ESANN 2010 proceedings, European Symposium on Artificial Neural Networks - Computational Intelligence and Machine Learning. Bruges (Belgium), 28-30 April 2010, d-side publi., ISBN 2-930307-10-2.

A randomized algorithm for spectral clustering

Nicola Rebagliati[1] and Alessandro Verri[1] *

1- Dipartimento di Informatica e Scienze dell'Informazione, Università di Genova
Via Dodecaneso 36, Genoa, 16146, Italy

Abstract. Spectral Clustering has reached a wide level of diffusion among unsupervised learning applications. Despite its practical success we believe that for a correct usage one has to face a difficult problem: given a target number of classes K the optimal K-dimensional subspace is not necessarily spanned by the first K eigenvectors of the graph Normalized Laplacian. The contribution of this paper is twofold. First, we show a bound for choosing a correct number of eigenvectors. Second, we propose a randomized spectral algorithm able to find a clustering solution. We show the efficacy of the algorithm with experiments on real world graphs. Our proposal is a scheme that naturally extends the current usage of Spectral Clustering.

1 Introduction

Spectral Clustering techniques constitute an effective toolbox for many unsupervised learning applications. With simple elements from linear algebra they give heuristic algorithms for combinatorial problems, usually NP-complete, arising in image segmentation, clustering and graph layout. In this work we focus on partitioning a dataset into disjoint sets so that the resulting partition minimizes the normalized cut. This functional aims at minimizing the sum of similarity weights across K clusters and at the same time to maximize the sum of similarity weights in each cluster, called volume. Minimizing the normalized cut is an NP-complete problem [1], but when the similarity matrix is, after a convenient rearrangement of the input dataset, block diagonal with K blocks we can easily find the desired K partitions by inspecting the first K eigenvectors of the Normalized Laplacian matrix L_N of the graph. This scheme is shared by most of Spectral Clustering algorithms. The general case is not so easy because we do not have any guarantee about the content of the first eigenvectors. On the contrary in [2] a graph construction is proposed such that for every K there is a graph where the first K eigenvectors of the Combinatorial Laplacian encode a cut completely different from the best one. The same construction can also be used for the Normalized Laplacian [3].

In this work we argue that despite counterexamples [2, 3] we can still rely on Spectral Clustering but introducing two modifications. Firstly we may need to consider more eigenvectors than the number of classes. We show a bound that may help in choosing this number of eigenvectors $M \geq K$. Secondly we propose a randomized algorithm to recover K classes using M eigenvectors of

*This work has been partially supported by the EU Integrated Project Health-e-Child IST-2004-027749.

ESANN 2010 proceedings, European Symposium on Artificial Neural Networks - Computational Intelligence and Machine Learning. Bruges (Belgium), 28-30 April 2010, d-side publi., ISBN 2-930307-10-2.

the Normalized Laplacian matrix. In the next section we give a brief overview of Spectral Clustering literature, in section 3 we present the normalized cut problem and a bound for choosing the number of eigenvectors of the Normalized Laplacian matrix, in section 4 we present a randomized algorithm and in section 5 there are experiments on real world graphs. We conclude with a discussion of the ideas presented.

2 Previous work and literature

The normalized cut problem was introduced in [1] together with a recursive spectral algorithm for image segmentation. After [1] many Spectral Clustering algorithms have been proposed in the literature, [4, 5, 6, 7], all of these consider the first K eigenvectors of the Normalized Laplacian matrix, perform some normalization on them and finally "round" to obtain a clustering. Spectral partitioning was introduced for cutting graphs with [8, 9], we refer to section 4 of [10] for a survey. Our work can be considered in between [11, 12, 13]. The idea of randomly searching in the span of the eigenvectors of the Combinatorial Laplacian for VLSI design was firstly proposed by [11], there they use one or two-dimensional random projections called "probes". Using many, if not all, eigenvectors has been advocated by [12]. A complementary point of view is that of discarding uninformative eigenvectors, as in [13].

3 The normalized cut

We consider symmetric weighted connected graphs. A symmetric weighted graph is a couple $G = (V, E)$, where V is a finite set of vertices and E is the set of edges between vertices, together with a weight function $w\colon V \times V \to [0, 1]$ such that $w(v_i, v_j) = w(v_j, v_i)$. $|V|$ is the number of vertices. We denote $w(v_i, v_j)$ with $w_{i,j}$. We define the matrix $W_G \in R^{n\times n}$ such that $(W_G)_{i,j} = w_{i,j}$ and the diagonal matrix D_G with the degree of each vertex, i.e. $d_i = \sum_{j=1}^{|V|} w_{i,j}$, on the diagonal, $(D_G)_{i,i} = d_i$ and 0 otherwise.

A cut of a graph is a partition of the vertices into disjoint, non-empty sets $(A_1, \ldots, A_K)$ such that $\cup_{k=1,\ldots,K} A_k = V$. The *cut* value of a subset A_k of V is $cut(W_G, A_k) = \sum_{i\in A_k, j\notin A_k} w_{i,j}$. The volume value of a subset A_k of V is $vol(W_G, A_k) = \sum_{i\in A_k} d_i$. The problem of minimizing the normalized cut is the following:

$$NCut^* = \min_{A_1,\ldots,A_K} NCut(W_G, A_1, \ldots, A_K) = \min_{A_1,\ldots,A_K} \sum_{k=1}^{K} \frac{cut(W_G, A_k)}{vol(W_G, A_k)}$$

This problem is NP-complete to solve [1], but we have an interesting heuristic argument at our disposal. Let $L_N = I - D_G^{-1/2} W_G D_G^{-1/2}$ be the Normalized Laplacian of graph G and $0 = \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_{|V|} \leq 2$ its eigenvalues in ascending order. Following [14] any given partition can be represented as a matrix

H_A such that $(H_A)_{i,k} = \sqrt{d_i/vol(W_G, A_k)}$ if $v_i \in A_k$ and 0 otherwise. It is straightforward to check that $NCut(W_G, H_A) = Tr(H_A^T L_N H_A)$ and this characterization of the *NCut* allows us to use Fan Inequality and obtain a lower bound on its minimal value, $NCut^* \geq \sum_{k=1}^{K} \lambda_k$. If the equality holds and $\lambda_{K+1} > \lambda_K$ the subspace spanned by the first K eigenvectors is the same of H_{A^*}. Even if we perturb the weights of the similarity matrix of the ideal case with a small quantity of noise, the subspace spanned by the first eigenvectors will substantially overlap the subspace spanned by the columns of H_{A^*} and we can find it using "rounding" techniques. These techniques are algorithms that take as input a noisy version of H_{A^*} and try to recover it.

There are examples [2] where the subspace spanned by the first K eigenvectors of L_N may not contain information about the best partition. Informally our observation is that if the general gap $\lambda_{M+1} - \lambda_K$ is large the information about the best cut can still be found in the first $M \geq K$ eigenvectors. This fact is formalized by the following theorem:

Theorem 1. Let $G = (V, E)$ be a graph and let L_N be its Normalized Laplacian. Let $L_N = U \Lambda U^T$ be the spectral decomposition of L_N, with eigenvalues $0 = \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_{|V|} \leq 2$ ordered in ascending order. Let M be an integer, $K \leq M \leq |V| - 1$. Let U_M be the first M eigenvectors of L_N. For any clustering H_A decomposed as $H_A = [U_M U_e] \left[R_A^T \; E_A^T\right]^T$, with $U = [U_M U_e]$, $R_A \in I\!R^{M \times M}$ and $E_A \in I\!R^{|V|-M \times M}$, the following upper bound on $\|E_A\|_F^2$ holds:

$$\|E_A\|_F^2 \leq \frac{NCut(W_G, A) - \sum_{k=1}^{K} \lambda_k}{\lambda_{M+1} - \lambda_K} \tag{1}$$

The proof can be found in [3]. This bound is a generalization of a bound in [15] and applies for any cut H_A, in particular for the best one H_{A^*}. Bound 1 tells us that when $\lambda_{M+1} - \lambda_K$ is large we can filter out eigenvectors with $\lambda \geq \lambda_{M+1}$ because we know that the frobenius norm of H_{A^*} is well approximated by the first M eigenvectors. In the next section we present a randomized algorithm that may use bound 1 for computing H_A^* .

4 A randomized spectral algorithm

Current Spectral Clustering algorithms [4, 5, 7] exploit the trace characterization of the normalized cut by considering the first K eigenvectors of the Normalized Laplacian matrix as a noisy representation of the K classes correct partition. However counterexamples in [2, 3] show that the first K eigenvectors of the Normalized Laplacian may not contain the information required for computing the best cut. Motivated by these examples we propose a randomized spectral algorithm that generalizes current spectral algorithms and considers $M \geq K$ eigenvectors. Bound 1 may help in choosing the value of M. We have to make two important observations in the case M is strictly greater than K. Firstly we do not know where H_{A^*} exactly is among the span of the first M eigenvectors, see examples in [3]. Secondly, we do not know the nature of the information contained in the remaining $M - K$-dimensional linear subspace. In the examples

given in [2] this information can be considered adversarial, as the related non optimal cuts are substantially different from the optimal ones. These observations suggest that rounding directly the first M eigenvectors into K classes may not work and a different scheme is needed. The algorithm **spectral cut** considers as inputs a symmetric weight matrix W_G of a graph $G = (V, E)$, a number of target clusters K, a number $M \geq K$ of eigenvectors of the Normalized Laplacian to be rounded, denoted as U_M. To recover H_{A^*} the randomized spectral algorithm samples random matrices $P \in I\!R^{M \times K}$, with $P^T P = I$, $(PP^T)_{1,1} = 1$ and apply a rounding scheme, e.g. K-means, to $U_M P$. The partition with the lowest normalized cut value among the `max_rep` samples is returned as a solution. It is necessary that $\|E_{A^*}\|_F^2$ is small enough to satisfy the conditions of the *rounding* technique for which $rounding(U_M R_{A^*}, K) = A^*$. On the other side M should be close to K because the probability of sampling a matrix close to R^* decreases exponentially with $M - K$. The number of iterations `max_rep` can be estimated using results on the principal angle between random subspaces, like [16]. A pseudocode of the algorithm is shown below. The time complexity of the algorithm depends on the computation of U_M, on the *rounding* procedure and on the graph's sparsity. For further details we refer to [3].

Alg. 1 spectral cut Rounding the first M eigenvectors of the Normalized Laplacian of W_G into K classes.

Input: A matrix W_G of a weighted graph $G = (V, E)$, a number of clusters K, a number M, $K \leq M \leq |V| - 1$, of eigenvectors.
Output: a partition of the graph $A = (A_1, \ldots, A_K)$

Let $L_N := I - D_G^{-1/2} W_G D_G^{-1/2}$ be the Normalized Laplacian of G,
let $\Lambda_M \in I\!R^{M \times M}$ be the diagonal matrix of its first M eigenvalues
$\lambda_1 \leq, \ldots, \leq \lambda_M$ and let $U_M \in I\!R^{|V| \times M}$ be the corresponding
eigenvectors.
for $k = 1, \ldots,$ `max_rep`
    Let $P \in I\!R^{M \times K}$ be a random matrix such that:
        First row of P has norm 1, $(PP^T)_{1,1} = 1$
        $P^T P = I$
    $(A_1, \ldots, A_K)_k := rounding(U_M P, K)$
    $NCut\_val_k := NCut(W_G, (A_1, \ldots, A_K)_k)$
$MinPos :=$ position of the minimum of $NCut\_val$
return $(A_1, \ldots, A_K)_{MinPos}$

We should notice that, depending on the problem, different good cuts may arise among the sampled matrices . If it is the case it should be reasonable to modify the algorithm and return a set of solutions instead of a single one.

5 Experiments

We tested the **spectral cut** on graphs from the Graph Partitioning Archive [1]. This repository contains a variety of graphs and updates the best cuts obtained with many state of the art algorithms. It was created as a follow up of [17]. We focus only on balanced bipartition, that is $K = 2$. The balance condition means that the solution classes must have the same number of vertices, so that the cardinality of each partition $|A_k|$ is $\leq \lceil |V|/2 \rceil$. In this case the **spectral cut** uses the eigenvectors of the Combinatorial Laplacian $L_C = D_G - W_G = U \Lambda U^T$ and the matrix H_A representing a cut is such that $(H_A)_{i,k} = 1/\sqrt{|A_k|}$ if $v_i \in A_k$ and 0 otherwise. The rounding technique we use is simply the median cut, i.e. comparison w.r.t. the median value of the vector. Unfortunately bound 1 applied to these graphs gives large values of M, on the other side for many graphs the norm of the benchmark solution B is well approximated by the first $M = 10$ eigenvectors, that is $\|E_B\|_F^2 \leq 0.15$, so we used this value in the experiments. In table 1 we report the results on graphs with less than $3 * 10^4$ vertices. The obtained cut values are within 105% from the benchmark and substantially improve the Fiedler Median Cut which considers only the second eigenvector.

graph	$\|V\|$	$\|E\|$	Min	Max	Mean	FMC	B	ME
data	2851	15093	**189**	204	197	260	189	0
3elt	4720	13722	**92**	112	100	117	90	40*
uk	4824	6837	**21**	28	26	31	20	8
add32	4960	9462	**11**	21	15	23	11	312*
whitaker3	9800	28989	**132**	136	136	136	127	370
crack	10240	30380	**190**	215	203	233	184	2244
fe_4elt2	11143	32818	**130**	130	130	130	130	676*
bcsstk29	13992	302748	**2863**	2978	2974	2978	2843	288
4elt	15606	45878	**145**	188	162	194	139	138*
bcsstk30	28924	1007284	**6408**	6620	6470	6620	6394	252

Table 1: Results for balanced cut on graphs from the Graph Partitioning Archive, two classes ($K = 2$) and 0% imbalance. For each graph we report its name, the number of vertices and edges. We repeated the **spectral cut** 10^2 times with `max_rep` $= 10^4$ and we show the minimum, maximum and mean (rounded to the nearest integer) cut values over the repetitions, the minimum often matches or nearly matches (≤ 1.05*B) the benchmark. In column B we show the benchmark value. In column ME we show the number of vertices different between the benchmark and the **spectral cut**, cuts with * have multiple solutions with the same value, in that case we report the greatest difference. In column FMC we report the results of the Fiedler Median Cut. The benchmark is considered on 26 february 2009.

[1] http://staffweb.cms.gre.ac.uk/∼wc06/partition/

6 Discussion

Our work goes in the direction of generalizing current Spectral Clustering algorithms. Many papers in Spectral Clustering, e.g. [4, 5, 7], show conditions for which the first K eigenvectors of the graph laplacian suffice for recovering the K classes we are interested in. These conditions are not always met in practice and, in the general case, the optimal solution may lie in some number of eigenvectors M strictly greater than the number of classes K. The **spectral cut** we propose is designed to recover such a solution and we also propose a bound to choose such M. The method is not a panacea, if $M - K$ is too large the time needed to find the solution can be quite high. Still experimental results empirically show that if the solution is in the span of the first M eigenvectors then it can be found or well-approximated. Other approaches do not have this guarantee.

References

[1] J. Shi and J. Malik. Normalized cuts and image segmentation. *PAMI*, 22, 1997.

[2] S. Guattery and G. L. Miller. On the performance of spectral graph partitioning methods. In *SODA*, 1995.

[3] N. Rebagliati and A. Verri. A randomized approach to spectral clustering. Technical Report DISI-TR-09-08, DISI, 2009.

[4] M. Meilă and J. Shi. Learning segmentation by random walks. In *NIPS*, 2000.

[5] A. Y. Ng, M. I. Jordan, and Y. Weiss. On spectral clustering: Analysis and an algorithm. In *NIPS*, 2001.

[6] C. H. Q. Ding and X. He. K-means clustering via principal component analysis. In *ICML*, 2004.

[7] F. R. Bach and M. I. Jordan. Learning spectral clustering, with application to speech separation. *JMLR*, 7, 2006.

[8] M. Fiedler. A property of eigenvectors of nonnegative symmetric matrices and its application to graph theory. *Czechoslovak Mathematical Journal*, 25, 1975.

[9] W.E. Donath and A.J. Hoffman. Lower bounds for the partitioning of graphs. *IBM J.Res.Develop.*, 17, 1973.

[10] C. J. Alpert and A. B. Kahng. Recent directions in netlist partitioning: A survey. *Integration: The VLSI Journal*, 19, 1995.

[11] J. A. Frankle and R. M. Karp. Circuit placement by eigenvector decomposition. In *ICCAD*, 1986.

[12] C. J. Alpert and S. Yao. Spectral partitioning: The more eigenvectors, the better. In *DAC*, 1995.

[13] T. Xiang and S. Gong. Spectral clustering with eigenvector selection. *Pattern Recogn.*, 41(3), 2008.

[14] U. von Luxburg. A tutorial on spectral clustering. *Statistics and Computing*, 17, 2007.

[15] M. Meilă. The uniqueness of a good optimum for k-means. In *ICML*, 2006.

[16] P.-A. Absil, A. Edelman, and P. Koev. On the largest principal angle between random subspaces. *Linear Algebra Appl.*, 414(1), 2006.

[17] A. J. Soper, C. Walshaw, and M. Cross. A combined evolutionary search and multilevel approach to graph partitioning. *London SE10 9LS*, 29, 2000.