# The Multicut Lemma

Marina Meilă
Department of Statistics
University of Washington

August, 2001 (Revised October, 2002)

## 1 Preliminaries

In the following, $W$ will be a symmetric $n \times n$ matrix with non-negative elements. The assumption is that $W_{ij}$ are weights corresponding to the edges $ij$ of a (complete) graph $(V, E)$ with vertices indexed by the numbers $1, \ldots n$. The indices $k, l$ will be used to index subsets of $V$ in a partition; we will call these subsets *clusters*. The indices $i, j$ will index elements of $V$.

We denote

$$d_i = \sum_{j \in V} W_{ij} \tag{1}$$

the degree of node $i \in V$ and by $D$ the diagonal matrix with $d_i, i \in V$ on the diagonal. The volume of $V$ is $Vol\ V = \sum_{i \in V} d_i$. By normalizing the rows of $W$ to sum to 1 we obtain the stochastic matrix $P$.

$$P = D^{-1}W \tag{2}$$

The Laplacian [Chung, 1997] of $W$ is

$$L = I - D^{-1/2} W D^{-1/2} \tag{3}$$

where $I$ is the unit matrix.

**Definition 1** *The $K$-way normalized cut associated to a partition $\Delta = (C_1, \ldots C_K)$ of $V$ is defined as*

$$MNCut(\Delta) \;=\; \sum_{k=1}^{K} \left[ 1 - \frac{\sum_{i \in C_k} \frac{d_i}{Vol V} \sum_{j \in C_k} P_{ij}}{\sum_{i \in C_k} \frac{d_i}{Vol V}} \right] \tag{4}$$

The intuition is that $\frac{d_i}{Vol V}$ represents the stationary distribution of the Markov chain defined by $P$ and consequently each term of the sum represents the probability of leaving cluster $C_k$ given that the Markov chain is in $C_k$, under the stationary distribution. The multiway normalized cut is the sum of these conditional probabilities.

**Definition 2** *If $\Delta = (C_1, \ldots C_K)$ is a partition of $V$, we say that a vector $x$ is* piecewise constant *w.r.t $\Delta$ if for all pairs $i, j$ in the same cluster $C_k$ we have $x_i = x_j$.*

Other results that will be used are grouped in the following lemmas. The first three lemmas were proved in [Meilă and Shi, 2001], while lemma 6 is proved in the Appendix.

**Lemma 3 (Relationship between Laplacian and Markov random walk)** *Denote by $1 = \lambda_1 \geq \lambda_2 \geq \ \ldots \ \lambda_n \geq -1$ the eigenvalues of $P$ and by $v^1, \ldots v^n$ the corresponding eigenvectors. Denote by $0 = \mu_1 \leq \mu_2 \leq \ \ldots \ \mu_n$ the eigenvalues of $L$ and by $u^1, \ldots u^n$ the corresponding eigenvectors. Then,*

$$\mu_i \;=\; 1 - \lambda_i \tag{5}$$

$$u^i \;=\; D^{1/2} v^i \tag{6}$$

*for all $i = 1, \ldots n$.*

Note that this lemma ensures that the eigenvalues of $P$ are always real and the eigenvectors lineraly independent.

**Lemma 4 (Lumpability)** *Let $P$ be a matrix with rows and columns indexed by $V$ that has independent eigenvectors. Let $\Delta = (C_1, C_2, \ldots C_k)$ be a partition of $V$. Then, $P$ has $K$ eigenvectors that are piecewise constant* *w.r.t.* $\Delta$ *and correspond to non-zero eigenvalues if and only if the sums* $P_{ik} = \sum_{j \in C_k} P_{ij}$ *are constant for all* $i \in C_l$ *and all* $k, l = 1, \ldots k$ *and the matrix* $\hat{P} = [\hat{P}_{kl}]_{k,l=1,\ldots K}$ *(with* $\hat{P}_{kl} = \sum_{j \in C_k} P_{ij}$, $i \in C_l$*) is non-singular.*

**Lemma 5 (Relationship between $P$ and $\hat{P}$)** *Assume that the conditions of Lemma 4 hold. Let* $v^1, \ldots v^K$ *and* $1 = \lambda_1 \geq \lambda_2 \geq \ \ldots \ \lambda_K$ *be the piecewise constant eigenvectors of* $P$ *and their eigenvalues. Denote by* $1 = \hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \ldots \ \hat{\lambda}_K$ *and*$\hat{v}^1, \ldots \hat{v}^K$ *the eigenvalues and eigenvectors of* $\hat{P}$*. Then*

$$\hat{\lambda}_k = \lambda_k \quad (7)$$

$$\hat{v}_l^k = v_i^k \quad \text{for } l = 1, \ldots K \quad \text{and } i \in C_l \quad (8)$$

**Lemma 6 (A generalized Rayleigh quotient)** *Let* $L$ *be a symmetric matrix* $n \times n$ *with eigenvalues* $0 = \mu_1 \leq \mu_2 \leq \ \ldots \ \mu_n$ *and corresponding eigenvectors* $u^1, \ldots u^n$*. Assume that the eigenvectors are normalized to have length 1. Then*

$$\min \sum_{k=1}^{K} (y^k)^T L y^k \ \text{ s.t. } (y^l)^T y^k \ = \ \delta_{kl}$$

*is* $\sum_{k=1}^{K} \mu_k$ *and the minimizing attained for* $y^k$, $k = 1, \ldots K$ *lie in the subspace spanned by* $u^1, \ldots u^K$*.*

## 2 The multicut lemma

**Lemma 7 The Multicut Lemma** *Let* $W$ *be an* $n \times n$ *symmetric matrix with nonnegative elements, and let* $P$ *be the stochastic matrix obtained by normalizing the rows of* $W$ *to sum to 1. Assume that* $P$ *has* $K$ *piecewise constant eigenvectors w.r.t a partition* $\Delta^* = (C_1^*, \ldots C_K^*)$*. Denote these eigenvectors by* $v^1, \ldots v^K$ *and their corresponding eigenvalues by* $\lambda_1, \ldots \lambda_K$*. Assume also that* $P$ *has* $n$ *distinct eigenvalues and that* $\lambda_1, \ldots \lambda_K$ *are the* $K$ *largest eigenvalues of* $P$ *and are all non-zero.*

*Then the minimum K-way normalized cut for* $W$ *is given by the partition* $\Delta^*$*.*

**Proof.** First let us compute the value of $MNCut(\Delta^*)$.

$$\begin{aligned} MNCut(\Delta^*) &= \sum_{k=1}^{K}\left[1 - \sum_{j \in C_k^*} P_{i_k j}\right] \quad \text{for some } i_k \in C_k && (9)\\ &= \sum_{k=1}^{K}(1 - \hat{P}_{kk}) && (10)\\ &= K - \text{trace}\hat{P} && (11)\\ &= K - \sum_{k=1}^{K} \lambda_k && (12) \end{aligned}$$

In the above derivations I used the results of Lemma 4.

We will show that there is no $K$-way cut that achieves a value smaller than $K - \sum_{k=1}^{K} \lambda_k$. Consider an arbitrary partition $\Delta = (C_1, \ldots C_K)$. Denote by $x^k$ the indicator vector of cluster $C_k$ for $k = 1, \ldots K$.

Let us massage the expression of $MNCut(\Delta)$ into a convenient form:

$$\begin{aligned} MNCut(\Delta) &= K - \sum_{k=1}^{K} Pr[C_k \to C_k \,|\, C_k] && (13)\\ &= K - \sum_{k=1}^{K} \frac{\sum_{i \in C_k} d_i \sum_{j \in C_k} P_{ij}}{\sum_{i \in C_k} d_i} && (14)\\ &= K - \sum_{k=1}^{K} \frac{\sum_{i,j \in C_k} W_{ij}}{\sum_{i \in C_k} d_i} && (15) \end{aligned}$$

Noting that

$$\sum_{i \in C_k} d_i = \sum_{i \in V} (x_i^k)^2 d_i \qquad (16)$$

and

$$\begin{aligned} \sum_{i,j \in C_k} W_{ij} &= \sum_{i,j \in V} W_{ij} x_i^k x_j^k && (17)\\ &= \frac{1}{2} \sum_{i,j \in V} W_{ij}[(x_i^k)^2 + (x_j^k)^2 - (x_i^k - x_j^k)^2] && (18)\\ &= \sum_{i \in V} (x_i^k)^2 d_i - \sum_{ij \in E} W_{ij}(x_i^k - x_j^k)^2 && (19) \end{aligned}$$

we obtain that

$$MNCut(\Delta) = \sum_{k=1}^{K} \frac{\sum_{ij\in E} W_{ij}(x_i^k - x_j^k)^2}{\sum_{i\in V}(x_i^k)^2 d_i} \quad (20)$$

$$= \sum_{k=1}^{K} R(x^k) \quad (21)$$

In the sums above, $i, j \in V$ means summation over the cartesian product $V \times V$ while $ij \in E$ means summation over all "edges", i.e all unordered pairs $(i, j)$ with $i \neq j$.

The expression $R(x)$ represents the Rayleigh quotient for the Laplacian of the weighted graph described by $W$ c.f.[Chung, 1997] equation (1.13). Therefore, if we denote

$$y^k = D^{1/2}x^k \quad (22)$$

we have that

$$R(x^k) = \frac{(y^k)^T L y^k}{(y^k)^T y^k} = \tilde{R}(y^k) \quad (23)$$

and

$$MNCut(\Delta) = \sum_{k=1}^{K} \tilde{R}(y^k) \quad (24)$$

Now we turn to the problem of minimizing $MNCut$. It is easy to see that minimizing $MNCut(\Delta)$ over all partitions $\Delta$ is equivalent to finding

$$\min \sum_{k=1}^{K} R(x^k) \text{ s.t. } x_i^k \in \{0, 1\} \ \forall\, k, i \text{ and } x^k \perp Dx^l \text{ for } k \neq l \quad (25)$$

This minimum is greater or equal to

$$\min \underbrace{\sum_{k=1}^{K} R(x^k)}_{J(x^1,\dots x^K)} \text{ s.t. } x^k \perp Dx^l \text{ for } k \neq l \quad (26)$$

Now we focus on $J$ and show that its minimum in (26) is equal to $MNCut(\Delta^*)$ which will prove that the latter is the smallest achievable $K$-way normalized cut.

With $y^1, \dots y^K$ defined as in (22) we have that

$$x^k \perp Dx^l \;\Leftrightarrow\; y^k \perp y^l. \tag{27}$$

In addition, we can assume w.l.o.g. that the vectors $y^k$ have unit length. Therefore, minimizing $J$ is equivalent to finding

$$\min \tilde{J}(y^1, \dots y^k) \;=\; \min \sum_{k=1}^{K} (y^k)^T L y^k \;\text{ s.t. }\; (y^l)^T y^k \;=\; \delta_{kl} \tag{28}$$

By Lemma 6, the minimum is attained whey $y^k = u^k$ the eigenvectors of $L$ corresponding to its $K$ smallest eigenvalues. Its value is $\sum_{k=1}^{K} \mu_k$.

Now we use Lemma 3 which says that $\mu_k = 1 - \lambda_k$. This proves that

$$\min \tilde{J} \;=\; \min J \;=\; MNCut(\Delta^*) \tag{29}$$

In addition, note that in this case the minimizing vectors $y^k$ are linear combinations of the first $K$ eigenvectors of $L$. Again, by Lemma 3 we have that $x^1, \dots x^K$ are linear combinations of the first $K$ eigenvectors of $P$ which are piecewise constant.

## Appendix: Proof of Lemma 6

Denote by $u^1, \dots u^n$ the orthonormal eigenvectors of $L$. The $y$'s are linear combinations of these vectors, i.e

$$y^k \;=\; \sum_{j=1}^{n} a_{jk} u^k \tag{30}$$

Denote by $A$ the $n \times K$ matrix $[a_{jk}]_{j,k}$. Because both $\{u^j\}$ and $\{y^k\}$ are orthonormal, the vectors $\{a_{.k}\}$ (i.e the colums of $A$) are orthonormal as well. We minimize $\tilde{J}$ w.r.t $\{a_{jk}\}$ by the Lagrange multiplier method.

$$\tilde{J}_\beta \;=\; \sum_{j=1}^{n} \sum_{k=1}^{K} \mu_j a_{jk}^2 - \sum_{k>l} \beta_{kl} a_{.l}^T a_{.k} - \sum_{k=1}^{k} \beta_k (a_{.k}^T a_{.k} - 1) \tag{31}$$

Because

$$(y^k)^T L y^k \;=\; \sum_{j=1}^{n} \mu_j a_{jk}^2. \tag{32}$$

$$\frac{\partial \tilde{J}}{\partial a_{jk}} = 2\mu_j a_{jk} - \sum_{k \neq l} a_{jl}\beta_{jk} - 2\sum_{k=1}^{K} \beta_k a_{jk} \tag{33}$$

Equating the above partial derivative with 0, and defining the matrix $B$ to be $B_{kl} = B_{lk} = 1/2\beta_{kl}$ for $k \neq l$ and $B_{kk} = \beta_k$ we obtain

$$Ba_{j\cdot} = \mu_j a_{j\cdot} \text{ for } j = 1, \dots n \tag{34}$$

Thus the vectors $a_{j\cdot}$ are either eigenvectors of $B$ or 0. There are $n$ distinct eigenvalues $\lambda_j$ while the matrix $B$ is $K \times K$ so it can have only $K$ eigenvalues and corresponding independent eigenvectors. Hence, at most $K$ rows of $A$ are non-zero. The non-zero rows are orthogonal, because $B$ is symmetric. On the other hand, $A$ has orthonormal columns, therefore there will be at least $K$ non-zero rows in $A$. Denote by $H$ the $K \times K$ matrix obtained from $A$ by eliminating the null rows. The columns of $H$ are orthonormal, therefore its rows must be orthonormal as well. Denote by $j_1, \dots j_K$ the indices of these rows in $A$.

The value of $\tilde{J}$ under the conditions above becomes

$$\tilde{J} = \sum_{i=1}^{K} \mu_{j_i} \tag{35}$$

Thus, $\tilde{J}$ has several local minima, one for each possible subset of indices $j_1, \dots j_K$. For each of them, the value of the minimum is the sum of the selected eigenvalues and the minimizing vectors $y^k$ are linear combinations of the eigenvectors $u^{j_1}, \dots u^{j_K}$. The lowest of the minima corresponds to chosing the $K$ smallest eigenvalues of the Laplacian, i.e. $\mu_1, \dots \mu_K$.

## Acknowledgement

This research was partially supported by University of Washington Royalty Research Fund grant 2684.

## References

[Chung, 1997] Chung, F. R. K. (1997). *Spectral Graph Theory*. Number Regional Conference Series in Mathematics in 92. American Mathematical Society, Providence, RI.

[Meilă and Shi, 2001] Meilă, M. and Shi, J. (2001). A random walks view of spectral segmentation. In Jaakkola, T. and Richardson, T., editors, *Artificial Intelligence and Statistics AISTATS.*