Network Community Detection

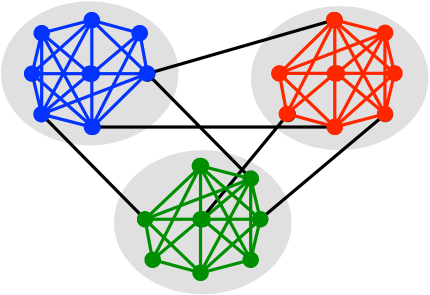
Gonzalo Mateos
Dept. of ECE and Goergen Institute for Data Science
University of Rochester
gmateosb@ece.rochester.edu
http://www.ece.rochester.edu/~gmateosb/

March 3, 2020

Community structure in networks

Communities within networks

- Networks play the powerful role of bridging the local and the global
  - $\Rightarrow$ Explain how processes at node/link level ripple to a population
- We often think of (social) networks as having the following structure



- Q: Can we gain insights behind this conceptualization?

Motivating context

- In the 60s., M. Granovetter interviewed people who changed jobs
  - Asked about how they discovered their new jobs
  - Many learned about opportunities through personal contacts
- Surprisingly, contacts where often acquaintances rather than friends

  $\Rightarrow$ Close friends likely have the most motivation to help you out
- Q: Why do distant acquaintances convey the crucial information?
- M. Granovetter, *Getting a job: A study of contacts and careers.* University of Chicago Press, 1974

Granovetter's answer and impact

- Linked two different perspectives on distant friendships
  - Structural: focus on how friendships span the network
  - Interpersonal: local consequences of friendship being strong or weak
- Intertwining between structural and informational role of an edge

1) Structurally-embedded edges within a community:

   $\Rightarrow$ Tend to be socially strong; and

   $\Rightarrow$ Are highly redundant in terms of information access

2) Long-range edges spanning different parts of the network:

   $\Rightarrow$ Tend to be socially weak; and

   $\Rightarrow$ Offer access to useful information (e.g., on a new job)

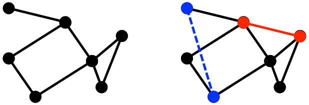
- General way of thinking about the architecture of social networks

  $\Rightarrow$ Answer transcends the specific setting of job-seeking

- A basic principle of network formation is that of triadic closure

  "If two people have a friend in common, then there is an increased likelihood that they will become friends in the future"

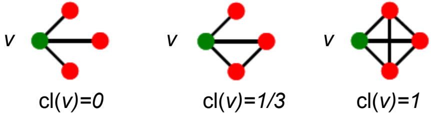
- Emergent edges in a social network likely to close triangles



$\Rightarrow$ More likely to see the red edge than the blue one

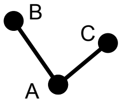
- Prevalence of triadic closure measured by the clustering coefficient

$$cl(v) = \frac{\#\text{pairs of friends of v that are connected}}{\#\text{pairs of friends of v}} = \frac{\#\triangle \text{ involving v}}{d_v(d_v - 1)/2}$$



Reasons for triadic closure

- Triadic closure is intuitively very natural. Reasons why it operates:



1) Increased opportunity for B and C to meet

   $\Rightarrow$ Both spend time with A

2) There is a basis for mutual trust among B and C

   $\Rightarrow$ Both have A as a common friend

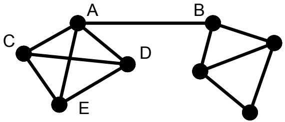
3) A may have an incentive to bring B and C together

   $\Rightarrow$ Lack of friendship may become a source of latent stress

- Premise based on theories dating to early work in social psychology
- F. Heider, *The Psychology of Interpersonal Relations.* Wiley, 1958

Bridges

- Ex: Consider the simple social network in the figure



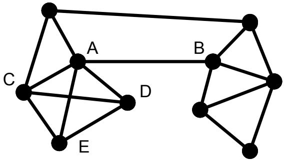
- A's links to C,D, and E connect her to a tightly knit group
  $\Rightarrow$ A,C,D, and E likely exposed to similar opinions

- A's link to B seems to reach to a different part of the network
  $\Rightarrow$ Offers her access to views she would otherwise not hear about

- A-B edge is called a bridge, its removal disconnects the network
  $\Rightarrow$ Giant components suggest that bridges are quite rare

Local bridges

- Ex: In reality, the social network is larger and may look as



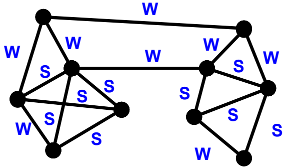
$\Rightarrow$ Without A, B knowing, may have a longer path among them

- **Def:** Span of (u, v) is the $u - v$ distance when the edge is removed
- **Def:** A local bridge is and edge with span > 2

  $\Rightarrow$ Ex: Edge A-B is a local bridge with span 3
- Local bridges with large spans $\approx$ bridges, but less extreme

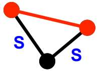
  $\Rightarrow$ Link with triadic closure: local bridges not part of triangles

- Categorize all edges in the network according to their strength
  - ⇒ Strong ties corresponding to friendship
  - ⇒ Weak ties corresponding to acquaintances



- Opportunity, trust, incentive act more powerfully for strong ties
  - ⇒ Suggests qualitative assumption termed strong triadic closure

"Two strong ties implies a third edge exists closing the triangle"



- Abstraction to reason about consequences of strong/weak ties

Local bridges and weak ties

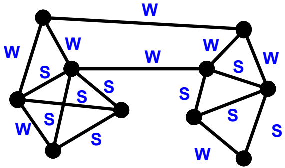
a) Local, interpersonal distinction between edges $\Rightarrow$ strong/weak ties

b) Global, structural notion $\Rightarrow$ local bridges present or absent

Theorem

If a node satisfies the strong triadic closure property and is involved in at least two strong ties, then any local bridge incident to it is a weak tie.

- Links structural and interpersonal perspectives on friendships

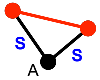


- Back to job-seeking, local bridges connect to new information
  - $\Rightarrow$ Conceptual span is related to their weakness as social ties
  - $\Rightarrow$ Surprising dual role suggests a "strength of weak ties"

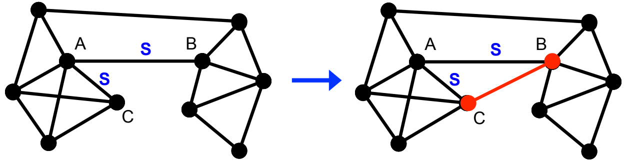
Proof by contradiction

Proof.

- We will argue by contradiction. Suppose node A has 2 strong ties
- Moreover, suppose A satisfies the strong triadic closure property



- Let A-B be a local bridge as well as a strong tie



$\Rightarrow$ Edge B-C must exist by strong triadic closure

- This contradicts A-B is a local bridge (cannot be part of a triangle)

$\square$

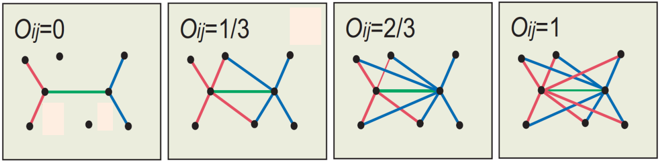
Tie strength and structure in large-scale data

- Q: Can one test Granovetter's theory with real network data?
  - $\Rightarrow$ Hard for decades. Lack of large-scale social interaction surveys
- Now we have "who-calls-whom" networks with both key ingredients
  - $\Rightarrow$ Network structure of communication among pairs of people
  - $\Rightarrow$ Total talking time, i.e., a proxy for tie strength
- Ex: Cell-phone network spanning $\approx 20\%$ of country's population
- J. P. Onella et al., "Structure and tie strengths in mobile communication networks," *PNAS*, vol. 104, pp. 7332-7336, 2007

Generalizing weak ties and local bridges

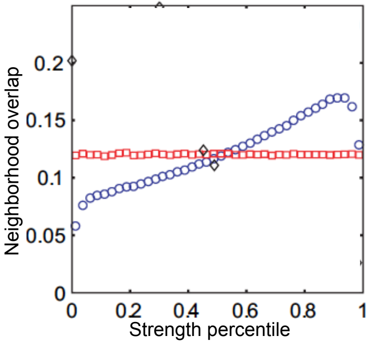
- Model described so far imposes sharp dichotomies on the network
  - $\Rightarrow$ Edges are either strong or weak, local bridges or not
  - $\Rightarrow$ Convenient to have proxies exhibiting smoother gradations
- Numerical tie strength $\Rightarrow$ Minutes spent in phone conversations
  - $\Rightarrow$ Order edges by strength, report their percentile occupancy
- Generalize local bridges $\Rightarrow$ Define neighborhood overlap of edge (i, j)

$$O_{ij} = \frac{|n(i) \cap n(j)|}{|n(i) \cup n(j)|}; \quad n(i) := \{j \in V : (i, j) \in E\}$$



$\Rightarrow$ Desirable property: $O_{ij} = 0$ if (i, j) is a local bridge

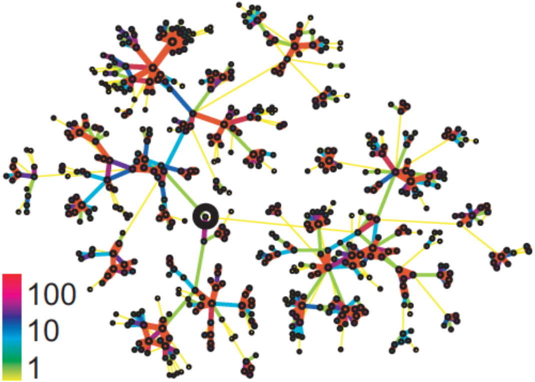
- Strength of weak ties prediction: O_{ij} grows with tie strength
  - $\Rightarrow$ Dependence borne out very cleanly by the data (o points)



- Randomly permuted tie strengths, fixed network structure (□ points)
  - $\Rightarrow$ Effectively removes the coupling between O_{ij} and tie strength

Phone network and tie strengths

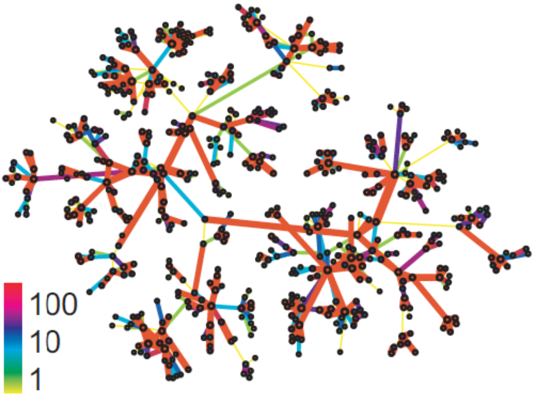
- Cell-phone network with color-coded tie strengths



1) Stronger ties more structurally-embedded (within communities)
2) Weaker ties correlate with long-range edges joining communities

Randomly permuted tie strengths

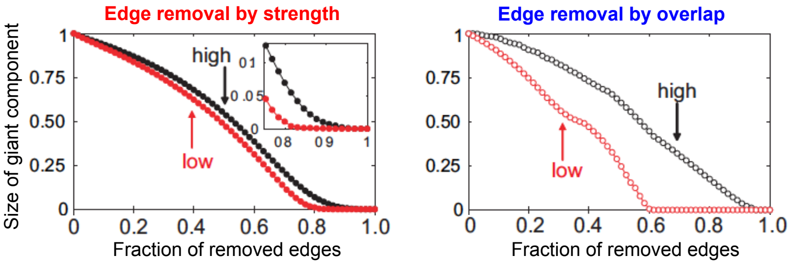
- Same cell-phone network with randomly permuted tie strengths



- No apparent link between structural and interpersonal roles of edges

Weak ties linking communities

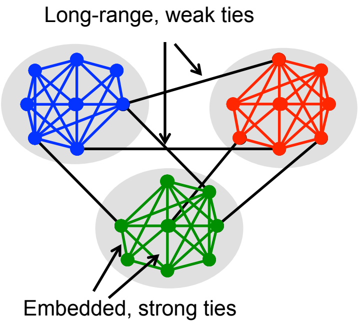
- Strength of weak ties prediction: long-range, weak ties bridge communities

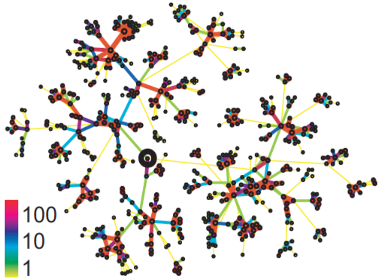


- Delete decreasingly weaker (small overlap) edges one at a time
  - $\Rightarrow$ Giant component shrinks rapidly, eventually disappears
- Repeat with strong-to-weak tie deletions $\Rightarrow$ slower shrinkage observed

Closing the loop

- We often think of (social) networks as having the following structure





- Conceptual picture supported by Granovetter's strength of weak ties

Network communities

Community structure in networks

Examples of network communities

Network community detection

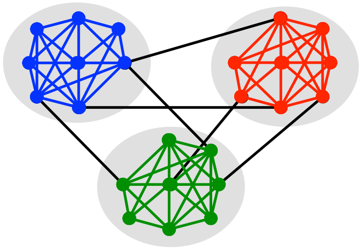
Modularity maximization

Spectral graph partitioning

Communities

- Nodes in real-world networks organize into communities
  Ex: families, clubs, political organizations, proteins by function, ...

- Supported by Granovetter's strength of weak ties theory

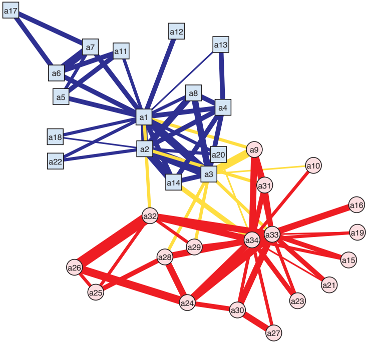


- Community (a.k.a. group, cluster, module) members are:
  $\Rightarrow$ Well connected among themselves
  $\Rightarrow$ Relatively well separated from the rest

- Exhibit high cohesiveness w.r.t. the underlying relational patterns

Zachary's karate club

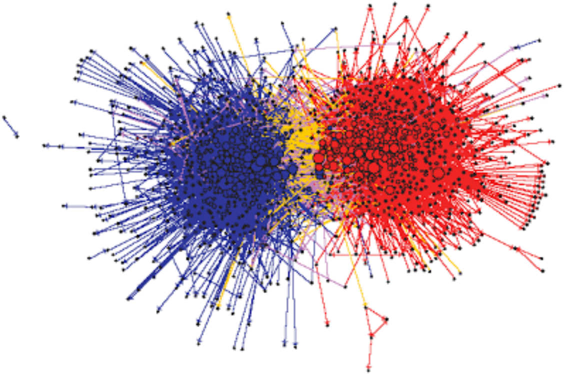
- Social interactions among members of a karate club in the 70s



- Zachary witnessed the club split in two during his study
  - ⇒ Toy network, yet canonical for community detection algorithms
  - ⇒ Offers "ground truth" community membership (a rare luxury)

Political blogs

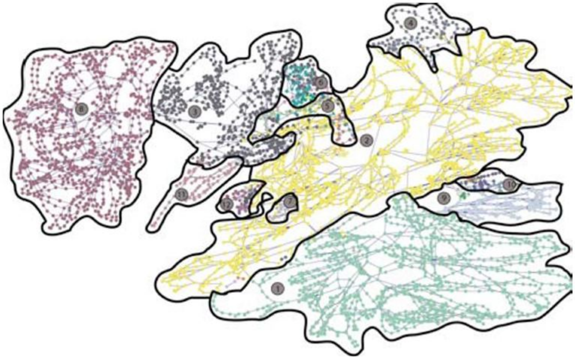
- The political blogosphere for the US 2004 presidential election



- Community structure of liberal and conservative blogs is apparent
  - $\Rightarrow$ People have a stronger tendency to interact with "equals"

Electrical power grid

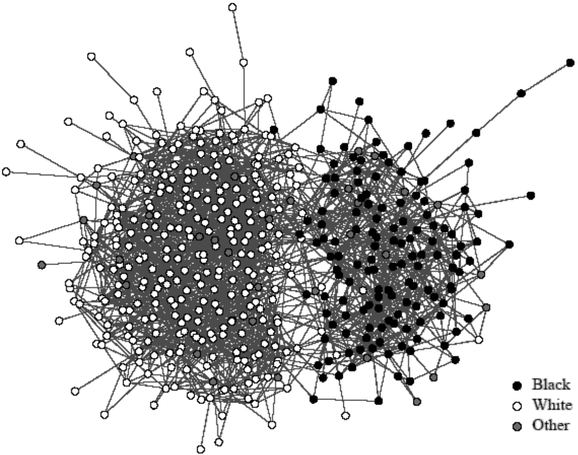
- Split power network into areas with minimum inter-area interactions



- **Applications:**
  - Decide control areas for distributed power system state estimation
  - Parallel computation of power flow
  - Controlled islanding to prevent spreading of blackouts

High-school students

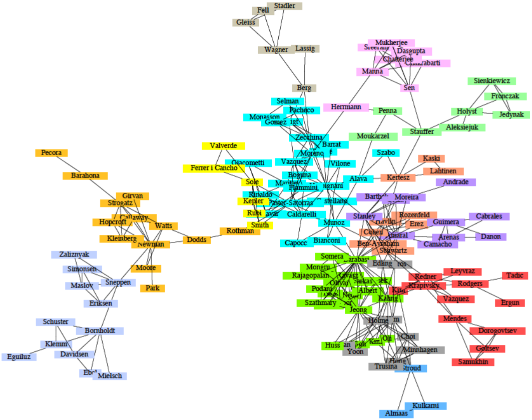
- Network of social interactions among high-school students



- Strong assortative mixing, with race as latent characteristic

Physicists working on Network Science

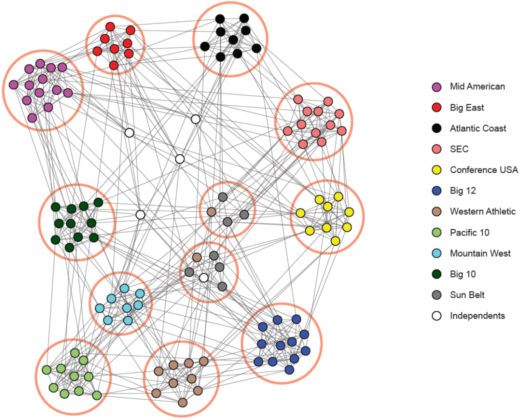
- Coauthorship network of physicists publishing networks' research



- Tightly-knit subgroups are evident from the network structure

College football

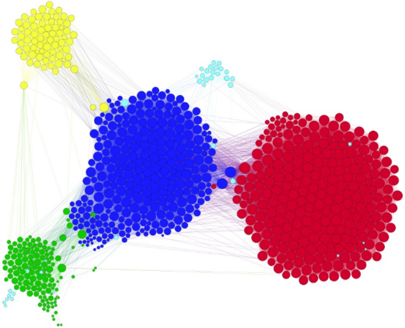
- Vertices are NCAA football teams, edges are games during Fall'00



- Communities are the NCAA conferences and independent teams

Facebook friendships

- Facebook egonet with 744 vertices and 30K edges



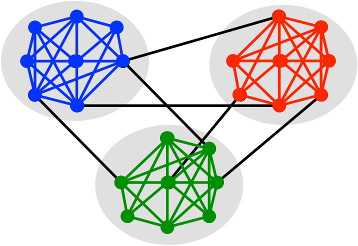
- Asked "ego" to identify social circles to which friends belong
  - ⇒ Company, high-school, basketball club, squash club, family

Network community detection

Unveiling network communities

- Nodes in real-world networks organize into communities

  Ex: families, clubs, political organizations, proteins by function, . . .



- Community (a.k.a. group, cluster, module) members are:

  ⇒ Well connected among themselves

  ⇒ Relatively well separated from the rest

- Exhibit high cohesiveness w.r.t. the underlying relational patterns

- Q: How can we automatically identify such cohesive subgroups?

Community detection and graph partitioning

- Community detection is a challenging clustering problem
  - C1) No consensus on the structural definition of community
  - C2) Node subset selection often intractable
  - C3) Lack of ground-truth for validation

- Useful for exploratory analysis of network data

  Ex: clues about social interactions, content-related web pages

Graph partitioning

Split V into given number of non-overlapping groups of given sizes

- **Criterion:** number of edges between groups is minimized (more soon)

  Ex: task-processor assignment for load balancing

- Number and sizes of groups unspecified in community detection

  $\Rightarrow$ Identify the natural fault lines along which a network separates

Graph partitioning is hard

- Ex: Graph bisection problem, i.e., partition V into two groups
  - Suppose the groups V_1 and V_2 are non-overlapping
  - Suppose groups have equal size, i.e., $|V_1| = |V_2| = N_v/2$
  - Minimize edges running between vertices in different groups
- Simple problem to describe, but hard to solve

$$\text{Number of ways to partition } V: \quad \binom{N_v}{N_v/2} \approx \frac{2^{N_v}}{\sqrt{N_v}}$$

$\Rightarrow$ Used Stirling's formula $N_v! \approx \sqrt{2\pi N_v}(N_v/e)^{N_v}$

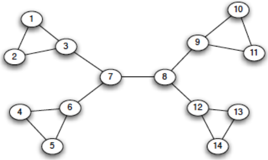
$\Rightarrow$ Exhaustive search intractable beyond toy small-sized networks

- No smart (i.e., polynomial time) algorithm, NP-hard problem

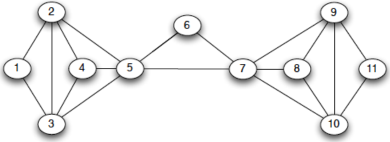
$\Rightarrow$ Seek good heuristics, e.g., relaxations of natural criteria

Strength of weak ties motivation

- Local bridges connect weakly interacting parts of the network



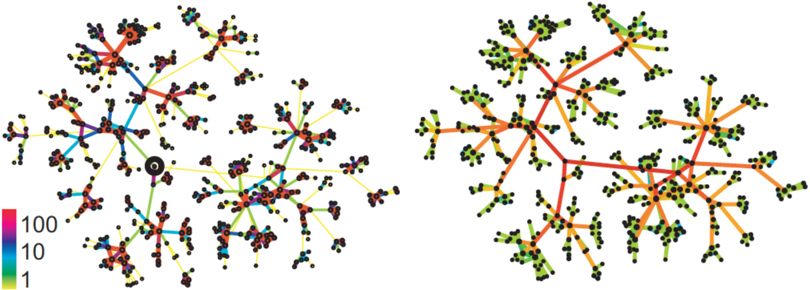
- Q: What about removing those to reveal communities?



- **Challenges**
  - Multiple local bridges. Some better that others? Which one first?
  - There might be no local bridge, yet an apparent natural division

Edge betweenness centrality

- Idea: high edge betweenness centrality to identify weak ties
  - High $c_{Be}(e)$ edges carry large traffic volume over shortest paths
  - Position at the interface between tightly-knit groups
- Ex: cell-phone network with colored edge strength and betwenness

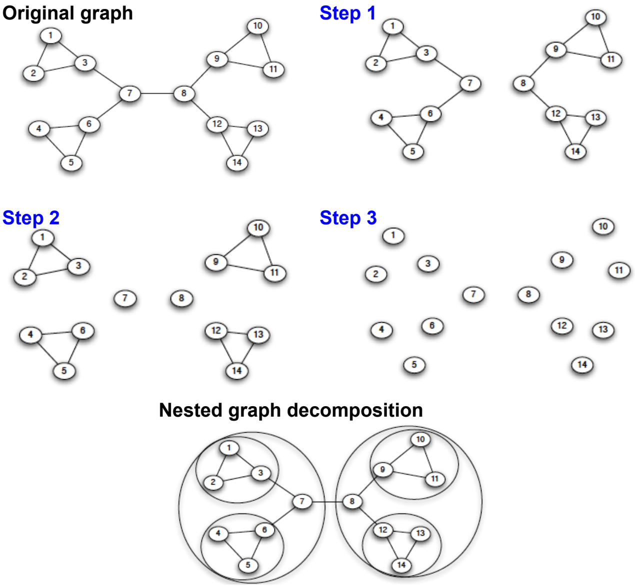


Edge strength　　　　Edge betweenness

Girvan-Newman's method

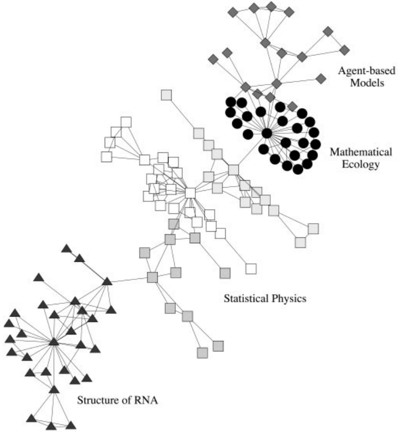
- Girvan-Newmann's method extremely simple conceptually
  - $\Rightarrow$ Find and remove "spanning links" between cohesive subgroups
- **Algorithm:** Repeat until there are no edges left
  - $\Rightarrow$ Calculate the betweenness centrality $c_{Be}(e)$ of all edges
  - $\Rightarrow$ Remove edge(s) with highest $c_{Be}(e)$
- Connected components are the communities identified
  - Divisive method: network falls apart into pieces as we go
  - Nested partition: larger communities potentially host denser groups
  - Recompute edge betweenness in $O(N_v N_e)$-time per step
- M. Girvan and M. Newman, "Community structure in social and biological networks," *PNAS*, vol. 99, pp. 7821-7826, 2002

Example: The algorithm in action



Scientific collaboration network

- Ex: Coauthorship network of scientists at the Santa Fe Institute



- Communities found can be traced to different disciplines

Hierarchical clustering

- Greedy approach to iteratively modify successive candidate partitions
  - Agglomerative: successive coarsening of partitions through merging
  - Divisive: successive refinement of partitions through splitting
- Per step, partitions are modified in a way that minimizes a cost
  - Measures of (dis)similarity x_{ij} between pairs of vertices v_i and v_j
  - Ex: Euclidean distance dissimilarity

$$x_{ij} = \sqrt{\sum_{k \neq i,j} (A_{ik} - A_{jk})^2}$$

- Method returns an entire hierarchy of nested partitions of the graph

  $\Rightarrow$ Can range fully from $\{\{v_1\}, \ldots, \{v_{N_v}\}\}$ to V

Agglomerative clustering

- An agglomerative hierarchical clustering algorithm proceeds as follows

  S1: Choose a dissimilarity metric and compute it for all vertex pairs

  S2: Assign each vertex to a group of its own

  S3: Merge the pair of groups with smallest dissimilarity

  S4: Compute the dissimilarity between the new group and all others

  S5: Repeat from S3 until all vertices belong to a single group

- Need to define group dissimilarity from pairwise vertex counterparts
  - Single linkage: group dissimilarity $x^{SL}_{G_i,G_j}$ follows single most dissimilar pair

$$x^{SL}_{G_i,G_j} = \max_{u \in G_i, v \in G_j} x_{uv}$$

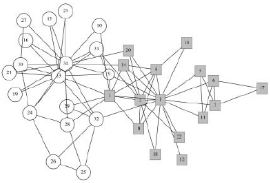
  - Complete linkage: every vertex pair highly dissimilar to have high $x^{CL}_{G_i,G_j}$

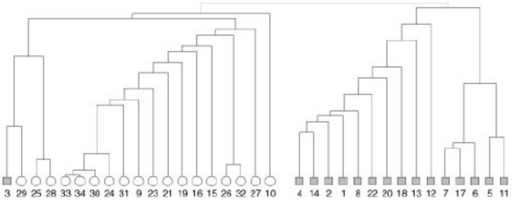
$$x^{CL}_{G_i,G_j} = \min_{u \in G_i, v \in G_j} x_{uv}$$

Dendrogram

- Hierarchical partitions often represented with a dendrogram
- Shows groups found in the network at all algorithmic steps

  $\Rightarrow$ Split the network at different resolutions
- Ex: Girvan-Newman's algorithm for the Zachary's karate club





- Q: Which of the divisions is the most useful/optimal in some sense?
- A: Need to define metrics of graph clustering quality

Modularity maximization

Modularity

- Size of communities typically unknown $\Rightarrow$ Identify automatically
- Modularity measures how well a network is partitioned in communities
  - Intuition: density of edges in communities higher than expected
- Consider a graph G and a partition into groups $s \in S$. Modularity:

$$Q(G,S) \propto \sum_{s\in S} [(\#\text{ of edges within group } s) - \mathbb{E}\left[\#\text{ of such edges}\right]]$$

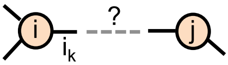
- Formally, after normalization such that $Q(G,S) \in [-1,1]$

$$Q(G,S) = \frac{1}{2N_e} \sum_{s\in S} \sum_{i,j\in s} \left[A_{ij} - \frac{d_i d_j}{2N_e}\right]$$

$\Rightarrow$ Null model: randomize edges, preserving degree distribution

Expected connectivity among nodes

- Null model: randomize edges preserving degree distribution in G
  - $\Rightarrow$ Random variable $A_{ij} := \mathbb{I}\left\{(i,j) \in E\right\}$
  - $\Rightarrow$ Expectation is $\mathbb{E}\left[A_{ij}\right] = \mathrm{P}\left((i,j) \in E\right)$
- Suppose node i has degree d_i, node j has degree d_j
  - $\Rightarrow$ Degree is "# of spokes" per node, $2N_e$ spokes in G



- Probability spoke i_k connected to j is $\frac{d_j}{2N_e-1} \approx \frac{d_j}{2N_e}$, hence

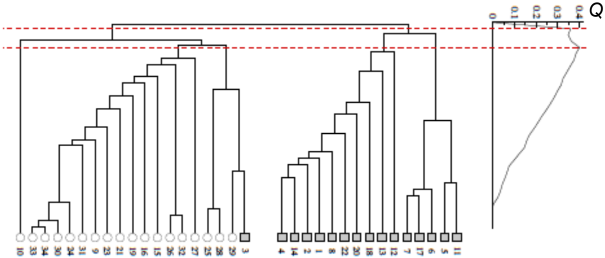
$$\mathrm{P}\left((i,j) \in E\right) = \mathrm{P}\left(\bigcup_{i_k=1}^{d_i} \{\text{spoke } i_k \text{ connected to } j\}\right)$$

$$= \sum_{i_k=1}^{d_i} \mathrm{P}\left(\text{spoke } i_k \text{ connected to } j\right) = \frac{d_i d_j}{2N_e}$$

- Can evaluate the modularity of each partition in a dendrogram

  $\Rightarrow$ Maximum value gives the "best" community structure

- Ex: Girvan-Newman's algorithm for the Zachary's karate club



- Q: Why not optimize $Q(G, S)$ directly over possible partitions S?

Modularity revisited

- Recall our definition of modularity

$$Q(G,S) = \frac{1}{2N_e}\sum_{s\in S}\sum_{i,j\in s}\left[A_{ij} - \frac{d_i d_j}{2N_e}\right]$$

- Let g_i be the group membership of vertex i, and rewrite

$$Q(G,S) = \frac{1}{2N_e}\sum_{i,j\in V}\left[A_{ij} - \frac{d_i d_j}{2N_e}\right]\mathbb{I}\left\{g_i = g_j\right\}$$

- Define for convenience the summands $B_{ij} := A_{ij} - \frac{d_i d_j}{2N_e}$

  $\Rightarrow$ Both marginal sums of B_{ij} vanish, since e.g.,

$$\sum_j B_{ij} = \sum_j A_{ij} - \frac{d_i}{2N_e}\sum_j d_j = d_i - \frac{d_i}{2N_e}2N_e = 0$$

Graph bisection

- Consider (for simplicity) dividing the network in two groups
- Binary community membership variables per vertex

$$s_i = \begin{cases} +1, & \text{vertex } i \text{ belongs to group 1} \\ -1, & \text{vertex } i \text{ belongs to group 2} \end{cases}$$

- Using the identity $\frac{1}{2}(s_i s_j + 1) = \mathbb{I}\{g_i = g_j\}$, the modularity is

$$Q(G,S) = \frac{1}{2N_e} \sum_{i,j \in V} \left[A_{ij} - \frac{d_i d_j}{2N_e} \right] \mathbb{I}\{g_i = g_j\}$$

$$= \frac{1}{4N_e} \sum_{i,j \in V} B_{ij}(s_i s_j + 1)$$

- Recall $\sum_j B_{ij} = 0$ to obtain the simpler expression

$$Q(G,S) = \frac{1}{4N_e} \sum_{i,j \in V} B_{ij} s_i s_j$$

Optimizing modularity

- Let $\mathbf{B} \in \mathbb{R}^{N_v \times N_v}$ be the modularity matrix with entries $B_{ij} := A_{ij} - \frac{d_i d_j}{2N_e}$

  $\Rightarrow$ Any partition S is defined by the vector $\mathbf{s} = [s_1, \ldots, s_{N_v}]^\top$

- Modularity is a quadratic form

$$Q(G, S) = \frac{1}{4N_e} \sum_{i,j \in V} B_{ij} s_i s_j = \frac{1}{4N_e} \mathbf{s}^\top \mathbf{B} \mathbf{s}$$

- Modularity as criterion for graph bisection yields the formulation

$$\hat{\mathbf{s}} = \arg \max_{\mathbf{s} \in \{\pm 1\}^{N_v}} \mathbf{s}^\top \mathbf{B} \mathbf{s}$$

  $\Rightarrow$ Nasty binary constraints $\mathbf{s} \in \{\pm 1\}^{N_v}$ (hypercube vertices)

  $\Rightarrow$ Modularity optimization is NP-hard [Brandes et al '06]

Just relax!

- Relax the constraint $\mathbf{s} \in \{\pm 1\}^{N_v}$ to $\mathbf{s} \in \mathbb{R}^{N_v}$, $\|\mathbf{s}\|_2 = \sqrt{N_v}$

$$\hat{\mathbf{s}} = \arg\max_{\mathbf{s}} \mathbf{s}^\top \mathbf{B}\mathbf{s}, \quad \text{s. to} \;\; \mathbf{s}^\top \mathbf{s} = N_v$$

- Associate a Langrange multiplier λ to the constraint $\mathbf{s}^\top \mathbf{s} = N_v$
  - $\Rightarrow$ Optimality conditions yields

$$\nabla_{\mathbf{s}} \left[\mathbf{s}^\top \mathbf{B}\mathbf{s} + \lambda(N_v - \mathbf{s}^\top \mathbf{s}) \right] = \mathbf{0} \Rightarrow \mathbf{B}\mathbf{s} = \lambda \mathbf{s}$$

- Conclusion is that $\mathbf{s}$ is an eigenvector of $\mathbf{B}$ with eigenvalue λ

- Q: Which eigenvector should we pick?
  - $\Rightarrow$ At optimum $\mathbf{B}\mathbf{s} = \lambda \mathbf{s}$ so objective becomes

$$\mathbf{s}^\top \mathbf{B}\mathbf{s} = \lambda \mathbf{s}^\top \mathbf{s} = \lambda$$

- A: To maximize modularity pick the dominant eigenvector of $\mathbf{B}$

Spectral modularity maximization

- Let $\mathbf{u}_1$ be the dominant eigenvector of $\mathbf{B}$, with i-th entry $[\mathbf{u}_1]_i$
  - $\Rightarrow$ Cannot just set $\mathbf{s} = \sqrt{N_v}\mathbf{u}_1$ because $\mathbf{u}_1 \neq \{\pm 1\}^{N_v}$
  - $\Rightarrow$ Best effort: maximize their similarity $\mathbf{s}^\top \mathbf{u}_1$ which gives

$$s_i = \text{sign}([\mathbf{u}_1]_i) := \begin{cases} +1, & [\mathbf{u}_1]_i > 0 \\ -1, & [\mathbf{u}_1]_i \leq 0 \end{cases}$$

- Spectral modularity maximization algorithm

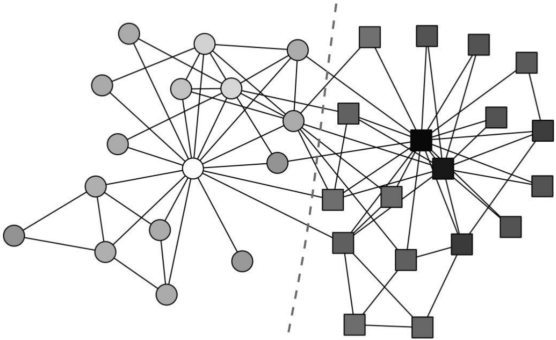
S1: Compute modularity matrix $\mathbf{B}$ with entries $B_{ij} = A_{ij} - \frac{d_i d_j}{2N_e}$

S2: Find dominant eigenvector $\mathbf{u}_1$ of $\mathbf{B}$ (e.g., power method)

S3: Cluster membership of vertex i is $s_i = \text{sign}([\mathbf{u}_1]_i)$

- Multiple (> 2) communities through e.g., repeated graph bisection

Example: Zachary's karate club

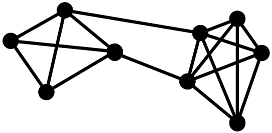


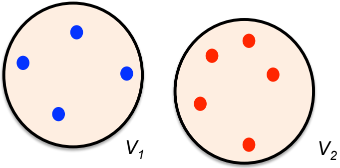
- Spectral modularity maximization
  - Shapes of vertices indicate community membership
  - Dotted line indicates partition found by the algorithm
  - Vertex colors indicate the strength of their membership

Spectral graph partitioning

Graph bisection

- Consider an undirected graph $G(V, E)$
- Ex: Graph bisection problem, i.e., partition V into two groups
  - Groups V_1 and $V_2 = V_1^C$ are non-overlapping
  - Groups have given size, i.e., $|V_1| = N_1$ and $|V_2| = N_2$

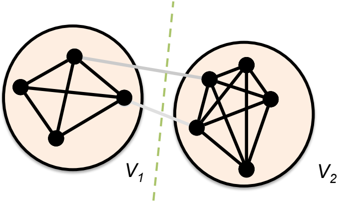




- Q: What is a good criterion to partition the graph?
- A: We have already seen modularity. Let's see a different one

Graph cut

- Desiderata: Community members should be
  - ⇒ Well connected among themselves; and
  - ⇒ Relatively well separated from the rest of the nodes



- **Def:** A cut C is the number of edges between groups V_1 and $V \setminus V_1$

$$C := \text{cut}(V_1, V_2) = \sum_{i \in V_1, j \in V_2} A_{ij}$$

- Natural criterion: minimize cut, i.e., edges across groups V_1 and V_2

From graph cuts ...

- Binary community membership variables per vertex

$$s_i = \begin{cases} +1, & \text{vertex } i \text{ belongs to } V_1 \\ -1, & \text{vertex } i \text{ belongs to } V_2 \end{cases}$$

- Let g_i be the group membership of vertex i, such that

$$\mathbb{I}\{g_i \neq g_j\} = \frac{1}{2}(1 - s_i s_j) = \begin{cases} 1, & i \text{ and } j \text{ in different groups} \\ 0, & i \text{ and } j \text{ in the same group} \end{cases}$$

- Cut expressible in terms of the variables s_i as

$$C = \sum_{i \in V_1, j \in V_2} A_{ij} = \frac{1}{2} \sum_{i,j \in V} A_{ij}(1 - s_i s_j)$$

...to the graph Laplacian matrix

- First summand in $C = \frac{1}{2}\sum_{i,j} A_{ij}(1 - s_i s_j)$ is

$$\sum_{i,j \in V} A_{ij} = \sum_{i \in V} d_i = \sum_{i \in V} d_i s_i^2 = \sum_{i,j \in V} d_i s_i s_j \mathbb{I}\left\{i = j\right\}$$

- Used $s_i^2 = 1$ since $s_i \in \{\pm 1\}$. The cut becomes

$$C = \frac{1}{2}\sum_{i,j \in V} \color{red}{(d_i \mathbb{I}\left\{i = j\right\} - A_{ij})} \color{black}{s_i s_j} = \frac{1}{2}\sum_{i,j \in V} \color{red}{L_{ij}} \color{black}{s_i s_j}$$

- Cut in terms of L_{ij}, entries of the graph Laplacian $\mathbf{L} = \mathbf{D} - \mathbf{A}$, i.e.,

$$C(\mathbf{s}) = \frac{1}{2}\mathbf{s}^\top \mathbf{L}\mathbf{s}, \quad \mathbf{s} := [s_1, \ldots, s_{N_v}]^\top$$

- Maximize modularity $Q(\mathbf{s}) \propto \mathbf{s}^\top \mathbf{B}\mathbf{s}$ vs. Minimze cut $C(\mathbf{s}) \propto \mathbf{s}^\top \mathbf{L}\mathbf{s}$

Graph cut minimization

- Since $|V_1| = N_1$ and $|V_2| = N_2 = N - N_1$, we have the constraint

$$\sum_{i \in V} s_i = \sum_{i \in V_1} (+1) + \sum_{i \in V_2} (-1) = N_1 - N_2 \;\Rightarrow\; \mathbf{1}^\top \mathbf{s} = N_1 - N_2$$

- Minimum-cut criterion for graph bisection yields the formulation

$$\hat{\mathbf{s}} = \arg \min_{\mathbf{s} \in \{\pm 1\}^{N_v}} \mathbf{s}^\top \mathbf{L} \mathbf{s}, \quad \text{s. to } \mathbf{1}^\top \mathbf{s} = N_1 - N_2$$

- Again, binary constraints $\mathbf{s} \in \{\pm 1\}^{N_v}$ render cut minimization hard
  $\Rightarrow$ Relax binary constraints as with modularity maximization

Laplacian matrix properties revisited

- Smoothness: For any vector $\mathbf{x} \in \mathbb{R}^{N_v}$ of "vertex values", one has

$$\mathbf{x}^\top \mathbf{L}\mathbf{x} = \sum_{i,j \in V} L_{ij} x_i x_j = \sum_{(i,j) \in E} (x_i - x_j)^2$$

  which can be minimized to enforce smoothness of functions on G

- Positive semi-definiteness: Follows since $\mathbf{x}^\top \mathbf{L}\mathbf{x} \geq 0$ for all $\mathbf{x} \in \mathbb{R}^{N_v}$

- Spectrum: All eigenvalues of $\mathbf{L}$ are real and non-negative

  $\Rightarrow$ Eigenvectors form an orthonormal basis of $\mathbb{R}^{N_v}$

- Rank deficiency: Since $\mathbf{L}\mathbf{1} = \mathbf{0}$, $\mathbf{L}$ is rank deficient

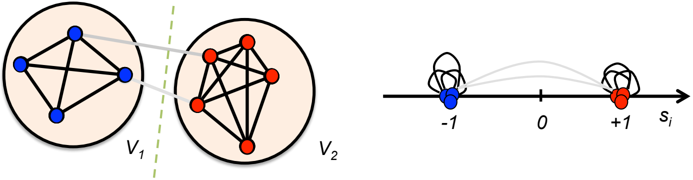
- Spectrum and connectivity: The smallest eigenvalue λ_1 of $\mathbf{L}$ is 0
  - If the second-smallest eigenvalue $\lambda_2 \neq 0$, then G is connected
  - If $\mathbf{L}$ has n zero eigenvalues, G has n connected components

Further intuition

- Since $\mathbf{s}^\top \mathbf{L}\mathbf{s} = \sum_{(i,j)\in E}(s_i - s_j)^2$, the minimum-cut formulation is

$$\hat{\mathbf{s}} = \arg \min_{\mathbf{s}\in\{\pm 1\}^{N_v}} \sum_{(i,j)\in E} (s_i - s_j)^2, \quad \text{s. to } \mathbf{1}^\top \mathbf{s} = N_1 - N_2$$

- Q: Does this equivalent cost function make sense? A: Absolutely!
  - $\Rightarrow$ Edges joining vertices in the same group do not add to the sum
  - $\Rightarrow$ Edges joining vertices in different groups add 4 to the sum



- Minimize cut: assign values s_i to nodes i such that few edges cross 0

Minimum-cut relaxation

- Relax the constraint $\mathbf{s} \in \{\pm 1\}^{N_v}$ to $\mathbf{s} \in \mathbb{R}^{N_v}$, $\|\mathbf{s}\|_2 = 1$

$$\hat{\mathbf{s}} = \arg\min_{\mathbf{s}} \mathbf{s}^\top \mathbf{L}\mathbf{s}, \quad \text{s. to } \mathbf{1}^\top \mathbf{s} = N_1 - N_2 \text{ and } \mathbf{s}^\top \mathbf{s} = 1$$

$\Rightarrow$ Straightforward to solve using Lagrange multipliers

- Characterization of the solution $\hat{\mathbf{s}}$ [Fiedler '73]:

$$\hat{\mathbf{s}} = \mathbf{v}_2 + \frac{N_1 - N_2}{N_v}\mathbf{1}$$

$\Rightarrow$ The 'second-smallest' eigenvector $\mathbf{v}_2$ of $\mathbf{L}$ satisfies $\mathbf{1}^\top \mathbf{v}_2 = 0$

$\Rightarrow$ Minimum cut is $C(\hat{\mathbf{s}}) = \hat{\mathbf{s}}^\top \mathbf{L}\hat{\mathbf{s}} = \mathbf{v}_2^\top \mathbf{L}\mathbf{v}_2 \propto \lambda_2$

- If the graph G is disconnected then we know $\lambda_2 = 0 = C(\hat{\mathbf{s}})$

$\Rightarrow$ If G is amenable to bisection, the cut is small and so is λ_2

Theoretical guarantee

- Consider a partition of G into V_1 and V_2, where $|V_1| \leq |V_2|$
- If G is connected, then the Cheeger inequality asserts

$$\frac{\alpha^2}{2d_{max}} \leq \lambda_2 \leq 2\alpha$$

where $\alpha = \frac{C}{|V_1|}$ and d_{max} is the maximum node degree

$\Rightarrow$ Certifies that λ_2 gives a useful bound

- F. Chung, "Four proofs for the Cheeger inequality and graph partition algorithms," *Proc. of ICCM*, 2007

Spectral graph bisection

- Q: How to obtain the binary cluster labels $\mathbf{s} \in \{\pm 1\}^{N_v}$ from $\hat{\mathbf{s}} \in \mathbb{R}^{N_v}$?

  $\Rightarrow$ Again, maximize the similarity measure $\mathbf{s}^\top \hat{\mathbf{s}}$

$$s_i = f(\mathbf{v}_2) := \begin{cases} +1, & [\mathbf{v}_2]_i \text{ among the } N_1 \text{ largest entries of } \mathbf{v}_2 \\ -1, & \text{otherwise} \end{cases}$$

- Spectral graph bisection algorithm

  S1: Compute Laplacian matrix $\mathbf{L}$ with entries $L_{ij} = D_{ij} - A_{ij}$

  S2: Find 'second smallest' eigenvector $\mathbf{v}_2$ of $\mathbf{L}$

  S3: Candidate membership of vertex i is $\bar{s}_i = f([\mathbf{v}_2])$ (or $\underline{s}_i = f([-\mathbf{v}_2])$)

  S4: Among $\bar{\mathbf{s}}$ and $\underline{\mathbf{s}}$ pick the one that minimizes $C(\mathbf{s})$

- Complexity: efficient Lanczos algorithm variant in $O(\frac{N_e}{\lambda_3 - \lambda_2})$ time

- Nomenclature: $\mathbf{v}_2$ is known as the Fiedler vector

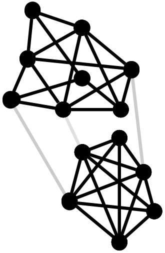
  $\Rightarrow$ Eigenvalue λ_2 is Fiedler value, or algebraic connectivity of G

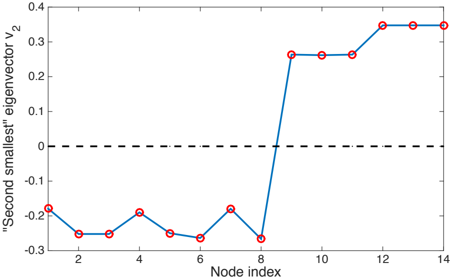
Spectral gap in Fiedler vector entries

- Suppose G is disconnected and has two connected components
  - $\mathbf{L}$ is block diagonal, two smallest eigenvectors indicate groups, i.e.,

$$\mathbf{v}_1 = [1, 1, \ldots, 1, 0, \ldots, 0]^\top \text{ and } \mathbf{v}_2 = [0, 0, \ldots, 0, 1, \ldots, 1]^\top$$

- If G is connected but amenable to bisection, $\mathbf{v}_1 = \mathbf{1}$ and $\lambda_2 \approx 0$
  - Also, $\mathbf{1}^\top \mathbf{v}_2 = \sum_i [\mathbf{v}_2]_i = 0 \;\Rightarrow$ Positive and negative entries in $\mathbf{v}_2$

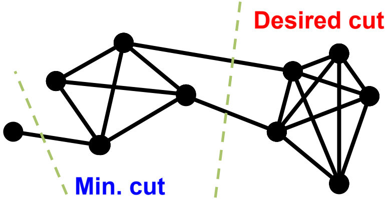




Unknown community sizes

- Consider the graph bisection problem with unknown group sizes

  $\Rightarrow$ Minimizing the graph cut may be no longer meaningful!



  $\Rightarrow$ Cost $C := \sum_{i \in V_1, j \in V_2} A_{ij}$ agnostic to groups' internal structure

- Better criterion is the ratio cut R defined as

$$R := \frac{C}{|V_1|} + \frac{C}{|V_2|}$$

  $\Rightarrow$ Balanced partitions: small community is penalized by the cost

Ratio-cut minimization

- Fix a bisection S of G into groups V_1 and V_2
- Define $\mathbf{f} : \mathbf{f}(S) = [f_1, \ldots, f_{N_v}]^\top \in \mathbb{R}^{N_v}$ with entries

$$f_i = \begin{cases} \sqrt{\frac{|V_2|}{|V_1|}}, & \text{vertex } i \text{ belongs to } V_1 \\ -\sqrt{\frac{|V_1|}{|V_2|}}, & \text{vertex } i \text{ belongs to } V_2 \end{cases}$$

- One can establish the following properties:

  P1: $\mathbf{f}^\top \mathbf{L} \mathbf{f} = N_v R(S)$;
  P2: $\sum_i f_i = 0$, i.e., $\mathbf{1}^\top \mathbf{f} = 0$; and
  P3: $\|\mathbf{f}\|^2 = N_v$

- From P1-P3 it follows that ratio-cut minimization is equivalent to

$$\min_{\mathbf{f}} \mathbf{f}^\top \mathbf{L} \mathbf{f}, \quad \text{s. to } \mathbf{1}^\top \mathbf{f} = 0 \text{ and } \mathbf{f}^\top \mathbf{f} = N_v$$

Ratio cut and spectral graph bisection

- Ratio-cut minimization is also NP-hard. Relax to obtain

$$\hat{\mathbf{s}} = \arg\min_{\mathbf{s}\in\mathbb{R}^{N_v}} \mathbf{s}^\top \mathbf{L}\mathbf{s}, \quad \text{s. to } \mathbf{1}^\top \mathbf{s} = 0 \text{ and } \mathbf{s}^\top \mathbf{s} = N_v$$

- Partition $\hat{S}$ also given by the spectral graph bisection algorithm

  S1: Compute Laplacian matrix $\mathbf{L}$ with entries $L_{ij} = D_{ij} - A_{ij}$

  S2: Find 'second smallest' eigenvector $\mathbf{v}_2$ of $\mathbf{L}$

  S3: Cluster membership of vertex i is $s_i = \text{sign}([\mathbf{v}_2]_i)$

- Alternative criterion is the normalized cut NC defined as

$$NC = \frac{C}{vol(V_1)} + \frac{C}{vol(V_2)}, \quad vol(V_i) := \sum_{v\in V_i} d_v,\ i = 1, 2$$

$\Rightarrow$ Corresponds to using the normalized Laplacian $\mathbf{D}^{-1}\mathbf{L}$

Glossary

- Network community
- (Strong) triadic closure
- Clustering coefficient
- Bridges and local bridges
- Tie strength
- Neighborhood overlap
- Strength of weak ties
- Zachary's karate club
- Community detection
- Graph partitioning and bisection
- Non-overlapping communities
- Edge betweenness centrality
- Girvan-Newmann method
- Hierarchical clustering
- Dendrogram
- Single and complete linkage
- Modularity
- Spectral modularity maximization
- Modularity and Laplacian matrices
- Minimum-cut partitioning
- Fiedler vector and value
- Ratio-cut minimization