Rethinking Ratio Cut

Quick note on notation: For a matrix M, $\lambda_j(M)$ is the j^{th} eigenvalue of M in ascending order. $v_j(M)$ is the corresponding eigenvector. m_{ij} is the $(ij)^{th}$ entry of M and m_j is the j^{th} column vector of M.

The Ratio Cut Problem

Let $G = (V, E)$ be a connected, undirected similarity graph with $|V| = n$ vertices. Let $L \in \mathbb{R}^{n\times n}$ be the combinatorial graph Laplacian of G. For a given $k > 1$, let $H \in \mathbb{R}^{n\times k}$ be a cluster membership matrix such that $h_{ij} = \begin{cases} n_j^{-1/2} & v_i \in C_j \\ 0 & \text{else} \end{cases}$ where C_j is the j^{th} cluster and n_j is the number of vertices in C_j. Then the optimal ratio cut partition $\{C_1, ..., C_k\}$ is given by:

$$\arg\min_H \mathrm{Tr}(H^\top LH)$$

under the constraint that H is of the cluster membership matrix specified above.

This discrete optimization problem is NP-hard, so a common approach is to relax the constraints on H. Notice that $H^\top H = I_k$. Replace the cluster membership matrix constraint on H and just use the constraint $H^\top H = I_k$. Then the solution is just the first k eigenvectors of L, $\begin{bmatrix} v_1(L) & \cdots & v_k(L) \end{bmatrix}$. Since this doesn't provide cluster memberships, it is treated as an embedding and k-means is used as a rounding step to obtain cluster memberships.

Justification for the Relaxation of the Ratio Cut Problem

Suppose we instead start with $G_{iso} = (V, E_{iso})$, a graph consisting of k subgraphs that are each connected but disconnected from each other. Let L_{iso} be its combinatorial graph Laplacian. Then $\lambda_1(L_{iso}) = \cdots = \lambda_k(L_{iso}) = 0$, and the corresponding eigenvectors form the cluster membership matrix H (although any orthonormal basis in the column space of H could be used as well). Then let $G_\epsilon = (V, E_\epsilon)$ be a connected graph constructed from G_{iso} such that the inter-cluster edges are characterized by some small ϵ (Ling and Strohmer uses $\epsilon = \max D_\delta$ where D_δ is the degree matrix of a graph constructed from just the inter-cluster edges), such that the optimal ratio cut clustering is still given by $H = \begin{bmatrix} v_1(L_{iso}) & \cdots & v_k(L_{iso}) \end{bmatrix}$. If ϵ is small, then $\begin{bmatrix} v_1(L_\epsilon) & \cdots & v_k(L_\epsilon) \end{bmatrix} \approx \begin{bmatrix} v_1(L_{iso}) & \cdots & v_k(L_{iso}) \end{bmatrix}$ (in some sense) by the Davis-Kahan $\sin\Theta$ theorem.

Alternative formulation for the case where $k = 2$

C. Ding[1] and U. von Luxburg[2] both describe another equivalent discrete objective function for the ratio cut problem when $k = 2$. Define $f \in \mathbb{R}^n$ such that $f_i = \begin{cases} \sqrt{\frac{n_2}{nn_1}} & v_i \in C_1 \\ -\sqrt{\frac{n_1}{nn_2}} & v_i \in C_2 \end{cases}$

Then the optimal 2-way ratio cut is given by:

$$\arg\min_f f^\top Lf$$

[1] http://ranger.uta.edu/~chqding/Spectral/
[2] https://arxiv.org/abs/0711.0189

subject to f of the form specified above.

This is again a discrete optimization problem that is NP-hard, so we first note that f has the following properties:

- $\sum f_i = 0$
- $||f|| = 1$
- $f^\top e = 0$

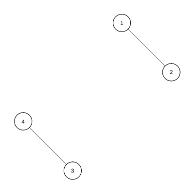
Then we remove the discreteness constraint on f and use these properties. The solution to this relaxed optimization problem is, similar to the arbitrary k case, is $v_2(L)$, the Fiedler vector. We can also note that as long as the graph that L describes is connected, $v_1(L) = e/\sqrt{n}$ no matter how small the inter-cluster edges are.

Proposition: Let $G_{iso} = (V, E_{iso})$ be a graph with two disconnected subgraphs, with corresponding weight matrix W_{iso}. Then let $\epsilon > 0$ and $G(\epsilon) = (V, E(\epsilon))$ be a connected graph constructed from G_{iso} such that $||D_\delta|| = \epsilon$ (where D and D_{iso} are the diagonal degree matrices of W and W_{iso} and $D_\delta = D - D_{iso}$, as described by Ling and Strohmer). Let $L(\epsilon)$ be the combinatorial graph Laplacian of $G(\epsilon)$.
Then as $\epsilon \to 0$, $v_2(L(\epsilon)) \to f$, where f is of the form described above.

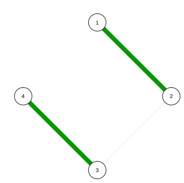
We typically think of the first two eigenvectors of L_{iso} as having the form $\begin{bmatrix} v_1(L_{iso}) & v_2(L_{iso}) \end{bmatrix} = H$, the solution to the ratio cut problem for L_ϵ, but since $\lambda_1(L_{iso}) = \lambda_2(L_{iso}) = 0$, we can really just take any two orthonormal vectors in the column space of H. However, we can still see that $\begin{bmatrix} v_1(L_\epsilon) & v_2(L_\epsilon) \end{bmatrix} \not\to H$, so perhaps it's more useful to think of f rather than H.

Example

Let G_{iso} be a graph with just four vertices and two edges of unit weight such that there is an edge between vertices 1 and 2 and an edge between vertices 3 and 4:



Let $\epsilon > 0$ and G_ϵ be a graph constructed from G_{iso} such that there is an edge between vertices 2 and 3 of weight ϵ (for the sake of making the optimal ratio cut solution be $\{\{1,2\},\{3,4\}\}$, we can let $\epsilon < 1$):



Let W_{iso} be the edge weight matrix of G_{iso} and L_{iso} be its combinatorial graph Laplacian. Then the first two eigenvalues of L_{iso} are $\lambda_1(L_{iso}) = \lambda_2(L_{iso}) = 0$ and the corresponding eigenvectors are $H = \begin{bmatrix} 1/\sqrt{2} & 0 \\ 1/\sqrt{2} & 0 \\ 0 & 1/\sqrt{2} \\ 0 & 1/\sqrt{2} \end{bmatrix}$.

On the other hand, let W_ϵ and L_ϵ be the edge weight matrix and combinatorial graph Laplacian of G_ϵ. Then the first eigenvalue of L_ϵ is $\lambda_1(L_\epsilon) = 0$ and the second eigenvalue is $\lambda_2(L_\epsilon) = 1 + \epsilon - \sqrt{1+\epsilon^2}$. The first eigenvector of L_ϵ is $v_1(L_\epsilon) = e/2$ while the second is $v_2(L_\epsilon) = \frac{1}{\sqrt{4+4\epsilon^2}} \begin{bmatrix} 1 \\ -\epsilon + \sqrt{1+\epsilon^2} \\ \epsilon - \sqrt{1+\epsilon^2} \\ -1 \end{bmatrix}$.

Taking the limit as $\epsilon \to 0$, we get:

- $v_1(L_\epsilon) = e/2$ (doesn't depend on ϵ)
- $v_2(L_\epsilon) \to \begin{bmatrix} 1/2 & 1/2 & -1/2 & -1/2 \end{bmatrix}^\top$

We can also see that $\lim\limits_{\epsilon \to 0} \lambda_2(L_\epsilon) = \lim\limits_{\epsilon \to 0} 1 + \epsilon - \sqrt{1+\epsilon^2} = 0 = \lambda_2(L_{iso})$.

Proof (sketch)

As $L_\epsilon \to L_{iso}$ ($\epsilon \to 0$), the subspace spanned by $v_1(L_\epsilon)$ and $v_2(L_\epsilon)$ must approach the subspace spanned by $v_1(L_{iso})$ and $v_2(L_{iso})$. One way to write an orthonormal basis for this subspace is $H \in \mathbb{R}^{n \times 2}$ such that $h_{ij} = \begin{cases} n_j^{-1/2} & v_i \in C_j \\ 0 & \text{else} \end{cases}$, where the C_j's correspond to the connected subgraphs of L_{iso} that are disconnected from each other and n_j corresponds to the number of vertices in each subgraph ($\sum_j n_j = n$). On the othre hand, we know that $\forall \epsilon > 0$, $v_1(L_\epsilon) = e/\sqrt{n}$, so $\begin{bmatrix} v_1(L_\epsilon) & v_2(L_\epsilon) \end{bmatrix} \not\to H$ as $\epsilon \to 0$ for $\epsilon > 0$. In order to find $\lim\limits_{\epsilon \to 0} \begin{bmatrix} v_1(L_\epsilon) & v_2(L_\epsilon) \end{bmatrix}$, we need to find an orthonormal basis for the subspace spanned by $v_1(L_{iso})$ and $v_2(L_{iso})$ that includes $v_1(L_\epsilon) = e/\sqrt{n}$. Since this subspace is two-dimensional, all we need to do are:

1. Verify that $e/\sqrt{n}$ is in this subspace.
2. Find the unique vector in this subspace that is orthogonal to $e/\sqrt{n}$.

For (1), we can see that $e/\sqrt{n} = \sqrt{\frac{n_1}{n}} v_1(L_{iso}) + \sqrt{\frac{n_2}{n}} h_2 = \sum_j^k \sqrt{\frac{n_j}{n}} h_j$, so $e/\sqrt{n}$ is in this subspace.

For (2), we already know that $||f|| = 1$ and $f \perp e/\sqrt{n}$ (recall $f_i = \begin{cases} \sqrt{\frac{n_2}{nn_1}} & v_i \in C_1 \\ -\sqrt{\frac{n_1}{nn_2}} & v_i \in C_2 \end{cases}$), so all that is left is to show that f is in this subspace. Sure enough, it can be shown that $f = \sqrt{\frac{n_2}{n}} h_1 - \sqrt{\frac{n_1}{n}} h_2$, so f must be in the subspace.

Additional Problems

Proximity between $v_2(L_\epsilon)$ and f

We noted that as $\epsilon \to 0$ (i.e., $L \to L_{iso}$ while keeping the graph generating L connected), v_2, the Fiedler vector of L, approaches f. Then it would be intuitive to believe that $||v_2(L_\epsilon) - f|| \le g(L_\epsilon)$ for some monotone increasing function g. We saw previously that performing 2-means clustering on v_2 (treating it as an embedding in $\mathbb{R}^1$) is equivalent to finding n_1 and n_2 such that $||v_2(L_\epsilon) - f||$ is minimized[3]. Furthermore, we saw that the 2-means clustering solution is not equivalent to the ratio cut solution. So minimizing $||v_2(L_\epsilon) - f||$ does not give us the ratio cut solution in the general case. However, if we can put an upper bound on $||v_2(L_\epsilon) - f||$, then perhaps we can define criteria for which the 2-means clustering solution on $v_2(L_\epsilon)$ is equivalent to the ratio cut solution (and hopefully this will be looser than the Ling-Strohmer criterion for RatioCut-SDP). In addition, if we can characterize $g(L_\epsilon)$ in some way, we can perhaps say how "correct" the spectral clustering approximation to ratio cut is depending on L_ϵ, since minimizing $||v_2(L_\epsilon) - f||$ is equivalent to the k-means objective for f.

Possible solution

By the Davis Kahan sine Θ theorem, we get:

$$||v_2(L_\epsilon) - f||^2 \le \frac{2||L_\epsilon - L_{iso}||^2}{(\lambda_2(L_\epsilon) - \lambda_2(L_{iso}))^2}$$

Since $\lambda_2(L_{iso}) = 0$, we get:

[3] https://github.com/johneverettkoo/summer-research-2018/blob/master/k2-example.pdf

$$||v_2(L_\epsilon) - f||^2 \leq 2\left(\frac{||L_\delta||}{\lambda_2(L_\epsilon)}\right)^2$$

As $\epsilon \to 0$, $||L_\delta|| \to 0$ and $\lambda_2(L_\epsilon) \to 0$, but numerical experiments show that the numerator should go to 0 faster.

Alternative SDP Problem

S. Ling and T. Strohmer[4] demonstrated that under certain conditions, a semidefinite program can solve ratio cut. Recall that ratio cut can be written as a discrete optimization problem $\arg\min_H \mathrm{Tr}(H^\top LH)$ where H is a cluster membership matrix as defined above. We can rewrite this as $\mathrm{Tr}(H^\top LH) = \mathrm{Tr}(LHH^\top) = \mathrm{Tr}(LZ)$ where $z_{ij} = \begin{cases} n_k^{-1} & x_i, x_j \text{ in same cluster } k \\ 0 & \text{else} \end{cases}$. This is still a discrete optimization problem that is NP-hard, but noting that Tr(Z) = k$ (the number of clusters), $Ze = e$, $Z \geq 0$ element-wise, and Z is positive semidefinite, Ling and Strohmer proposed another continuous relaxation for ratio cut:

$$\begin{aligned} \arg\min_Z & \mathrm{Tr}(LZ) \\ \text{s.t. } & \mathrm{Tr}(Z) = k \\ & Ze = e \\ & Z \geq 0 \text{ element-wise} \\ & Z \text{ is positive semidefinite} \end{aligned}$$

Based on this, we can try to form an alternative SDP problem using f instead of H (this is specific to $k = 2$)[5]:

Let $\Phi = ff^\top$. Then we can observe:

- $\Phi = \Phi^\top, \Phi \in \mathbb{R}^{n\times n}$
- $\Phi e = ff^\top e = f0 = \overrightarrow{0}$
- $\mathrm{Tr}(\Phi) = \mathrm{Tr}(ff^\top) = \mathrm{Tr}(f^\top f) = \mathrm{Tr}(1) = 1$ (this should be $k-1$ in the general case)
- Φ is positive semidefinite
  - Φ has rank 1 (or $k-1$ in the general case)
  - $\Phi f = ff^\top f = f1 = f$
- $\phi_{ij} = \begin{cases} \frac{n_2}{nn_1} & v_i, v_j \in C_1 \\ \frac{n_1}{nn_2} & v_i, v_j \in C_2 \\ -\frac{1}{n} & \text{else} \end{cases}$

Rewriting $\mathrm{Tr}(f^\top Lf) = \mathrm{Tr}(Lff^\top) = \mathrm{Tr}(L\Phi)$, we can state a relaxation of the ratio cut problem as:

$$\begin{aligned} \arg\min_\Phi & \mathrm{Tr}(L\Phi) \\ \text{s.t. } & \Phi \text{ is PSD} \\ & \Phi e = 0 \\ & \mathrm{Tr}(\Phi) = 1 \\ & \Phi \geq -\frac{1}{n} \text{ element-wise} \end{aligned}$$

... with the hope that this will let us relax the optimality criterion from Ling and Strohmer.

[4] https://arxiv.org/abs/1806.11429

[5] Note: I'm still working on generalizing $f \in \mathbb{R}^{n\times(k-1)}$ for the arbitrary k case. It's straightforward to describe this as a simplex in $\mathbb{R}^{k-1}$ with point masses on the vertices proportional to the cluster sizes such that the center of mass is the origin, but the actual closed form is taking some time to figure out.