Spectral Clustering and Transitivity

Francisco Barreras

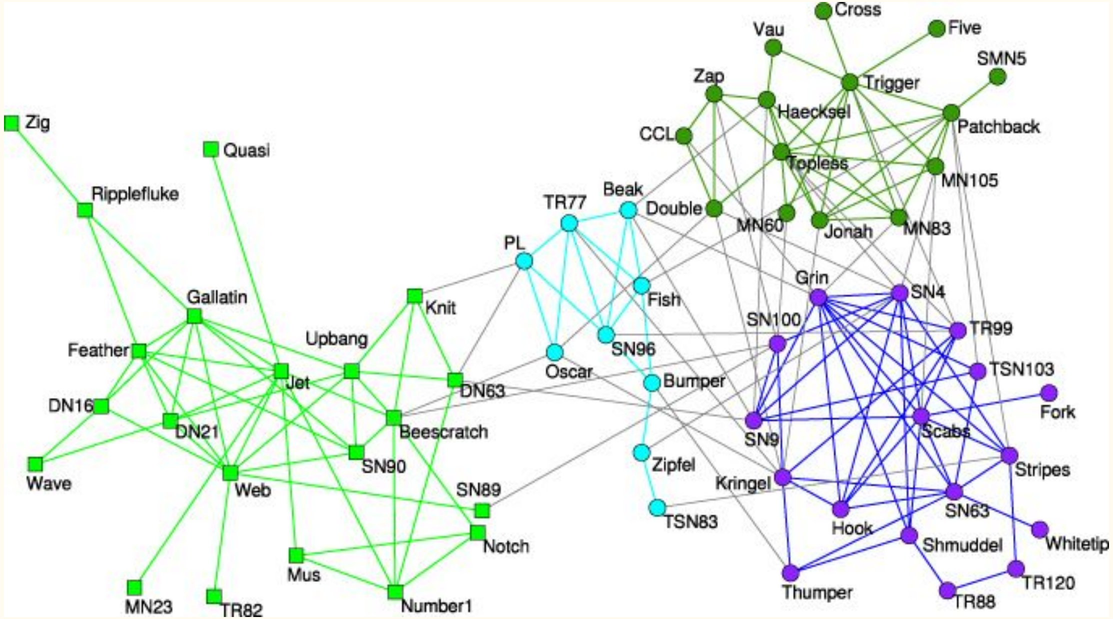
Different Approaches to Community Detection

What constitutes a community?

1. Balanced partition with min-cut.
2. Clusters (maximize density within)
3. Connectivity profile classes (SBM)
4. Groups where flows stay very long time ("units" for network dynamics)

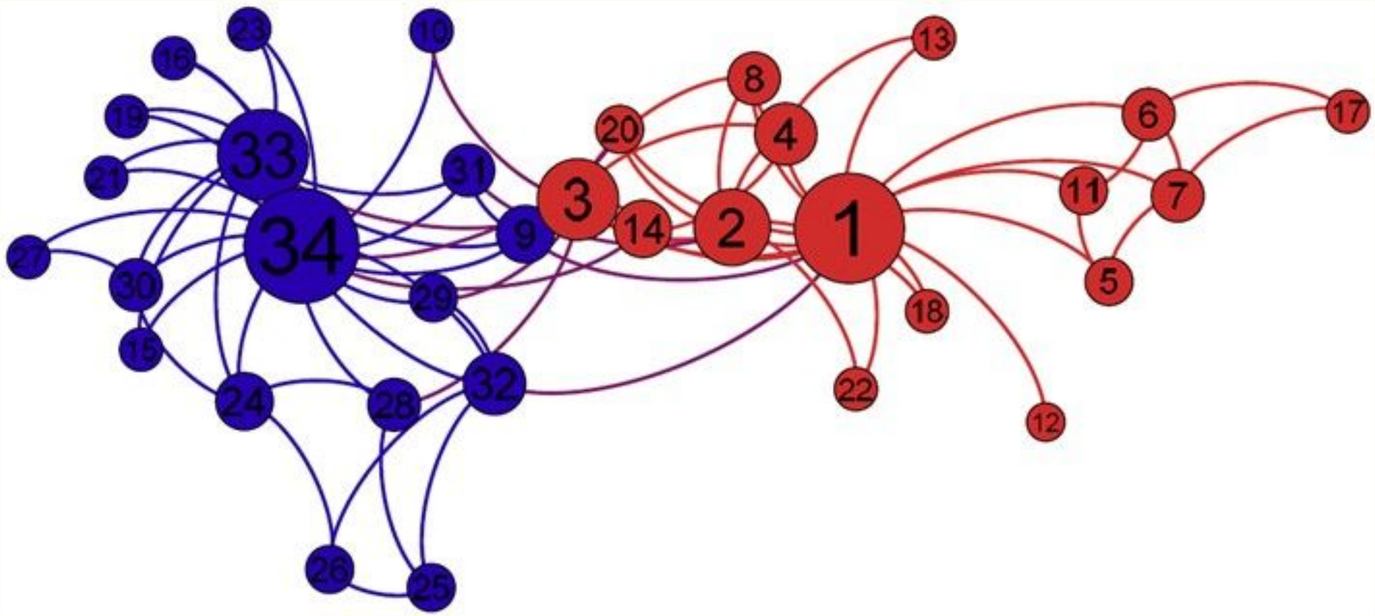
Networks with communities: Dolphins

- Undirected social network of frequent associations between 62 dolphins in a community living off Doubtful Sound, New Zealand. *D. Lusseau, K. Schneider, O. J. Boisseau, P. Haase, E. Slooten, and S. M. Dawson, Behavioral Ecology and Sociobiology 54, 396-405 (2003).*



Networks with communities: Karate

- Social network of friendships between 34 members of a karate club at a US university in the 1970s. *W. W. Zachary, An information flow model for conflict and fission in small groups, Journal of Anthropological Research 33, 452-473 (1977).*



Networks with communities: Adjective-Nouns

- Adjacency network of common adjectives and nouns in the novel David Copperfield by Charles Dickens. *M. E. J. Newman, Phys. Rev. E 74, 036104 (2006).*

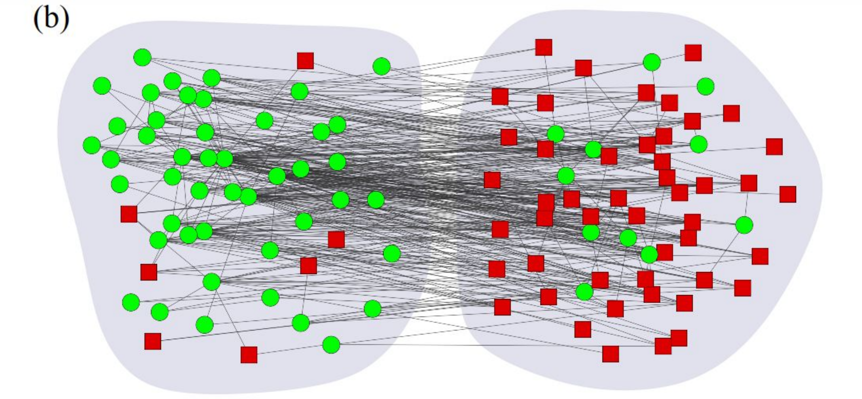


FIG. 7 (a) The network of commonly occurring English adjectives (circles) and nouns (squares) described in the text. (b) The same network redrawn with the nodes grouped so as to minimize the modularity of the grouping. The network is now revealed to be approximately bipartite, with one group consisting almost entirely of adjectives and the other of nouns.

Networks with communities: Political Blogs

- recent books on US politics, with edges connecting pairs of books that are frequently purchased by the same customers of the on-line bookseller Amazon.com. *compiled by V. Krebs (unpublished)*

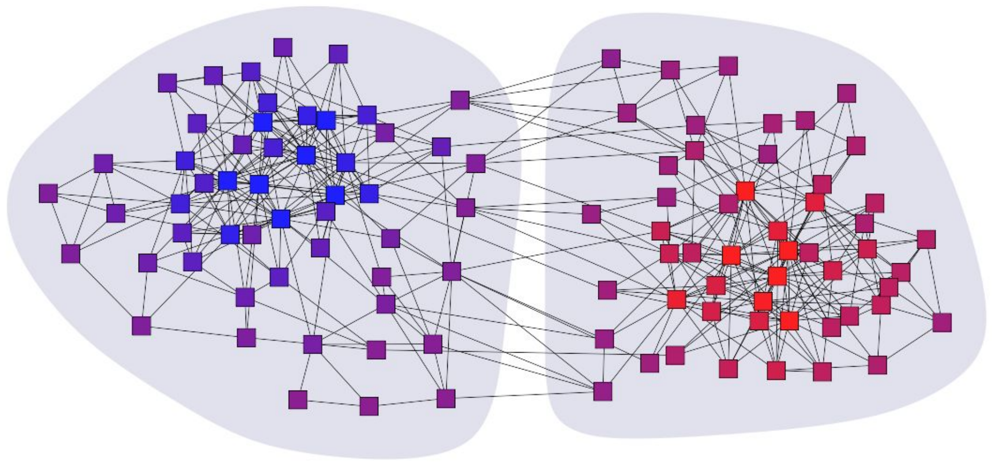


FIG. 3 The network of political books described in the text. Vertex colors range from blue to red to represent the values of the corresponding elements of the leading eigenvector of the modularity matrix.

In each example the definition of communities is different...

Is there a “**ground truth**”?

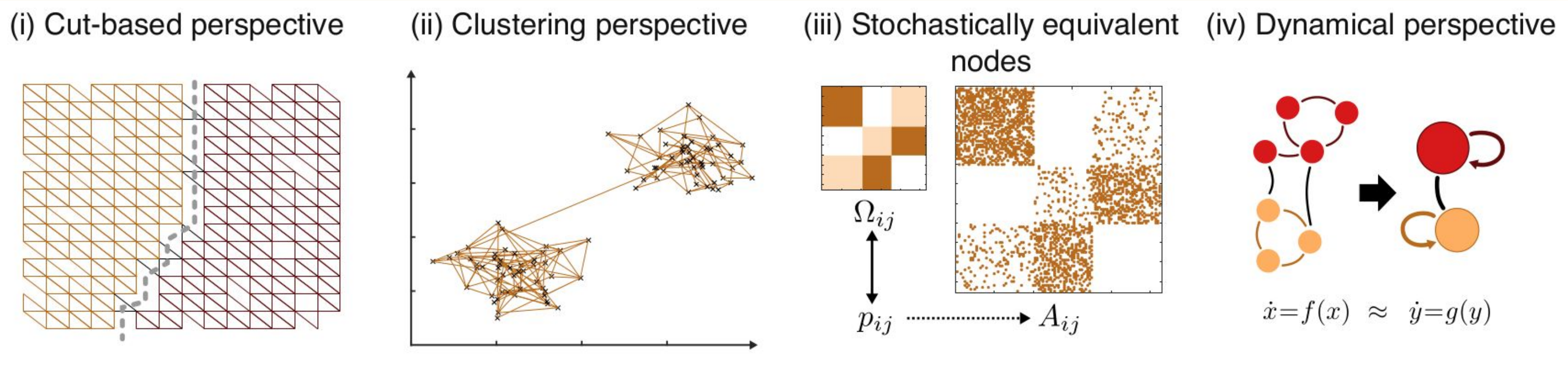
Quick recap of important concepts

- **Adjacency Matrix:** $A_{ij} = 1$
- **Laplacian Matrix:** $L = D - A$
- **Spectrum of a Matrix:** eigenvalues and eigenvectors, important because of stationary distribution and convergence to it.
- **Random Walk:** Stochastic process, in this case jumping from nodes to neighbors with homogeneous transition.
- **Stationary Distribution:** All random walks that are irreducible and aperiodic have a stationary distribution such that:

$$\pi P = \pi$$

- **Perron-Frobenius:** The largest eigenvalue is always real and corresponds to the stationary distribution. It has multiplicity 1.
- **Power Method:** $A^n w \to v_1$
- **High degree hub:** Fancy name for the neighborhood of a high degree node.
- **Erdos-Renyi random graph:** Simplest random graph. N nodes, and I connect each pair by flipping a coin.

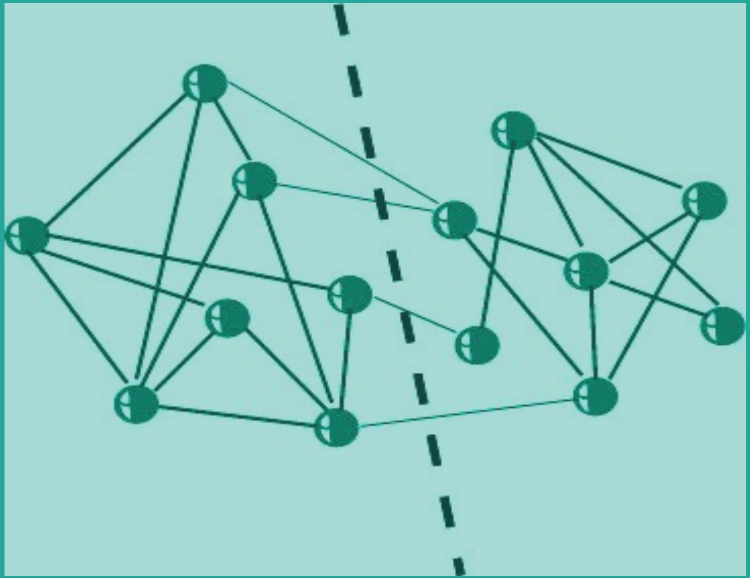
Min. cut | Statistical Inference | Max. Modularity | Longest mixing



- Ground-truth is **discretional** or at least dependent on the mathematical formulation for communities.
- However, all solutions are highly related.[1]

[1] “Spectral methods for network community detection and graph partitioning”, M.E.J. Newman

Min-cut approach



Graph Partition

We want to pick the partition that minimizes the edges that are "cut"...

- Long tradition in computer science.
- AplicatiDesign of Hardware and Distributed Computation.
- Can be solved exactly in polynomial time, unfortunately the complexity is:

$$o\left(n^{c^2}\right)$$

- Usually we want to approximate

Minimizing Ratio Cut

- The ratio cut for a given partition of the nodes, counts the edges that go across communities, penalizing small communities.

$$\min_{V} \sum_{V_u \in V} \frac{\mathrm{cut}(V_u, V_u^c)}{|V_u|}$$

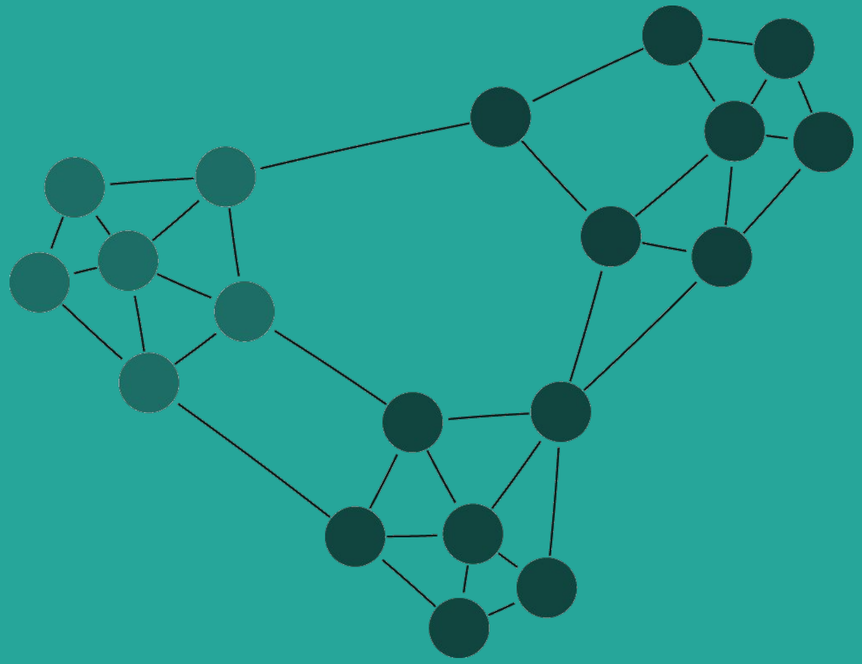
- Where the cut is defined as the number of edges across:

$$\mathrm{cut}(V, W) := \sum_{v \in V} \sum_{w \in W} A_{vw}$$

$$\min_{V} \sum_{V_u \in V} \frac{\mathrm{cut}(V_u, V_u^c)}{|V_u|} \sim \begin{array}{l} \min_{f} f^T L f \\ \mathrm{s.t.}\, f^T 1 = 0, \|f\| = \sqrt{n} \\ \text{where } f_u := \begin{cases} -\sqrt{|V_2|/|V_1|} & \text{if } u \in V_1 \\ \sqrt{|V_2|/|V_1|} & \text{if } u \in V_2 \end{cases} \end{array}$$

When we relax the second problem, it's a convex problem and the solution is related to the second (smallest) eigenvector of L. How?

Modularity approach



Considering Internal Density...

"Despite its evident success in the graph partitioning arena, spectral partitioning is a poor approach for detecting natural community structure in real-world networks [...]. The issue is with the condition that the **sizes of the groups** into which the network is divided be fixed. This condition is neither appropriate nor **realistic** for community detection problems" - M. E. J. Newman, *Finding community structure in networks using the eigenvectors of matrices*, 2006

- We want group sizes that are "**free**".
- Cut sizes are simply **not the right thing to optimize** because they don't accurately reflect our **intuitive concept of network communities**.

#Edges Within - $\mathbb{E}$(#Edges within)

- Minimizing the cut size doesn't really look into the **internal density** of the partition. Maximize that instead!
- Modularity and its variants can be seen as a trade-off between **cut-based** measures and **entropy**.

$$Q(\mathbf{V}) = \frac{1}{2m} \sum_{i=1}^{k} \sum_{u,v \in V_i} \left[A_{uv} - \frac{d_u d_v}{2m} \right]$$

- Modularity is the fraction of the edges that fall **within** the given groups minus the expected fraction **if edges were distributed at random** (with <u>fixed degree distribution</u> as in the configuration model).

Modularity Matrix and Spectral Methods

$$
\begin{aligned}
Q &= \frac{1}{2m}\sum_{ij}(A_{ij}-P_{ij})\,\delta(s_i,s_j)\\
&= \frac{1}{2m}\sum_{ij}(A_{ij}-P_{ij})\,\frac{1}{2}(s_is_j+1)\\
&= \frac{1}{4m}\sum_{ij}(A_{ij}-P_{ij})\,s_is_j\\
&= \frac{1}{4m}s^T\mathbf{B}s
\end{aligned}
$$

$$\mathbf{s} = \sum a_i \mathbf{u_i}$$

$$Q = \frac{1}{4m}\sum a_i^2\lambda_i$$

This would be easily maximized with a vector **s** proportional to the first eigenvector…But we're constrained.

It can easily be shown that all rows and columns of **B** add up to zero. So the vector **1** will be an eigenvector.

Modularity Matrix and Spectral Methods

$$\mathbf{s_i} = \begin{cases} +1 & \text{if } \ u_i^{(1)} \geq 0 \\ -1 & \text{if } \ u_i^{(1)} < 0 \end{cases}$$

This approximation is the "**most parallel**" to the first eigenvector of **B**.

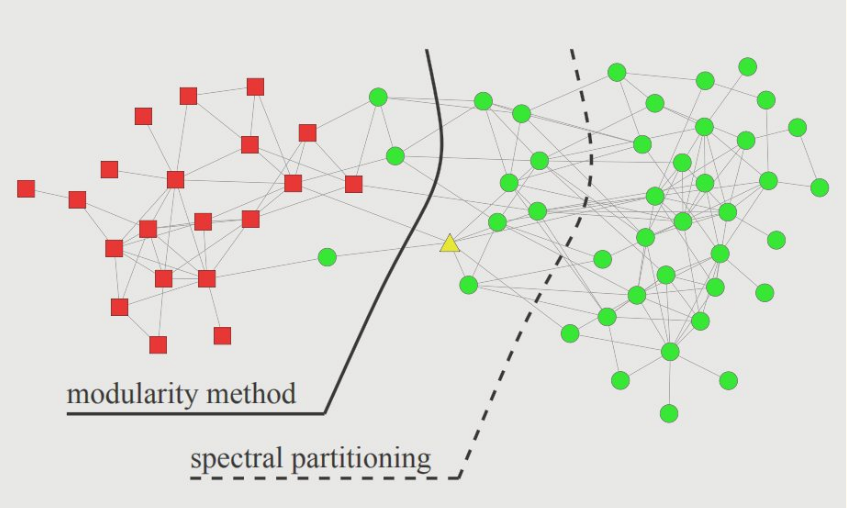


FIG. 2 The dolphin social network of Lusseau *et al.* [68]. The dashed curve represents the division into two equally sized parts found by a standard spectral partitioning calculation (Section II). The solid curve represents the division found by the modularity-based method of this section. And the squares and circles represent the actual division of the network observed when the dolphin community split into two as a result of the departure of a keystone individual. (The individual who departed is represented by the triangle.)

[1] "Finding community structure in networks using the eigenvectors of matrices", M.E.J. Newman

Random graphs with communities

Establishing a “ground truth”

SBM

Stochastic Block Model

- We construct k Erdós-Renyi graphs with parameters $\frac{c_{in}}{n}$
- We connect them with independent coin flips of probability $\frac{c_{out}}{n}$
- Assuming this is the model, we can find the **MAP** values of the parameters (statistical inference).
- Spectral methods are cheaper and make sense even when the true model isn't an **SBM**.

Statistical Inference for SBM

- Allows for a solid theoretical framework (For example proving **phase transition**)
- Provably optimal (attains detection and recovery to the theoretical limit)

Basic Idea

Estimate **MAP** parameters, like fitting a line through points. (Good heuristics)

Problem

If the true model isn't **SBM** then it's like fitting a line to a quadratic function, it can never be a good fit. SBM has **no triangles** and **Poisson degree distribution (unrealistic)**.

Solution?

Degree corrected block model or Configuration+Transitivity model >> Hard problem

After relaxation, the MAP problem is equivalent to **modularity maximization** which are solved with **spectral methods**

Phase transition and spectral method gap

- Using the statistical inference framework it was proven that detectability of communities isn't possible when:

$$c_{in} - c_{out} < 2\sqrt{\frac{c_{in} + c_{out}}{2}}$$

- Below that threshold the SBM is indistinguishable from an Erdós-Renyi graph with parameter the average degree.
- This was conjectured to happen due to the second eigenvalue being lost in the **bulk**.

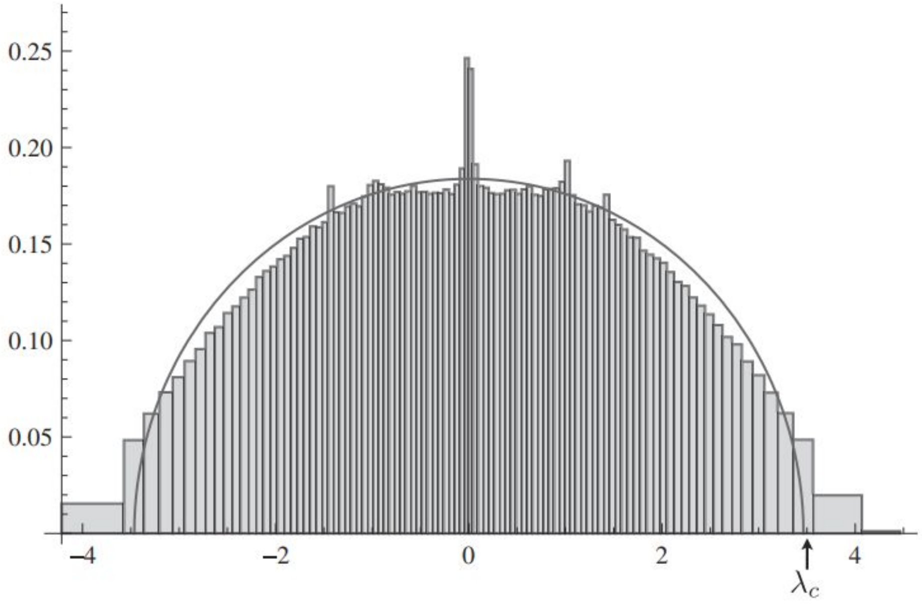
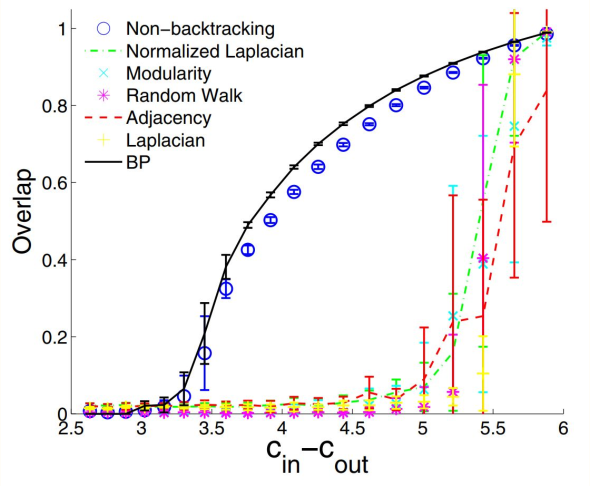


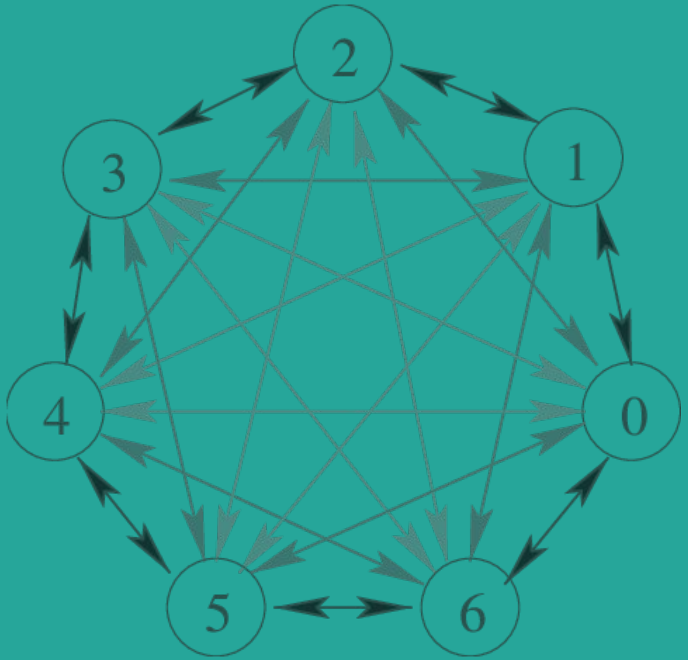
Fig. 1. Spectrum of the adjacency matrix of a sparse network generated by the block model (excluding the zero eigenvalues). Here, $n = 4{,}000$, $c_{\text{in}} = 5$, and $c_{\text{out}} = 1$, and we average over 20 realizations. Even though the eigenvalue $\lambda_c = 3.5$ given by Eq. **2** satisfies the threshold condition **1** and lies outside the semicircle of radius $2\sqrt{c} = 3.46$, deviations from the semicircle law cause it to get lost in the bulk, and the eigenvector of the second largest eigenvalue is uncorrelated with the community structure. As a result, spectral algorithms based on A are unable to identify the communities in this case.

Spectral Method Gap

- For a decade or so, spectral methods didn't get to the theoretical limit when statistical inference methods (like Belief Propagation) did. (**SPARSE CASE**)
- Spectral algorithms, based on the adjacency matrix, random walk matrix, and graph Laplacian, were assumed to be inferior.
- (Krzakala, Mossel, 2013) showed a "**Spectral Redemption**"



Flow Dynamics approach



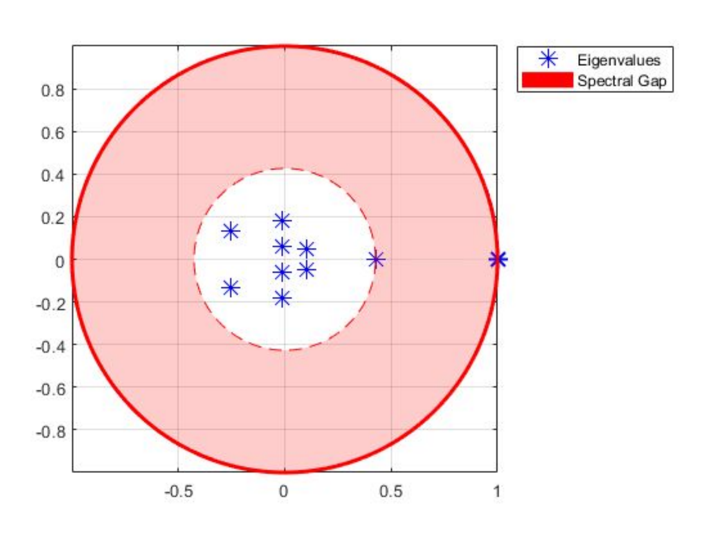
Random Walk

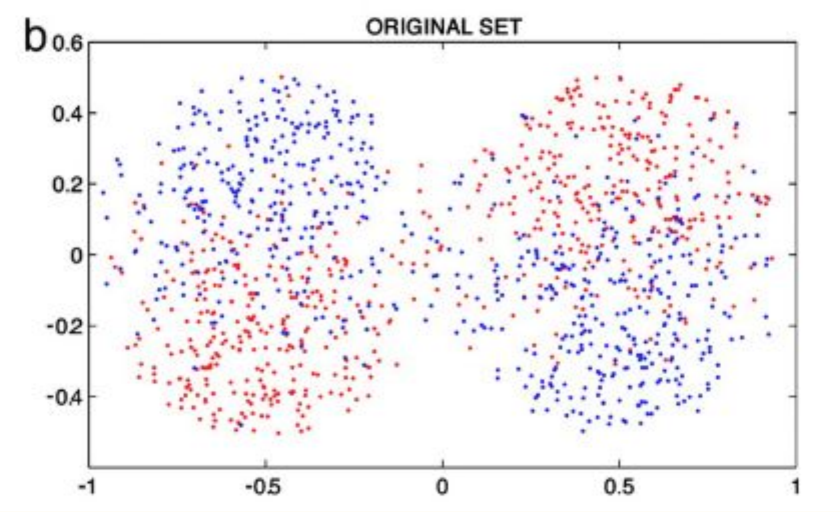
- We can associate an adjacency matrix to a **simple random walk** on the nodes.
- This random walk has a stationary distribution

$$\pi_i = \frac{d_i}{2m}$$

- Any probability measure on the nodes will **converge to it**. Such initial probability can be seen as an **excess of energy** in each node (when subtracting the stationary distribution).
- Which "**shock**" to the equilibrium takes longer to "**mix**"?

Conservative Flow ~ Community structure

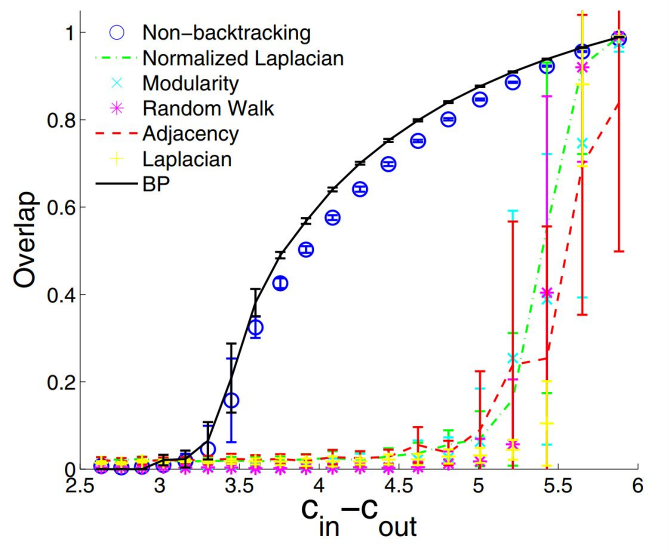




- To get the **slowest mode** of convergence, we want to put hot particles and cold particles "**away**" from each other. Make use of **Bottlenecks**.
- It makes sense that the **second eigenvector** of the transition matrix, gives the perturbation that vanishes at the **slowest rate**.
- In a **SBM** model this is precisely what communities are.

Conservative Flow approach

- An intuitive approach is to row-normalize the **Adjacency matrix** and look at the second eigenvector for the community structure.
- Other options include the normalized **Laplacian Matrix** and the **symmetric normalized Laplacian Matrix**. They all fail:



Spectral Redemption

- Krzakala et al. (2013) argue that the spectrum of these sparse matrices is very influenced by the high degree nodes. Other eigenvalues "swallow" the important one

Theorem 1.1. *Let $G = G(n, p)$ be a random graph and let Δ be the maximum degree of G. Then almost surely the largest eigenvalue of the adjacency matrix of G satisfies*

$$\lambda_1(G) = (1 + o(1)) \max\{\sqrt{\Delta}, np\},$$

where the $o(1)$ term tends to zero as $\max\{\sqrt{\Delta}, np\}$ tends to infinity.

- From a flow perspective; high degree nodes and their neighborhoods conserve the flow for a long time, but <u>they might not be totally included in one community.</u>
- The second eigenvector gets **focalized** around **high degree nodes**.

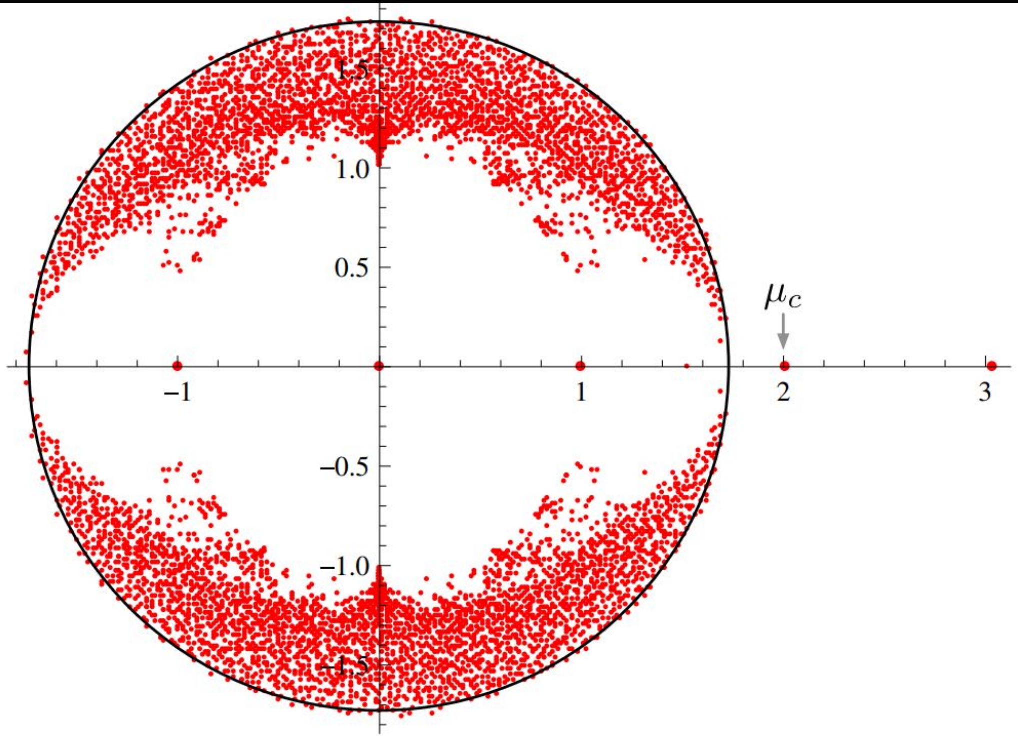


Fig. 2. Spectrum of the nonbacktracking matrix B for a network generated by the block model with same parameters as in Fig. 1. The leading eigenvalue is at $c = 3$, the second eigenvalue is close to $\mu_c = (c_{\text{in}} - c_{\text{out}})/2 = 2$, and the bulk of the spectrum is confined to the disk of radius $\sqrt{c} = \sqrt{3}$. Because μ_c is outside the bulk, a spectral algorithm that labels vertices according to the sign of B's second eigenvector (summed over the incoming edges at each vertex) labels the majority of vertices correctly.

Non-Backtracking Operator

Problem	Spectral Methods fail, because high degree “hubs” trap the flow for a long time, and could be **uncorrelated to communities.**
Idea	Force the random walk over the nodes to not stay in the high degree hubs. Backtracking is not allowed!
Operator	This non-markovian dynamics require an operator that encodes the past. The **non-backtracking operator** on directed **edges**.

The Non-Backtracking Matrix

- We represent this walk with a walk on the edges.
- First, we duplicate each **undirected edge** with two **opposing directed edges**.
- Then we create a matrix **B** on those 2m directed edges, with values 1 and 0 given by:

$$B_{i\to j,k\to l}=\delta_{jk}(1-\delta_{kl})$$

- This matrix' spectrum isn't influenced that much by high degree nodes. I.e. the bulk doesn't swallow the second eigenvalue that easily.
- From a **flow** perspective, we push the excess energy away from high degree hubs forcing it to travel the network.

Clustering using the Non-Backtracking Operator

1. Compute the Non-Backtracking matrix.
2. Get the second eigenvector (of dimension 2m)
3. Average the entries corresponding to edges that point to node i. That's the score for node i.
4. Classify the nodes according to the sign of their score.

Problem

Tree-like structures can only be "walked" with backtracking. The effect is that their nodes have 0 value in the eigenvector and are incorrectly classified. This is an issue with **very sparse** graphs.

Detection in networks with Triangles

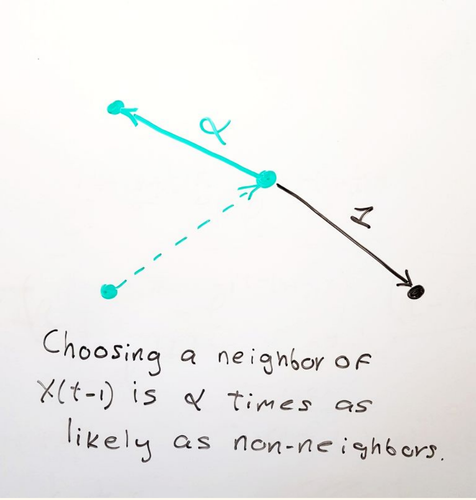
Transitivity

If $A_{ij} = 1$ and $A_{jk}=1$, then $A_{ik}=1$ with high probability

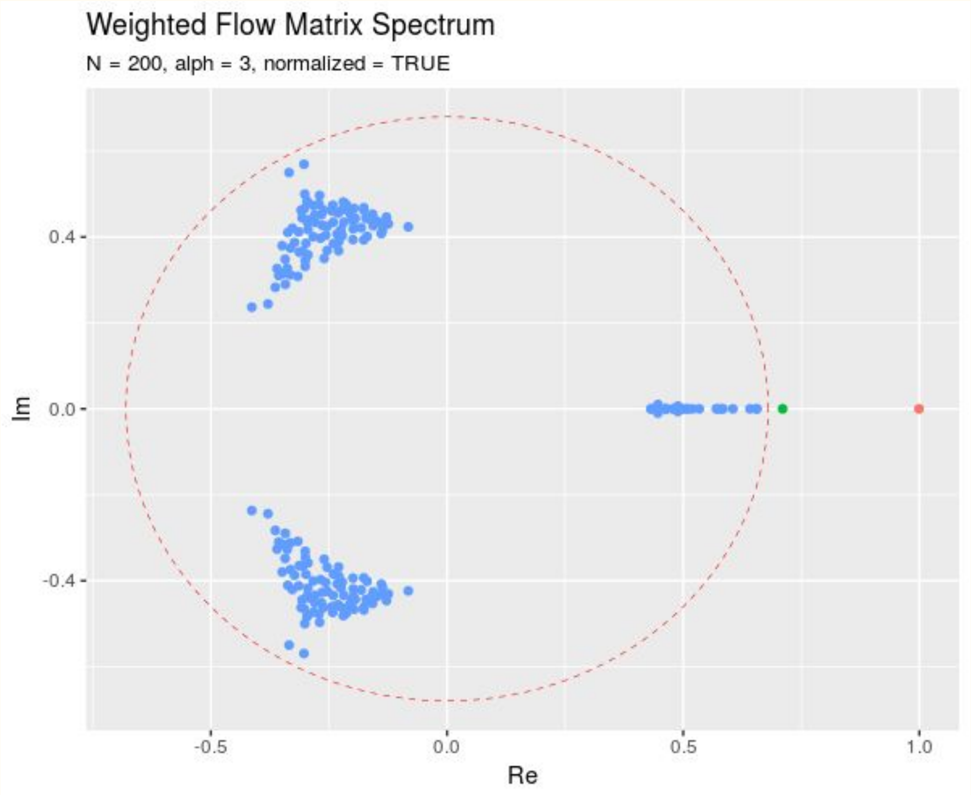
- What explains it?
- How can we generate random networks with it?
- SBM doesn't have it.
- Non-backtracking operator doesn't use it.
- Can we?

Main idea: Modify NBT operator

- If we know that within communities there are **more triangles** than across, then we can use that information to try and **stay within communities** as long as possible.
- If I'm biased towards the nodes that form a triangle with my current node and the previous one, then with high likelihood I'll stay in the community of those two nodes.



Spectrum of the BTT algorithm



Future Work

- Node dependent Alpha
- Analysis of the third spectral moment.
- Connection to the higher-order cut problem.
- Transitivity detectability threshold.

Thank you