrix representation of the graph
  2. Decomposition:
     - Compute eigenvalues and eigenvectors of the matrix
     - Map each point to a lower-dimensional representation based on one or more eigenvectors
  3. Grouping:
     - Assign points to two or more clusters, based on the new representation

Spectral Partitioning Algorithm

- Pre-processing
  - Build Laplacian matrix *L* of the graph



	1	2	3	4	5	6
1	3	-1	-1	0	-1	0
2	-1	2	-1	0	0	0
3	-1	-1	3	-1	0	0
4	0	0	-1	3	-1	-1
5	-1	0	0	-1	3	-1
6	0	0	0	-1	-1	2

- Decomposition
  - Find eigenvalues λ and eigenvectors *x* of the matrix *L*

$\lambda =$

0.0
1.0
3.0
3.0
4.0
5.0

X =

0.4	0.3	-0.5	-0.2	-0.4	-0.5
0.4	0.6	0.4	-0.4	0.4	0.0
0.4	0.3	0.1	0.6	-0.4	0.5
0.4	-0.3	0.1	0.6	0.4	-0.5
0.4	-0.3	-0.5	-0.2	0.4	-0.5
0.4	-0.6	-0.4	-0.4	-0.4	0.0

  - Map vertices to corresponding components of λ_2

1	0.3
2	0.6
3	0.3
4	-0.3
5	-0.3
6	-0.6

How do we find the clusters?

Spectral Partitioning (continued)

- Grouping
  - Sort components of reduced 1-dimensional vector.
  - Identify clusters by splitting the sorted vector in two.
- How to choose a splitting point?
  - Naïve approaches:
    - Split at 0, mean or median value
  - More expensive approaches
    - Attempt to minimise normalized cut criterion in 1-dimension

1	0.3
2	0.6
3	0.3
4	-0.3
5	-0.3
6	-0.6

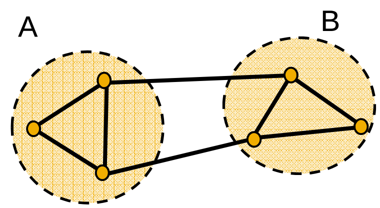
Split at 0

Cluster A: Positive points

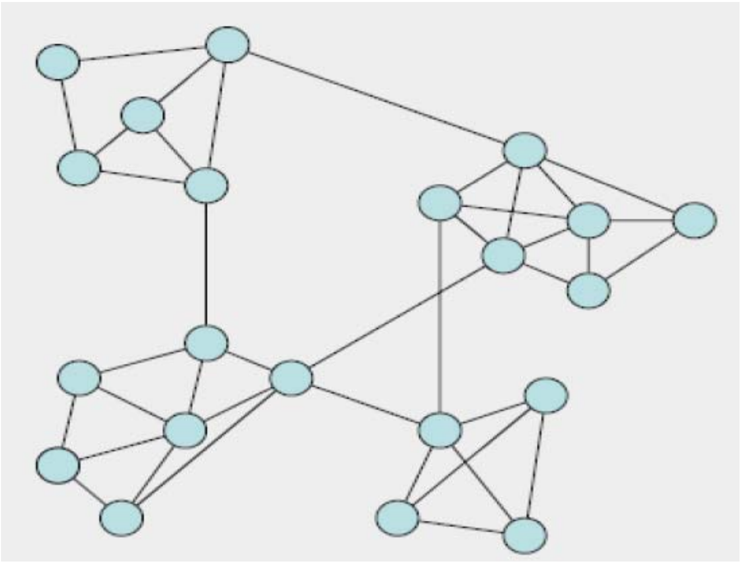
Cluster B: Negative points

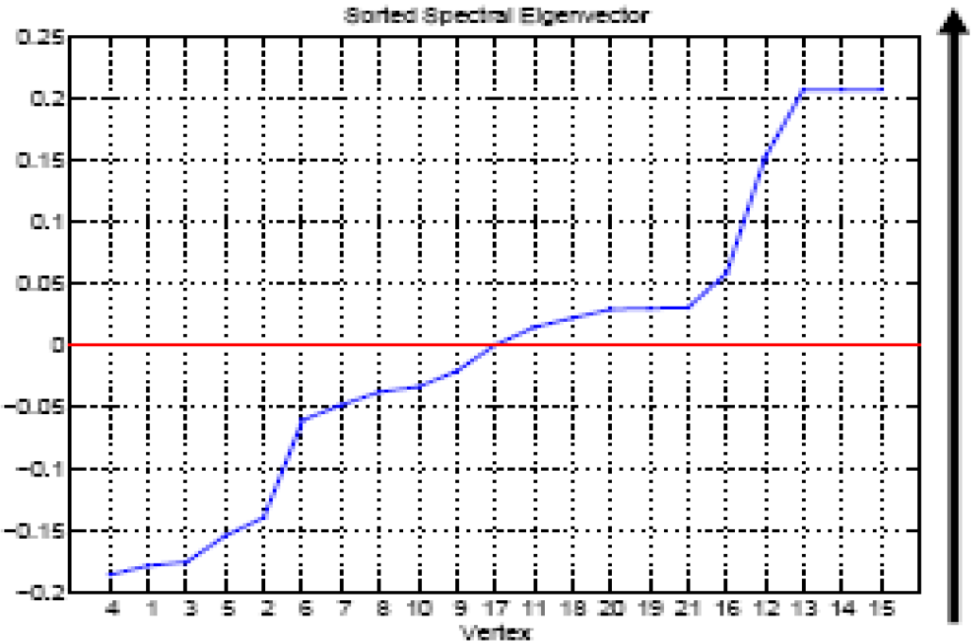
1	0.3
2	0.6
3	0.3

4	-0.3
5	-0.3
6	-0.6



Example

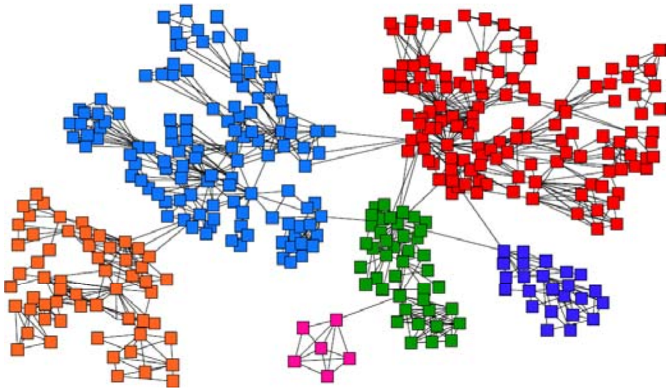




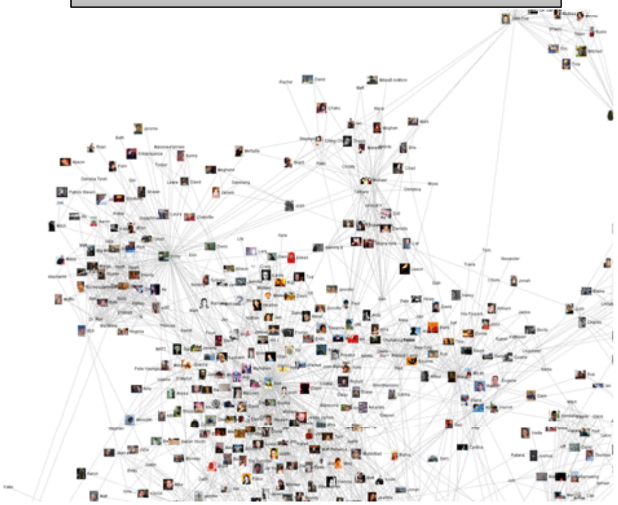
Sorted Spectral Eigenvector
0.25
0.2
0.15
0.1
0.05
0
−0.05
−0.1
−0.15
−0.2
4 1 3 5 2 6 7 8 10 9 17 11 18 20 19 21 16 12 13 14 15
Vertex

Community structure in networks

- What is the cluster structure of networks?
- How does it scale from small to very large networks?
- How to think about clusters in large networks?



Physics collaborations



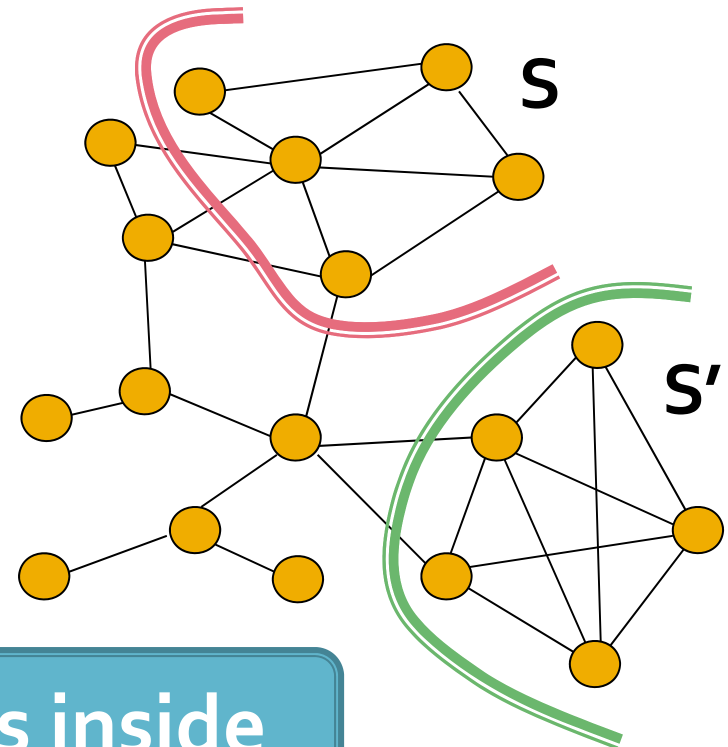
Part of a large social network

Objective function

What is a good cluster?

- Many edges internally
- Few pointing outside

Formally, **conductance/ normalized cut:**



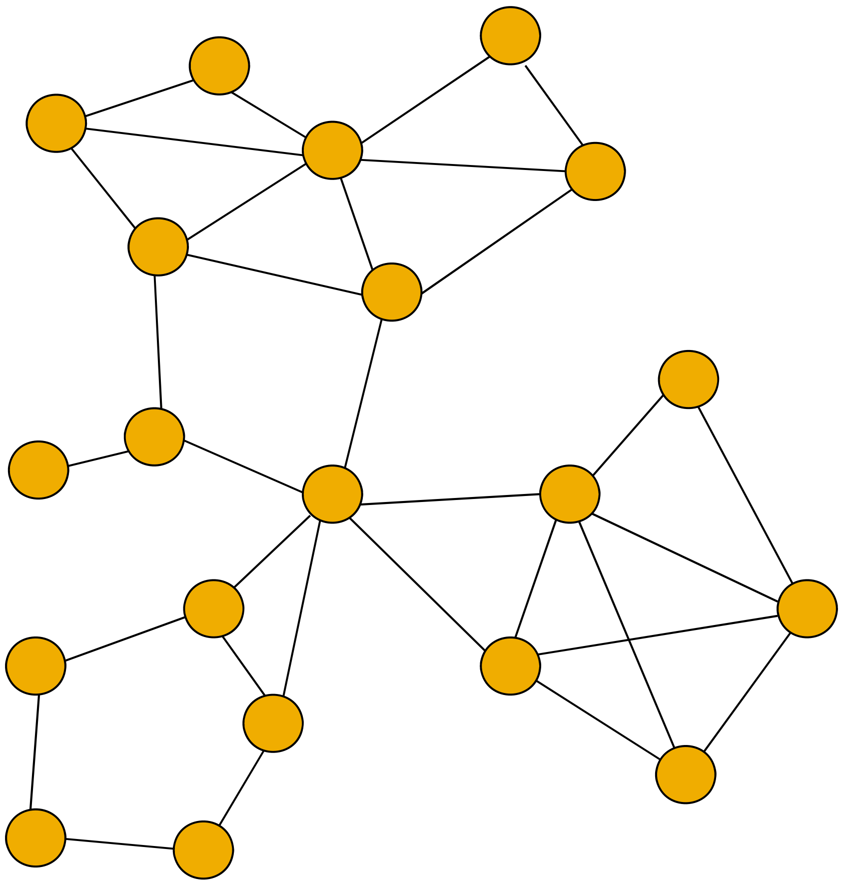
Φ(S) = # edges cut / # edges inside

Small Φ(S) corresponds to good clusters

- What method to use to optimize the objective function?

Community score (quality)

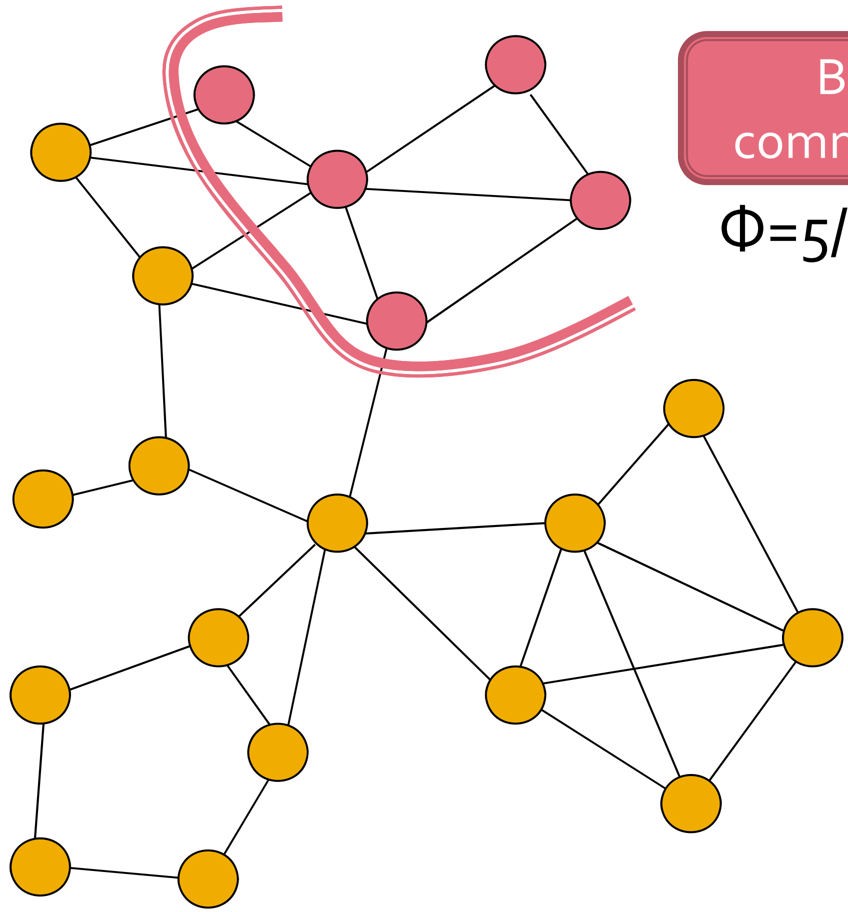
What is "best" community of 5 nodes?



Score: $\Phi(S)$ = **# edges cut / # edges inside**

Community score (quality)

What is "best" community of 5 nodes?



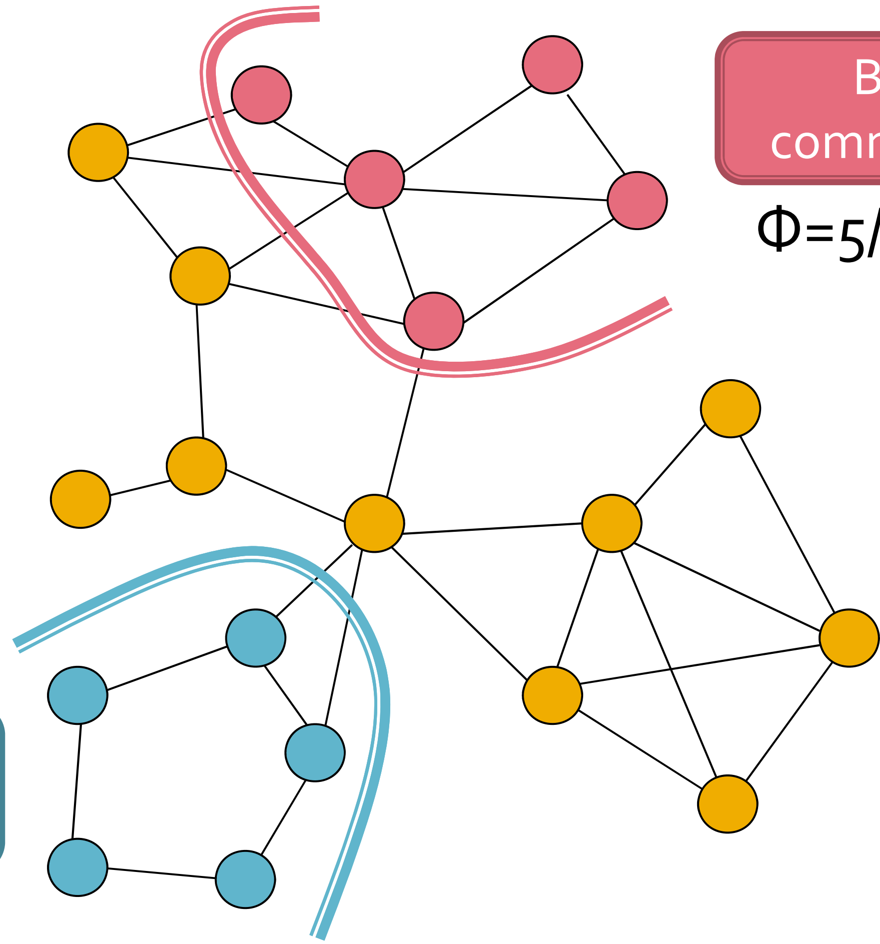
Bad community

Φ=5/6 = 0.83

Score: **Φ(S) = # edges cut / # edges inside**

Community score (quality)

What is "best" community of 5 nodes?



Bad community

Φ=5/7 = 0.7

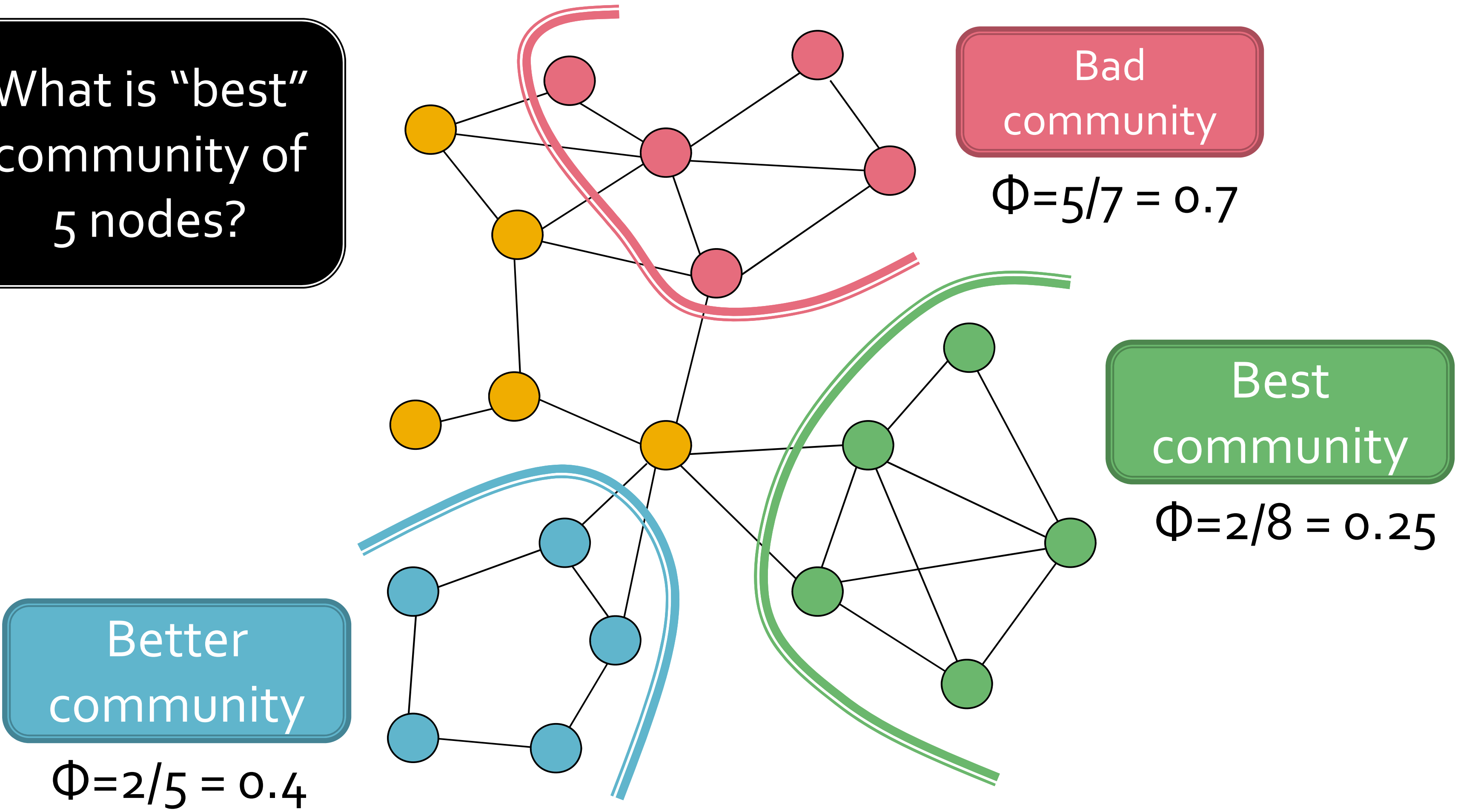
Better community

Φ=2/5 = 0.4

Score: Φ(S) = # edges cut / # edges inside

Community score (quality)

What is "best" community of 5 nodes?



Score: $\Phi(S)$ = # edges cut / # edges inside

Quantifying community structure

- So far we defined the measure that quantifies how cluster-like is a set of nodes

- Now, we want to define a measure of how expressed are the clusters in the network overall

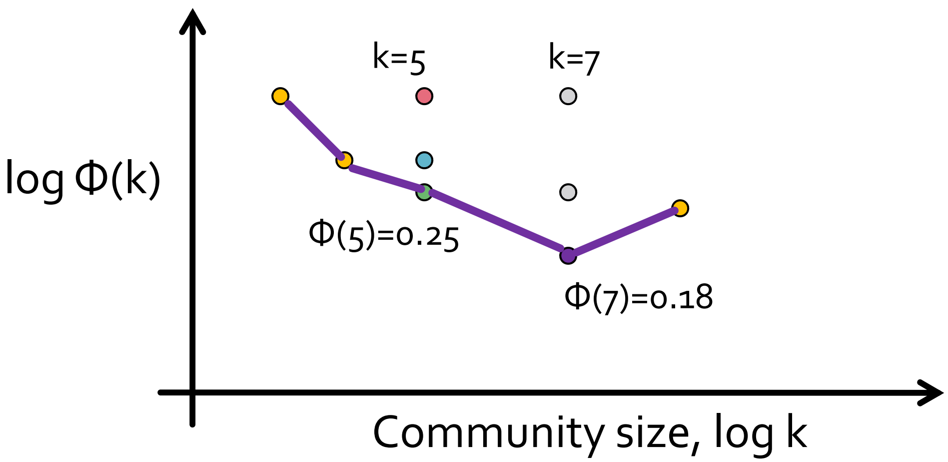
[Leskovec et al. WWW '08]

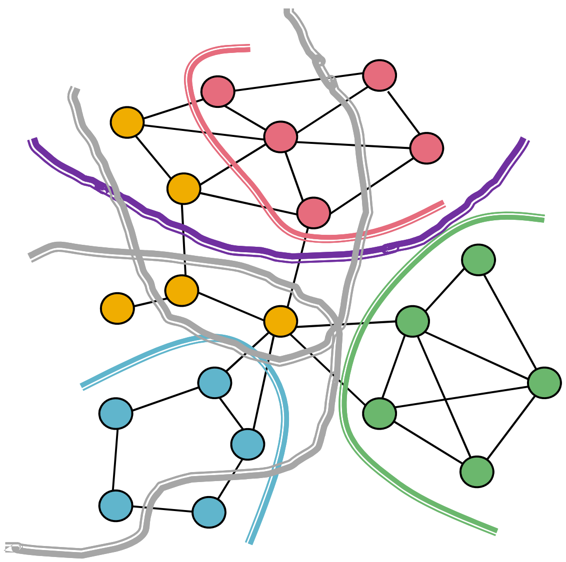
Network Community Profile Plot

- <u>Define:</u>
  Network community profile (**NCP**) plot

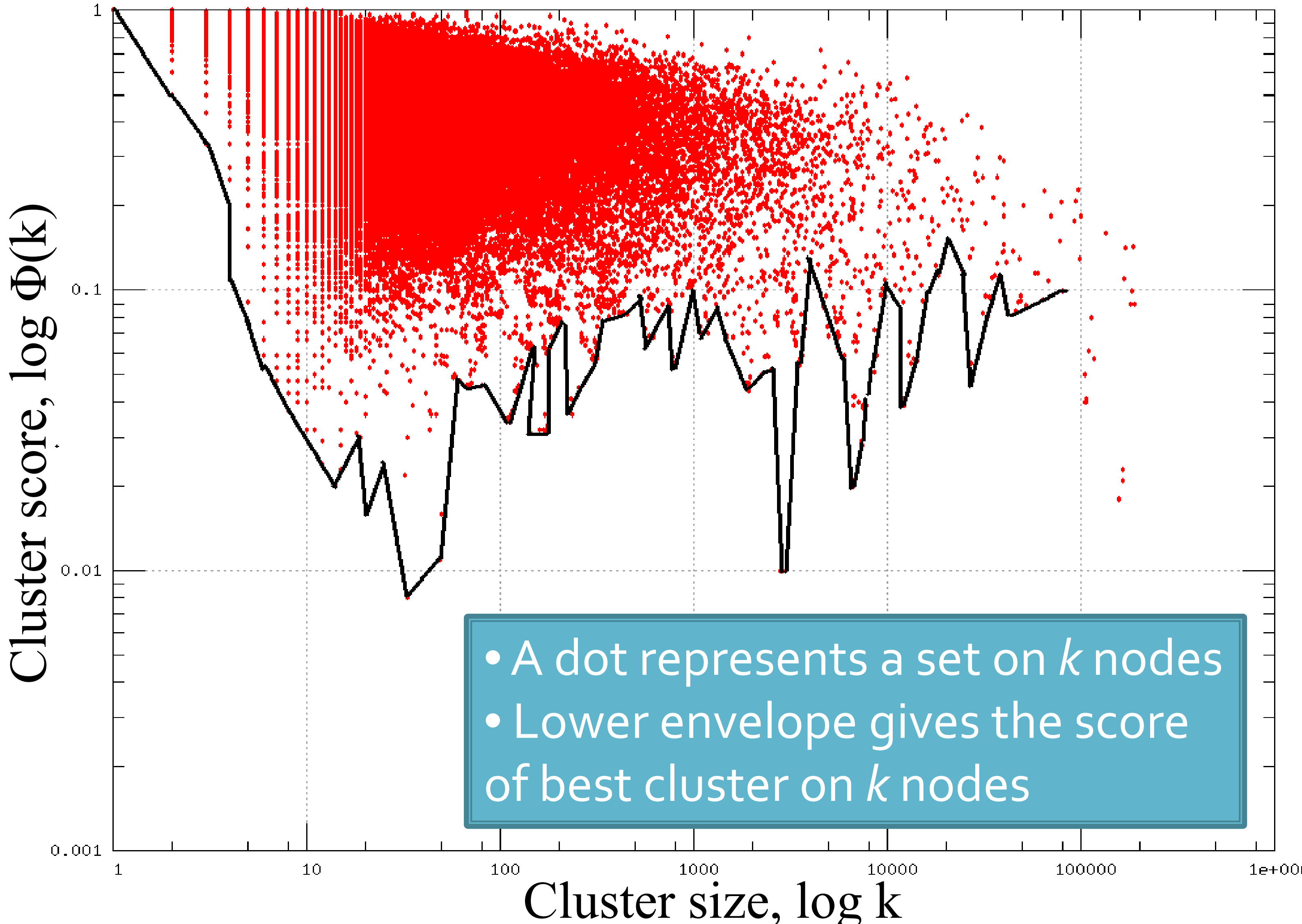
  Plot the score of best community of size *k*

$$\Phi(k) = \min_{S \subset V, |S|=k} \phi(S)$$





Network Community Profile Plot



- A dot represents a set on *k* nodes
- Lower envelope gives the score of best cluster on *k* nodes

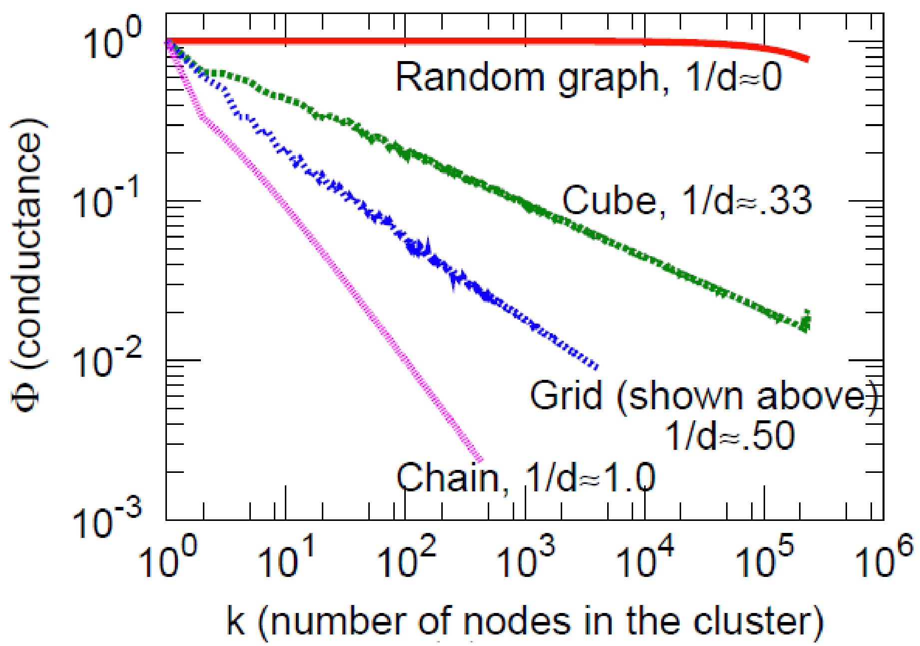
[Leskovec et al. WWW '08]

Probing networks

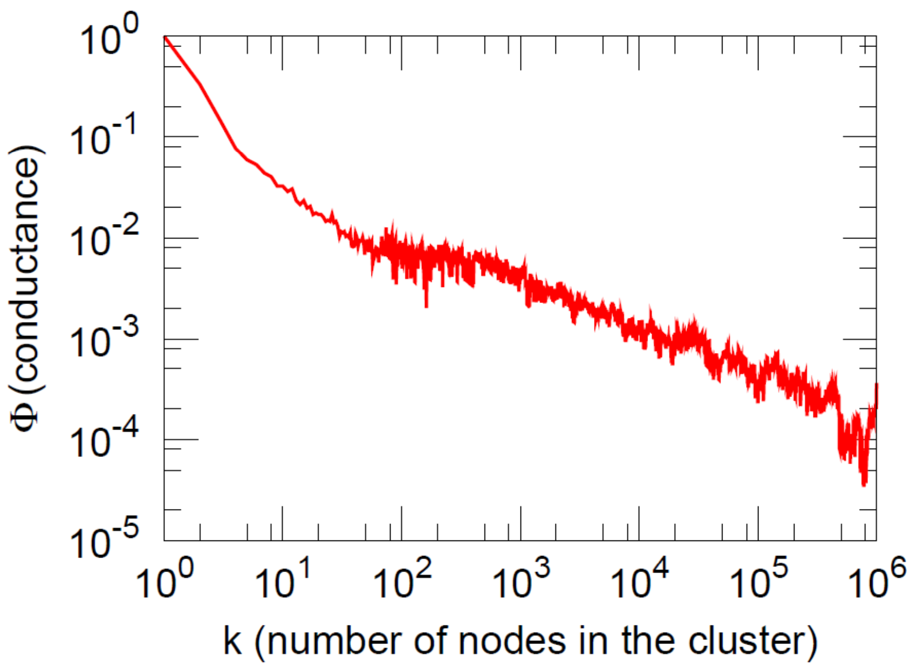
- Computing NCP is NP-hard
- Use algorithms to extract clusters:
  - Spectral clustering
  - Girvan-Newman/Modularity optimization: popular heuristics
  - Metis (multi-resolution heuristic): common in practice [Karypis
  - Local Spectral - connected and tighter sets [Andersen-Chung 07]

NCP plot: Meshes

- Meshes, grids, dense random graphs:



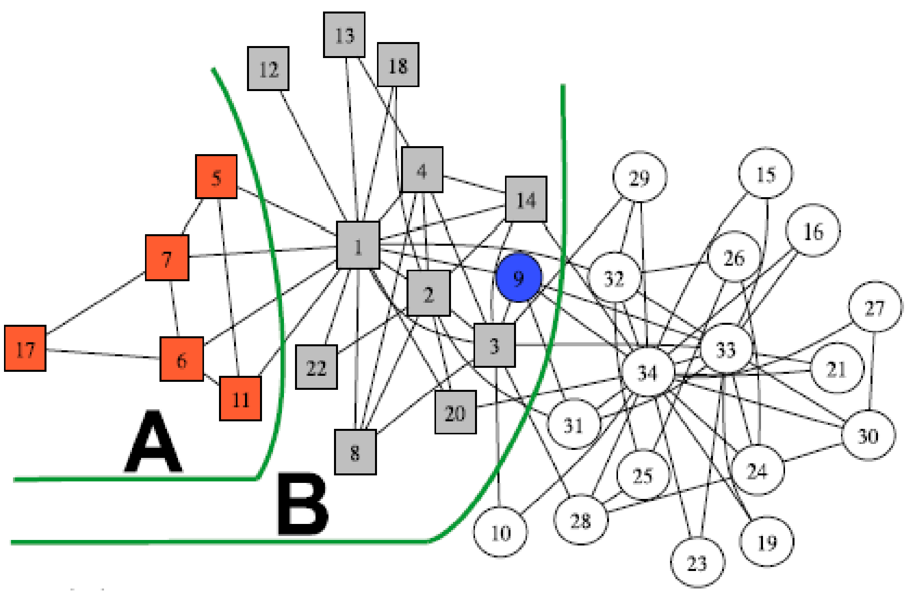
d-dimensional meshes

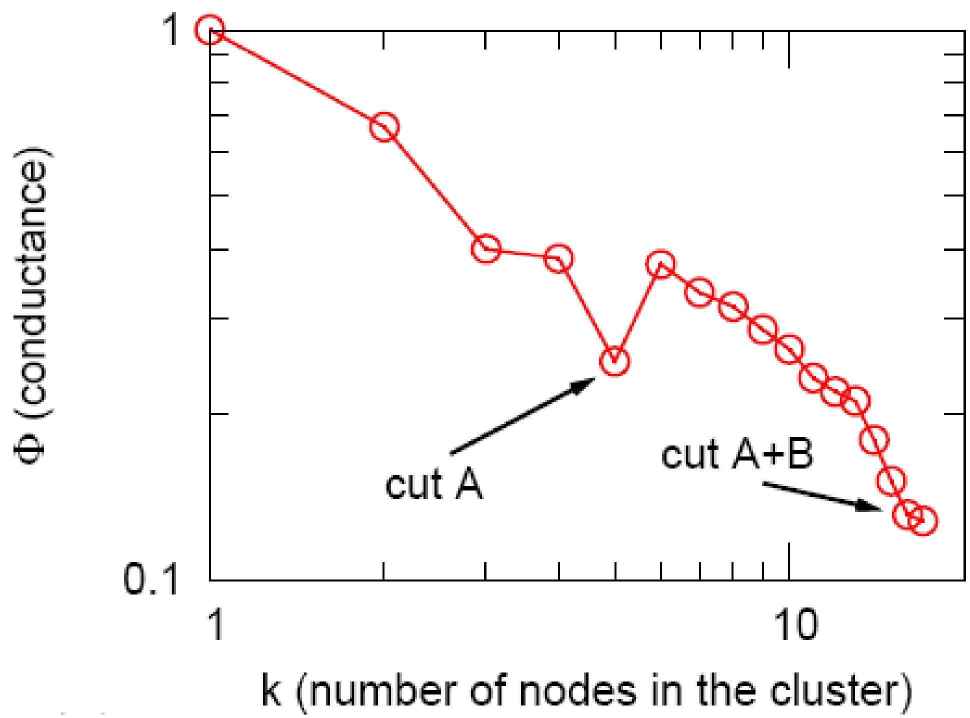


California road network

NCP plot: Small social networks

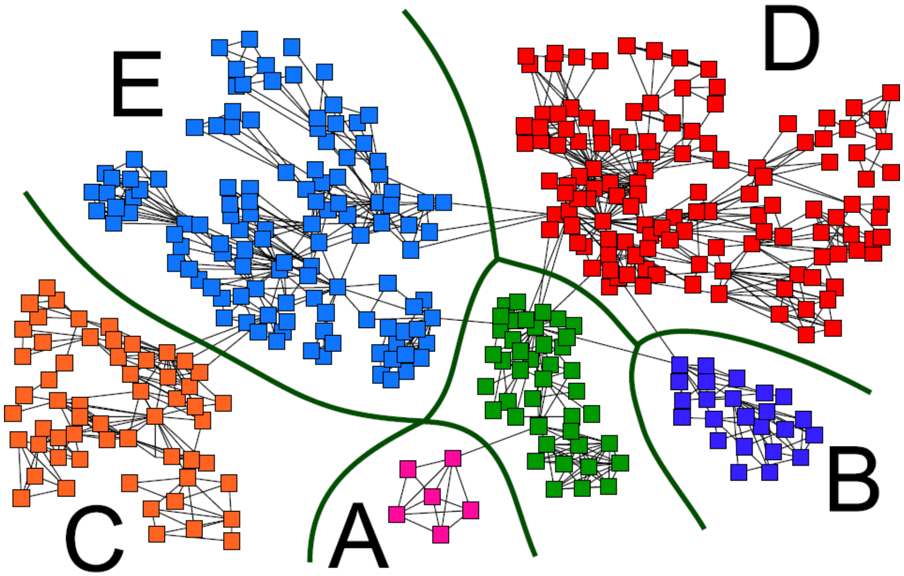
- Zachary's university karate club social network

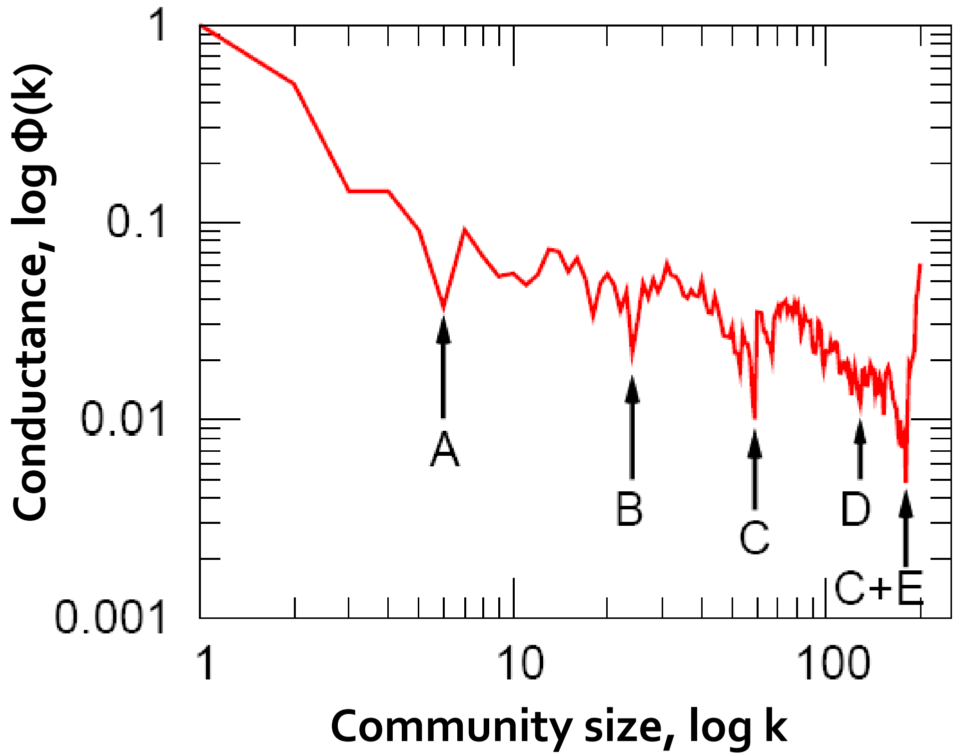




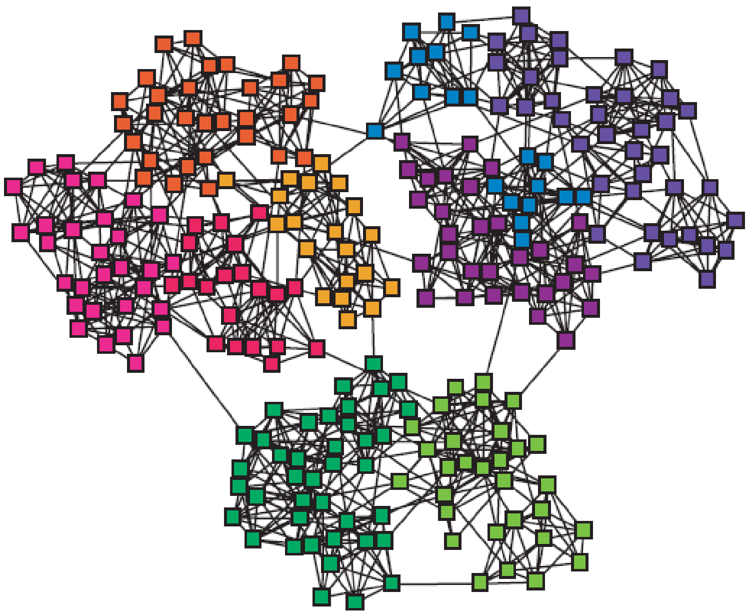
NCP plot: Network Science

- Collaborations between scientists in networks [Newman, 2005]

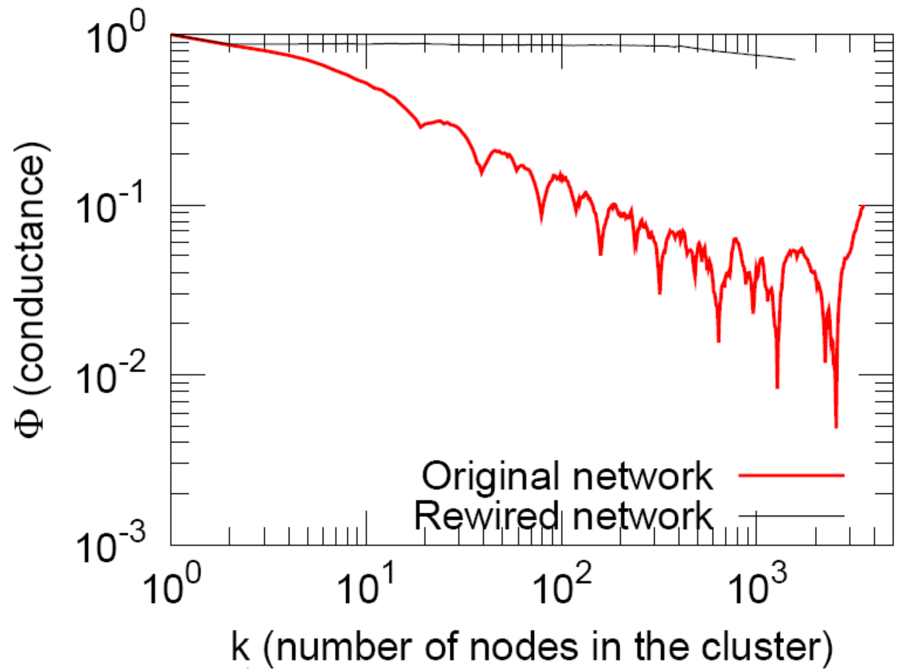




NCP plot: Hierarchical networks



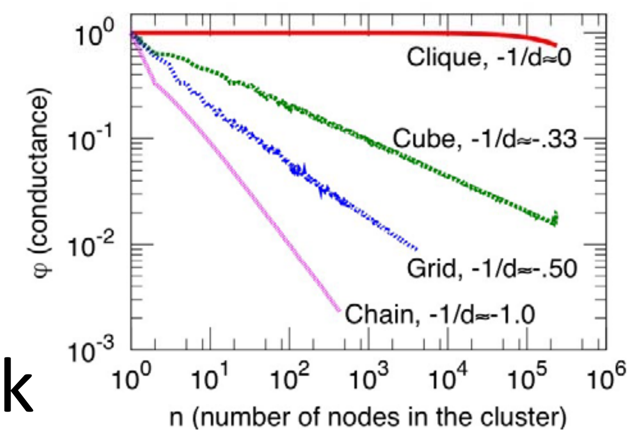
[Clauset-Moore-Newman 08]



Natural hypothesis

Natural hypothesis about NCP:

- NCP of real networks slopes downward
- Slope of the NCP corresponds to the dimensionality of the network



What about large networks?

• Social nets	Nodes	Edges	Description
LiveJournal	4,843,953	42,845,684	Blog friendships [5]
Epinions	75,877	405,739	Trust network [28]
CA-DBLP	317,080	1,049,866	Co-authorship [5]
• Information (citation) networks			
Cit-hep-th	27,400	352,021	Arxiv hep-th [14]
AmazonProd	524,371	1,491,793	Amazon products [8]
• Web graphs			
Web-google	855,802	4,291,352	Google web graph
Web-wt10g	1,458,316	6,225,033	TREC WT10G
• Bipartite affiliation (authors-to-papers) networks			
Atp-DBLP	615,678	944,456	DBLP [21]
AtM-Imdb	2,076,9		
• Internet networks			
AsSkitter	1,719,0		
Gnutella	62,5		

We examined more than **100 large networks**

[Leskovec et al. WWW '08]

Large networks: Very different

Typical example: General Relativity collaborations (n=4,158, m=13,422)