Graph partitioning using matrix differential equations

Nicola Guglielmi

Gran Sasso Science Institute, Italia

Madrid - Universidad Carlos III, April 6, 2018

Mainly inspired by joint researches with Christian Lubich (Univ. Tübingen).

Also based on PhD projects of Eleonora Andreotti (Univ. L'Aquila) and Dominik Edelmann (Univ. Tübingen).

Outline of the talk

1. **The problems**
2. **Two step methodology**
3. **Inner step.**
   - Deriving monotone ODEs
   - Qualitative properties of the gradient system of ODEs
4. **Outer step.**
   - Quadratically convergent iterations
5. **Additional constraints: membership and cardinality**
6. **Illustrative examples and computational considerations**
   - Computational considerations

The problems

Given a connected undirected weighted graph, we are concerned with problems related to partitioning the graph. First of all we look for the closest disconnected graph (the minimum cut problem).

We are interested in the case of constrained minimum cut problems, where constraints include cardinality or membership requirements, which leads to NP-hard combinatorial optimization problems.

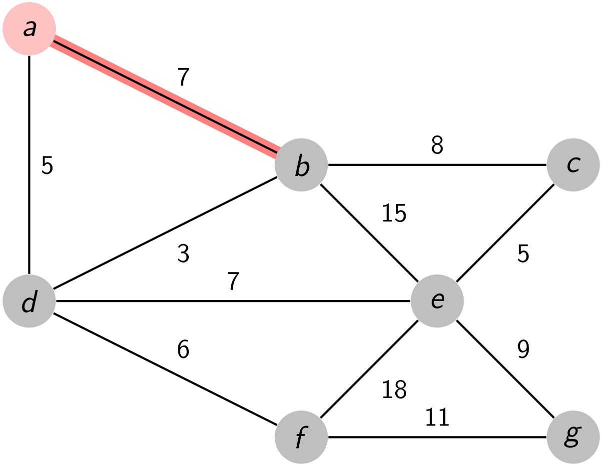
Also, we are interested in ambiguity issues, i.e. in the robustness of clustering algorithms that are based on spectral partitioning.

As opposed to combinatorial algorithms, the algorithm presented here modifies all weights of the graph as it proceeds, and only in the end arrives at the cut and the unchanged remaining weights.

The above-mentioned problems are restated as matrix nearness problems for the weight matrix of the graph.

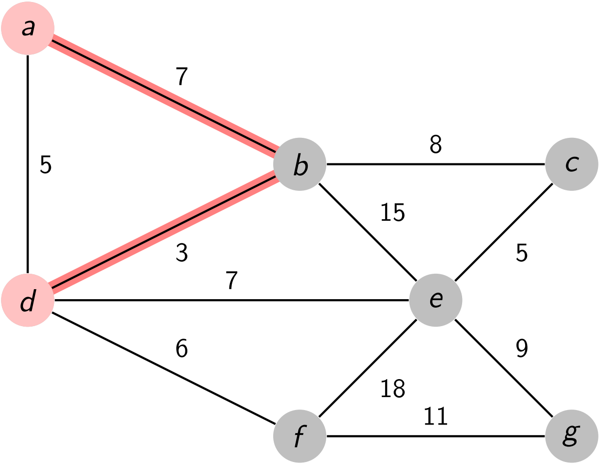
Edge-weighted graph connectivity

Minimum cut with bounded size (cardinality constraint), i.e. find the minimum cut which partitions the graph into two subsets S and T with $|S|, |T| \geq \bar{n}$ ($\bar{n} = 2$ in the example). Problem is NP-hard.



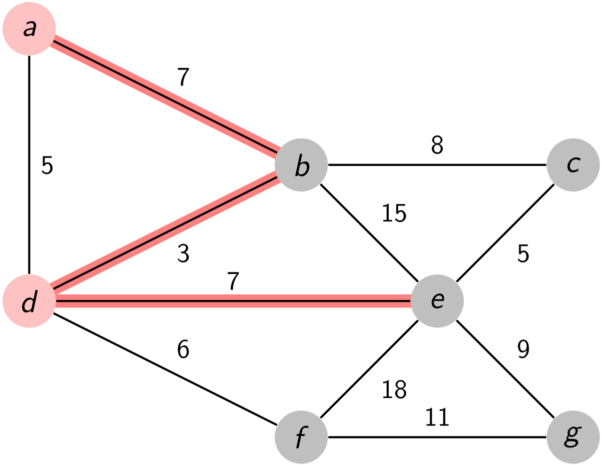
Edge-weighted graph connectivity

Minimum cut with bounded size (cardinality constraint), i.e. find the minimum cut which partitions the graph into two subsets S and T with $|S|, |T| \geq \bar{n}$ ($\bar{n} = 2$ in the example). Problem is NP-hard.



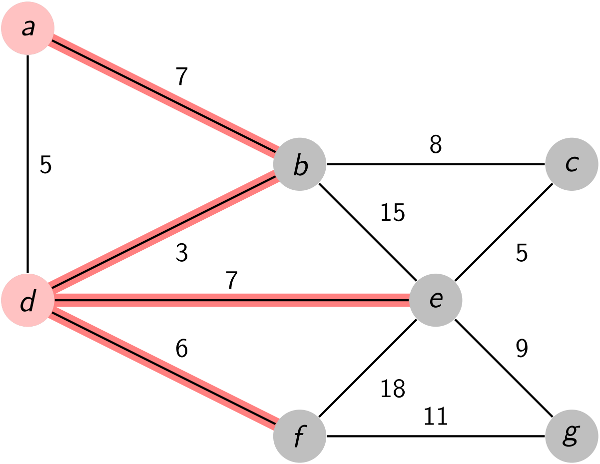
Edge-weighted graph connectivity

Minimum cut with bounded size (cardinality constraint), i.e. find the minimum cut which partitions the graph into two subsets S and T with $|S|, |T| \geq \bar{n}$ ($\bar{n} = 2$ in the example). Problem is NP-hard.



Edge-weighted graph connectivity

Minimum cut with bounded size (cardinality constraint), i.e. find the minimum cut which partitions the graph into two subsets S and T with $|S|, |T| \geq \bar{n}$ ($\bar{n} = 2$ in the example). Problem is NP-hard.



The minimum cut problem

Consider a graph with vertex set $\mathcal{V} = \{1, \ldots, n\}$ and edge set $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$. We assume that the graph is undirected: if $(i, j) \in \mathcal{E}$, also $(j, i) \in \mathcal{E}$. We associate *weights* w_{ij} for $(i, j) \in \mathcal{E}$, such that

$$w_{ij} = w_{ji} \geq 0 \quad \text{for all} \ \ (i, j) \in \mathcal{E}.$$

The graph is connected if for all $i, j \in \mathcal{V}$, there is a path of arbitrary length ℓ, $(i_0, i_1), (i_1, i_2), \ldots, (i_{\ell-1}, i_\ell) \in \mathcal{E}$, such that $i = i_0$ and $j = i_\ell$ and $w_{i_{k-1}, i_k} > 0$ for all $k = 1, \ldots, \ell$.

Minimum cut problem: Given a connected weighted undirected graph with weights w_{ij}, we aim to find a disconnected weighted undirected graph with the same edge set $\mathcal{E}$ and modified weights $\widehat{w}_{ij}$ such that

$$\frac{1}{2} \sum_{(i,j) \in \mathcal{E}} (\widehat{w}_{ij} - w_{ij})^2 \quad \text{is minimized}$$

plus some possible constraints (as a bounded size).

Graph Laplacian and graph connectivity

Setting $w_{ij} = 0$ for $(i,j) \notin \mathcal{E}$, we have the symmetric weight matrix

$$W = (w_{ij}) \in \mathbb{R}^{n \times n}.$$

The degrees $d_i = \sum_{j=1}^{n} w_{ij}$ are collected in the diagonal matrix

$$D = \mathrm{diag}(d_i) = \mathrm{diag}(W\mathbf{1}), \qquad \text{where } \mathbf{1} := (1, \ldots, 1)^T \in \mathbb{R}^n.$$

The *Laplacian matrix* $L = \mathrm{Lap}(W)$ is defined by

$$L = D - W$$

All eigenvalues of L are nonnegative, and $L\mathbf{1} = 0$, so that $\lambda_1 = 0$ is the smallest eigenvalue of L. Remarkably, the connectivity of the graph is characterized by the second-smallest eigenvalue of L.

Graph Laplacian and graph connectivity

Theorem (M. Fiedler, 1973)

Let $W \in \mathbb{R}^{n \times n}$ be the weight matrix of an undirected graph and L the associated Laplacian matrix with eigenvalues

$$0 = \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n.$$

Then, the graph is disconnected if and only if $\lambda_2 = 0$. Moreover, if $0 = \lambda_2 < \lambda_3$, then the entries of the corresponding eigenvector orthogonal to $\mathbf{1}$ assume only two different values, of different sign, which mark the membership to the two connected components.

Because of this result, the second smallest eigenvalue λ_2 of L is called *algebraic connectivity* of W.

If λ_2 is a simple eigenvalue, then the corresponding eigenvector is known as the *Fiedler vector* and is used for clustering purposes.

Considered problems

Constrained minimum cut:

- *Membership constraint minimum cut:* It is required that a given set of vertices $\mathcal{V}^+ \subset \mathcal{V}$ are in one connected component and another given set of vertices $\mathcal{V}^- \subset \mathcal{V}$ is in the other connected component.
- *Cardinality constraint minimum cut:* It is required that each of the connected components has a prescribed minimum number $\overline{n}$ of vertices.

Clustering robustness:

- *Constrained/unconstrained clustering:* if a small perturbation in the weights is able to determine coalescence of λ_2 and λ_3 the clustering based on λ_2 is not robust and an ambiguity occurs.

Outline of the talk

The underlying ideas

The approach of this talk takes basic ideas and techniques of recent algorithms for eigenvalue optimization via differential equations.

A common feature is a two-level procedure, where on the inner level a gradient flow drives perturbations to the original matrix of a fixed size into a (local) minimum of a functional that depends on eigenvalues and possibly eigenvectors, and in an outer iteration the perturbation size is determined such that the functional becomes zero.

Similarly to previous ones, the algorithms presented here cannot guarantee to find the global minimum of a non-smooth, non-convex optimization problem, or of an NP-hard combinatorial optimization problem.
Even with this caveat, the presented algorithm performs remarkably well in the examples from the literature on which we have tested it.

Two-level procedure stated for minimum cut

Minimum cut restated as matrix nearness problem: find $\Delta = \varepsilon E$ (with E of unit norm) such that $\mathrm{Lap}(W + \Delta)$ has eigenvalue $\lambda_2 = 0$.

Define the functional

$$F_\varepsilon(E) = \lambda_2\big(\mathrm{Lap}(W + \varepsilon E)\big)$$

(i) For given $\varepsilon > 0$, find $E = E(\varepsilon)$ of unit norm that minimizes $F_\varepsilon(E)$ under the constraints $W + \varepsilon E \geq 0$ and symmetry and sparsity pattern of E.

(ii) Find the smallest ε such that $F_\varepsilon(E) = 0$.

ODE approach: in order to compute $E(\varepsilon)$ for a given $\varepsilon > 0$, we use a constrained gradient system for the functional $F_\varepsilon(E)$

Outline of the talk

Feasible set

For a set of edges $\mathcal{E}$, we define $P_{\mathcal{E}}$ as the orthogonal projection from $\mathbb{R}^{n\times n}$ onto the sparsity pattern determined by $\mathcal{E}$: for $E = (e_{ij})$,

$$P_{\mathcal{E}}(E)\big|_{ij} := \begin{cases} e_{ij} & \text{if } (i,j) \in \mathcal{E}\,, \\ 0 & \text{otherwise.} \end{cases}$$

For a fixed given weight matrix W and for $\varepsilon > 0$, we call a matrix $E = (e_{ij}) \in \mathbb{R}^{n\times n}$ ε-feasible if the following conditions are satisfied:

(i) E is of unit Frobenius norm

(ii) E is symmetric

(iii) $E = P_{\mathcal{E}}(E)$

(iv) $W + \varepsilon E \geq 0$

The gradient of the functional

Lemma

Let $t \mapsto E(t)$ be a regular path of ε-feasible matrices, and $\lambda_2(t)$ the simple second smallest eigenvalue of $L(t) = \mathrm{Lap}(W + \varepsilon E(t))$ with associated eigenvector $x(t)$ and $\|x(t)\| = 1$. Then

$$\dot{\lambda}_2 = x^T \dot{L} x = \langle xx^T, \dot{L} \rangle = \varepsilon \langle G_\varepsilon(E), \dot{E} \rangle, \quad \textit{where}$$

$$G_\varepsilon(E) = P_{\mathcal{E}}(\mathrm{Sym}(x^2 \mathbf{1}^T) - xx^T)$$

Interpretation: since $F_\varepsilon(E) = \lambda_2(E)$, $G_\varepsilon(E)$ is the gradient of $F_\varepsilon(E)$

Notation: $\langle X, Y \rangle = \mathrm{trace}(X^T Y)$ denotes Frobenius inner product

$\mathrm{Sym}(A) = \frac{1}{2}(A + A^T)$ denotes the symmetric part of a matrix A

x^2 is intended componentwise; $\mathbf{1}$ denotes the vector with all entries 1

Admissible directions

Since $E(t)$ is of unit Frobenius norm by condition (i), we have

$$0 = \frac{1}{2}\,\frac{d}{dt}\|E(t)\|^2 = \langle E(t),\, \dot{E}(t)\rangle.$$

Condition (iv) ($W + \varepsilon E \geq 0$) requires that $\dot{E}_{ij} \geq 0$ for all $(i,j) \in \mathcal{E}_0$, where is the set of zero-weight edges defined by

$$\mathcal{E}_0 := \{(i,j) \in \mathcal{E} :\ w_{ij} + \varepsilon e_{ij} = 0\}.$$

Hence, for every ε-feasible matrix E, a matrix $Z = (z_{ij}) \in \mathbb{R}^{n\times n}$ is the derivative at $t = 0$ of some path of ε-feasible matrices starting at E if and only if the following four conditions are satisfied:

(i') $\langle E, Z\rangle = 0$

(ii') Z is symmetric

(iii') $Z = P_{\mathcal{E}}(Z)$

(iv') $P_{\mathcal{E}_0}(Z) \geq 0$

Steepest descent direction

Goal: determine the steepest descent ε-feasible direction Z for $F_\varepsilon(E)$.

The optimization problem (with $G = G_\varepsilon(E)$)

$$\min_Z \langle G, Z \rangle \quad \text{subject to (i')–(iv') and } \langle Z, Z \rangle = 1.$$

The additional constraint $\|Z\| = 1$ just normalizes the direction.

The solution satisfies the KKT conditions and is given by

$$Z = -G - \kappa E + \sum_{(i,j)\in\mathcal{E}_0} \mu_{ij} \mathrm{e}_i \mathrm{e}_j^T$$

$$\mu_{ij} z_{ij} = 0 \text{ for all } (i,j) \in \mathcal{E}_0$$

$$\mu_{ij} \geq 0 \text{ for all } (i,j) \in \mathcal{E}_0$$

Constrained gradient system

The constrained gradient flow of F_ε is the system of ODEs

$$\dot{E}(t) = Z(t),$$

where $Z(t)$ solves the KKT system, with set of edges $\mathcal{E}_0(t)$.

Lemma

On an interval where $\mathcal{E}_0(t)$ does not change, the gradient system becomes, with $P^+ = P_{\mathcal{E}\setminus\mathcal{E}_0}$

$$\dot{E} = -P^+ G_\varepsilon(E) - \kappa P^+ E \quad \textit{with} \quad \kappa = \frac{\langle -G_\varepsilon(E), P^+ E\rangle}{\|P^+ E\|^2}.$$

In a numerical solution of the system, we have to monitor the sets of edges where $w_{ij} + \varepsilon e_{ij} = 0$ and among them those edges where the sign of $-g_{ij} - \kappa e_{ij}$ changes. When the active set is changed, then also κ changes in a discontinuous way. It can happen that only a generalized solution in the Filippov sense exists.

Outline of the talk

Monotonicity and stationary points

Next result follows from the construction of the gradient system.

Theorem

The flow of the ODE has the following properties:

1. *Norm conservation:* $\|E(t)\|_F = 1$ *for all* t*;*
2. *Monotonicity:* $\lambda_2(t)$ *decreasing along solutions of ODE;*
3. *Stationary points: the following statements are equivalent:*
   $\dot{\lambda}_2 = 0 \iff \dot{E} = 0 \iff P^+E$ *is real multiple of* $P^+G_\varepsilon(E)$

Outline of the talk

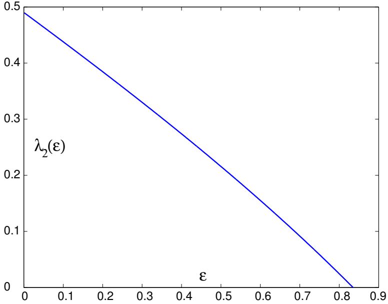
Summary of inner step

For fixed ε we compute optimal $E(\varepsilon)$ and $\lambda_2(\varepsilon)$, eigenvalue of

$$\mathrm{Lap}\,(W + \varepsilon E(\varepsilon)).$$

In order to determine the minimum cut we have to solve equation

$$\lambda_2(\varepsilon) = 0 \qquad \text{with respect to } \varepsilon.$$



Variational formula and outer step

Theorem

Under natural smoothness assumptions the function $f(\varepsilon) = F_\varepsilon(E(\varepsilon))$ is differentiable and its derivative equals (with $\;' = d/d\varepsilon$)

$$f'(\varepsilon) = -\|P^+ G_\varepsilon(E(\varepsilon))\| \, \|P^+ E(\varepsilon)\| - \frac{1}{\varepsilon^2} \frac{\|P^+ G_\varepsilon(E(\varepsilon))\|}{\|P^+ E(\varepsilon)\|} \, \|P_{\mathcal{E}_0} W\|^2.$$

This is the crucial formula for the the outer step of the methodology.

It allows to apply a Newton-bisection technique providing quadratic convergence from the left.

Membership/cardinality constrained minimum cut problem

Notation:

Let $x = (x_i) \in \mathbb{R}^n$ be the eigenvector to λ_2 of $\mathrm{Lap}(W + \varepsilon E)$.

Let $x^- = (x_i^-)$ with $x_i^- = \min(x_i, 0)$ and $x^+ = (x_i^+)$ with $x_i^+ = \max(x_i, 0)$ collect the negative and positive components of x.

Let n^-, n^+ be the numbers of negative and nonnegative components of x, respectively and denote the averages of x^- and x^+ by

$$\langle x^- \rangle = \frac{1}{n^-} \sum_{i=1}^{n} x_i^-, \qquad \langle x^+ \rangle = \frac{1}{n^+} \sum_{i=1}^{n} x_i^+.$$

Membership constraint: Let $\mathcal{V}^-$ and $\mathcal{V}^+$ be the set of indices whose membership to different components of the cut graph is prescribed.
Cardinality constraint: Here $\mathcal{V}^-$ and $\mathcal{V}^+$ are not given *a priori*, but are chosen depending on E, collecting the indices of the smallest and largest $\overline{n}$ components of the eigenvector x, respectively.

The constrained functional

Motivated by the special form of the eigenvector we consider

$$F_\varepsilon(E) = \lambda_2(\mathrm{Lap}(W + \varepsilon E)) + \frac{\alpha}{2} \sum_{i \in \mathcal{V}^-} (x_i - \langle x^- \rangle)^2 + \frac{\alpha}{2} \sum_{i \in \mathcal{V}^+} (x_i - \langle x^+ \rangle)^2,$$

where $\alpha > 0$ is a weight to be chosen.

Remark: computation of the gradient of F_ε requires differentiation of eigenvectors, implying computation of the solution of a linear system involving the pseudo-inverse of $L - \lambda_2 \mathrm{I}$.

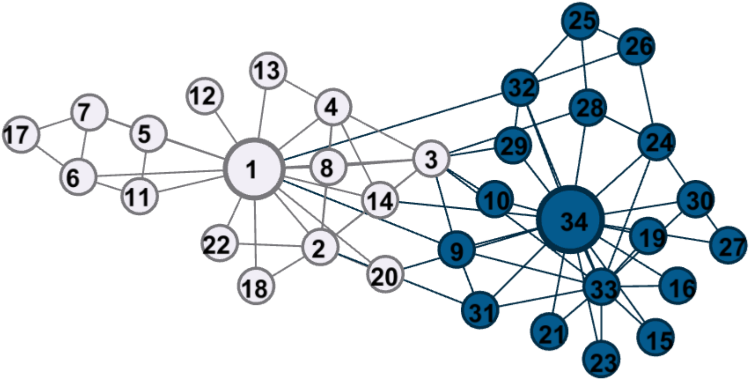
The choice of the sign of the eigenvector x is such that $F_\varepsilon(E)$ takes the smaller of the two possible values.

Outline of the talk

1. **The problems**
2. **Two step methodology**
3. **Inner step.**
   - Deriving monotone ODEs
   - Qualitative properties of the gradient system of ODEs
4. **Outer step.**
   - Quadratically convergent iterations
5. **Additional constraints: membership and cardinality**
6. **Illustrative examples and computational considerations**
   - Computational considerations

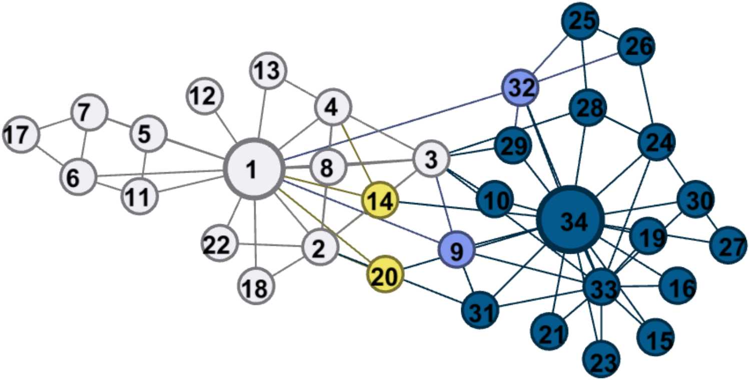
Illustrative example 1: unconstrained minimum cut

Zachary's karate club graph.



Illustrative example 1: membership constraint

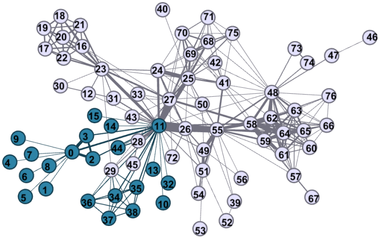
Zachary's karate club with membership constraint.



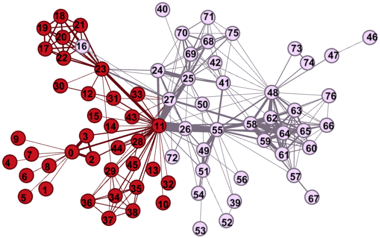
Illustrative example 2: cardinality constraint

In Figure is illustrated the graph character co-occurence in Les Miserables (Knuth, 2009), an undirected weighted network consisting of 77 vertices (representing characters), 22 of these belong to a part of Fiedler partition, and the remaining 55 belong to the other part.

Setting a threshold $\bar{n} = 35$ we obtain the result in the right picture.



(a) Les Miserables colored by Fiedler eigenvector.



(b) Les Miserables constrained graph colored by Fiedler eigenvector.

Computational considerations

The proposed algorithm is an iterative algorithm, where in each step the second eigenvalue and the associated eigenvector of the Laplacian of a graph with perturbed weights are computed.

In the cardinality- or membership-constrained cases, additionally a linear system with an extended shifted Laplacian is solved in each step.

For a large sparse connected graph (where the number of edges leaving any vertex is moderately bounded), these computations can be done in a complexity that is linear in the number of vertices. In the known (unconstrained) minimum cut algorithms, the computational complexity is at least quadratic (see Stoer and Wagner, 1997).

Computational considerations

It is thus conceivable that for large sparse connected graphs, the proposed iterative algorithm can favorably compete with the classical unconstrained minimum cut algorithms.

In constrained cases, it appears that the computational complexity is even more favorable in comparison with the existing heuristic combinatorial algorithms (see Bruglieri, Maffioli and Ehrgott, 2004).

However, as of now no detailed comparisons of the relative merits of the conceptually and algorithmically fundamentally different approaches have been made.