

---

This is an electronic reprint of the original article.  
This reprint may differ from the original in pagination and typographic detail.

Orponen, Pekka; Schaeffer, Satu Elisa; Gaytán, Vanesa Avalos

**Locally computable approximations for spectral clustering and absorption times of random walks**

Published: 01/01/2008

*Document Version*

Early version, also known as pre-print

*Please cite the original version:*

Orponen, P., Schaeffer, S. E., & Gaytán, V. A. (2008). Locally computable approximations for spectral clustering and absorption times of random walks. <http://arxiv.org/abs/0810.4061>

---

This material is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the repository collections is not permitted, except that material may be duplicated by you for your research use or educational purposes in electronic or print form. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone who is not an authorised user.

# Locally computable approximations for spectral clustering and absorption times of random walks

Pekka Orponen<sup>†</sup>, Satu Elisa Schaeffer<sup>‡,1</sup>, and Vanesa Avalos-Gaytán<sup>‡</sup>

<sup>†</sup> Department of Information and Computer Science,  
Helsinki University of Technology TKK,  
FI-02015 TKK Espoo, Finland  
`pekka.orponen@tkk.fi`

<sup>‡</sup> School of Mechanical and Electrical Engineering,  
Universidad Autónoma de Nuevo León (UANL),  
Ciudad Universitaria, San Nicolás de los Garza, NL 66450, Mexico,  
`{elisa,vanesa}@yalma.fime.uanl.mx`

<sup>1</sup> Corresponding author, tel. +52 81 1340 4000, fax +52 81 1052 3321

October 20, 2008

## Abstract

We address the problem of determining a natural local neighbourhood or “cluster” associated to a given seed vertex in an undirected graph. We formulate the task in terms of absorption times of random walks from other vertices to the vertex of interest, and observe that these times are well approximated by the components of the principal eigenvector of the corresponding fundamental matrix of the graph’s adjacency matrix. We further present a locally computable gradient-descent method to estimate this Dirichlet-Fiedler vector, based on minimising the respective Rayleigh quotient. Experimental evaluation shows that the approximations behave well and yield well-defined local clusters.

**Key words:** graph clustering, spectral clustering, random walk, absorption time, gradient method

**AMS Classification:** 05C50, 05C85, 68R10, 68W25, 90C27, 90C52, 90C59, 94C15

## 1 Introduction and motivation

### 1.1 Nonuniform networks

The field of natural-network study became popular when Watts and Strogatz [1] published their observations on the short average path length and the high

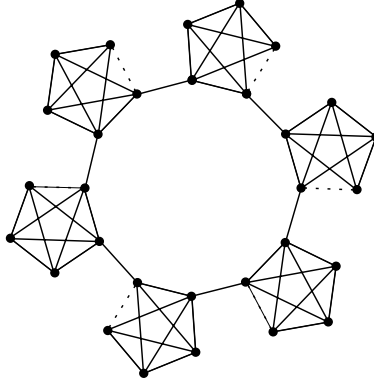


Figure 1: A *caveman graph* [13] composed of six near-cliques of five vertices each that have been connected into a circulant graph by “opening” one edge from each clique (the removed edge is shown with a dotted line).

clustering coefficient of many natural graphs, followed by the observations of scale-free distributions [2, 3] in the degrees and other structural properties of such networks. As a consequence of the resulting wide interest in the properties of natural networks, there now exist numerous models to meet the observations made on natural networks [4, 5, 6].

## 1.2 Graph clustering

One of the properties of interest in the field of natural graphs is the presence of *clusters* or *communities* [7], that is, the existence of dense induced subgraphs that have relatively few connections outside compared to the internal density [8].

*Graph clustering* is the task of grouping the vertices of the graph into clusters taking into consideration the edge structure of the graph in such a way that there should be many edges *within* each cluster and relatively few *between* the clusters. For an artificial example, see Figure 1 that illustrates a small graph with a clear six-cluster structure. Another classic example is a small real-world social network studied by Zachary [9] and often referred to in graph clustering papers [5, 10, 11]. It is a social network of a small karate club that was just about to split into two (see Figure 2), making it an ideal case for two-classification algorithms. For a survey on graph-clustering algorithms, see [12].

## 1.3 Local clustering

In *local clustering*, the goal is to find the cluster of a given *seed vertex*  $s \in V$ . Hence, essentially, it is the task of finding a *bipartition* of the graph  $G$  into two vertex sets  $S$  and  $V \setminus S$  such that  $s \in S$  and  $S$  makes a good cluster in some predefined sense. Common cluster quality criteria include *cut capacity* and

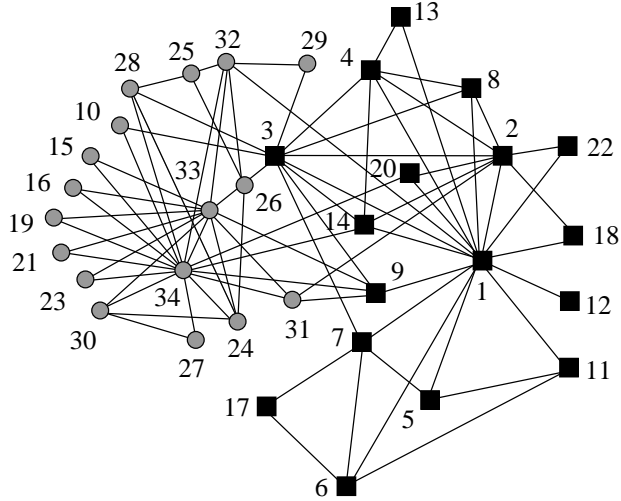


Figure 2: The karate club social network studied by Zachary [9]. The two groups into which the club split are indicated by the shape with which the vertices are drawn: the squares later formed their own club, and the circles formed another club.

related measures such as *conductance* [14] or density-based measures [15]. Also methods motivated by electric networks have been proposed for global and local clustering alike [10, 11, 16].

#### 1.4 Spectra of graphs

Let  $G = (V, E)$  be an unweighted undirected connected graph with at least two vertices. For simplicity, we focus on unweighted graphs, although much of what follows can easily be generalised to the weighted case. Denote the order of  $G$ , i.e. its number of vertices, by  $n$  and identify each vertex  $v$  with a label in  $\{1, 2, \dots, n\}$ . Denote the seed vertex by  $s$ . The *adjacency matrix* of  $G$  is the binary matrix  $\mathbf{A}$ , where  $a_{ij} = 1$  if edge  $\{i, j\}$  is in  $E$ , and otherwise  $a_{ij} = 0$ .

For a weighted graph, one would consider instead the analogous *edge-weight matrix*. Note also that for multigraphs, edge multiplicities can in the present context be considered simply as integer weights. For an undirected graph, the adjacency (resp. edge weight) matrix is symmetric, whereas directed graphs pose further complications in the algebraic manipulation — we refer the reader to the textbook and other works of Chung [17, 18, 19, 20] for properties and local clustering of directed graphs.

The *degree*  $d_v$  of a vertex  $v$  is the number (resp. total weight) of its incident edges; thus the components of the *degree vector*  $\mathbf{d}$  of  $G$  are the row sums of  $\mathbf{A}$ . Denote by  $\mathbf{D}$  the diagonal  $n \times n$  matrix formed by setting the diagonal elements to  $d_{ii} = d_i$  and all other elements to zero.

Let  $\mathbf{I}$  be the  $n \times n$  unit matrix. The *Laplacian* matrix of  $G$  is  $\mathbf{L} = \mathbf{D} - \mathbf{A}$  and the *normalised Laplacian* matrix of  $G$  is  $\mathcal{L} = \mathbf{D}^{-\frac{1}{2}} \mathbf{L} \mathbf{D}^{-\frac{1}{2}} = \mathbf{I} - \mathbf{D}^{-\frac{1}{2}} \mathbf{A} \mathbf{D}^{-\frac{1}{2}}$ . Since both  $\mathbf{L}$  and  $\mathcal{L}$  are symmetric, all their eigenvalues are real. It turns out that  $\mathcal{L}$  is in some respects a more natural object of study than  $\mathbf{L}$ , and we shall mostly focus on that. It is easy to see that zero is an eigenvalue of both  $\mathbf{L}$  and  $\mathcal{L}$ , and for  $\mathcal{L}$  it can be shown that all the other  $n - 1$  eigenvalues (counting multiplicities) lie in the interval  $[0, 2]$ . Denote these in increasing order as  $0 = \mu_0 \leq \mu_1 \leq \dots \leq \mu_{n-1} \leq 2$ , and let  $\mathbf{u}_i$  be some right eigenvector associated to  $\mu_i$ . We may assume that the distinct eigenvectors  $\mathbf{u}_i$  are orthogonal to each other. For more information on the spectral and algebraic properties of graphs, see e.g. the excellent monographs of Biggs [21] and Chung [17].

## 1.5 Random walks

The *simple random walk* on a graph  $G$  is a Markov chain where each vertex  $v \in V$  corresponds to a state and the transition probability from state  $i$  to state  $j$  is  $p_{ij} = d_i^{-1}$  if  $\{i, j\} \in E$  and zero otherwise. For a weighted graph,  $p_{ij}$  is the ratio of the weight of edge  $\{i, j\}$  to the total weight of edges incident to  $i$ .

Denote the transition probability matrix of this Markov chain by  $\mathbf{P} = \mathbf{D}^{-1} \mathbf{A}$ . Note that even for undirected graphs,  $\mathbf{P}$  is not in general symmetric. However, it is similar to the matrix

$$\mathcal{P} = \mathbf{D}^{\frac{1}{2}} \mathbf{P} \mathbf{D}^{-\frac{1}{2}} = \mathbf{D}^{-\frac{1}{2}} \mathbf{A} \mathbf{D}^{-\frac{1}{2}} \quad (1)$$

which *is* symmetric because  $\mathbf{A}$  is the adjacency matrix of an undirected graph. Thus,  $\mathbf{P}$  and  $\mathcal{P}$  have the same spectrum of eigenvalues, which are all real. Moreover,

$$\begin{aligned} \mathcal{L} &= \mathbf{D}^{-\frac{1}{2}} \mathbf{L} \mathbf{D}^{-\frac{1}{2}} = \mathbf{D}^{-\frac{1}{2}} (\mathbf{D} - \mathbf{A}) \mathbf{D}^{-\frac{1}{2}} \\ &= \mathbf{D}^{-\frac{1}{2}} (\mathbf{D} - \mathbf{D} \mathbf{P}) \mathbf{D}^{-\frac{1}{2}} = \mathbf{I} - \mathbf{D}^{\frac{1}{2}} \mathbf{P} \mathbf{D}^{-\frac{1}{2}} \\ &= \mathbf{I} - \mathcal{P}. \end{aligned} \quad (2)$$

Consequently,  $\lambda$  is an eigenvalue of the normalised transition matrix  $\mathcal{P}$  if and only if  $\mu = 1 - \lambda$  is an eigenvalue of the normalised Laplacian matrix  $\mathcal{L}$ . Thus,  $\mathbf{P}$ ,  $\mathcal{P}$  and  $\mathcal{L}$  have the following correspondence:  $\mathbf{v}$  is a right eigenvector associated to eigenvalue  $\lambda$  in  $\mathbf{P}$  if and only if  $\mathbf{u} = \mathbf{D}^{\frac{1}{2}} \mathbf{v}$  is a right eigenvector associated to the same eigenvalue in  $\mathcal{P}$ , and to eigenvalue  $1 - \lambda$  in  $\mathcal{L}$ .

Since in the case of Markov chains, *left* eigenvectors are also of interest, let us note in passing that the analogous correspondence holds between each left eigenvector  $\pi$  of  $\mathbf{P}$  and left eigenvector  $\rho = \pi \mathbf{D}^{-\frac{1}{2}}$  of  $\mathcal{P}$  or  $\mathcal{L}$ .

Denote the eigenvalues of  $\mathbf{P}$  in decreasing order as  $\lambda_0^{(\mathbf{P})} \geq \lambda_1^{(\mathbf{P})} \geq \dots \geq \lambda_{n-1}^{(\mathbf{P})}$ . Since  $\mathbf{P}$  is a stochastic matrix, it always has eigenvalue  $\lambda_0^{(\mathbf{P})} = 1$ , corresponding to the smallest Laplacian eigenvalue  $\mu_0^{(\mathcal{L})} = 0$ . All the other eigenvalues of  $\mathbf{P}$  satisfy  $|\lambda_i^{(\mathbf{P})}| \leq 1$ . If moreover  $G$  is connected and not bipartite, the Markov chain determined by  $\mathbf{P}$  is ergodic, in which case  $|\lambda_i^{(\mathbf{P})}| < 1$  for all  $i \geq 1$ . Without much loss of generality, we shall assume this condition, and moreover that all

the eigenvalues  $\lambda_i^{(\mathbf{P})}$  are nonnegative. Both of these conditions can be enforced by considering, if necessary, instead of  $\mathbf{P}$  the “lazy random walk” with transition matrix

$$\mathbf{P}' = \frac{1}{2}(\mathbf{I} + \mathbf{P}). \quad (3)$$

For a connected graph  $G$  this chain is ergodic, and has nonnegative eigenvalues

$$\lambda_i^{(\mathbf{P}')} = \frac{1}{2}(1 + \lambda_i^{(\mathbf{P})}), \quad (4)$$

with the same eigenvectors as  $\mathbf{P}$ .

Let us then consider a transition matrix  $\hat{\mathbf{P}}$  obtained from  $\mathbf{P}$  by making a given state, or vertex  $s$  *absorbing*. Thus,  $\hat{\mathbf{P}}$  is otherwise equal to  $\mathbf{P}$ , but all  $\hat{p}_{si} = 0$  except for  $\hat{p}_{ss} = 1$ . We shall henceforth assume, for simplicity of notation, that  $s = n$ , so that in particular  $\hat{\mathbf{P}}$  has the block structure:

$$\hat{\mathbf{P}} = \left( \begin{array}{c|c} & \begin{matrix} p_1 \\ \vdots \\ p_{n-1} \end{matrix} \\ \hline \mathbf{Q} & \\ \hline 0 \dots 0 & 1 \end{array} \right) \quad (5)$$

The *absorption time*  $m_i$  from vertex  $i \neq s$  to the seed vertex  $s$  is the expected number of steps that a walk initiated at  $i$  will take before hitting  $s$ . Intuitively, as the absorption time measures in a certain sense the proximity of vertex  $i$  to vertex  $s$ , vertices belonging to a good cluster  $S$  for  $s$ , if such a cluster exists, should have characteristically smaller absorption times to  $s$  than vertices in  $V \setminus S$ . Note that not all graphs exhibit a clustered structure, in which case no clustering method will be able to pinpoint a high-quality cluster [12].

It is well known that the absorption times to vertex  $s = n$  can be calculated as row sums

$$m_i = m_{i,1} + m_{i,2} + \dots + m_{i,n-1}. \quad (6)$$

from the *fundamental matrix*

$$\mathbf{M} = \mathbf{I} + \mathbf{Q} + \mathbf{Q}^2 + \mathbf{Q}^3 + \dots = (\mathbf{I} - \mathbf{Q})^{-1}, \quad (7)$$

where  $\mathbf{Q}$  is the matrix obtained from  $\hat{\mathbf{P}}$  (or equivalently from  $\mathbf{P}$ ) by eliminating the row and column corresponding to vertex  $s = n$  (as shown above in Equation (5)),

In Figure 3, we illustrate the absorption times in the caveman graph of Figure 1: we computed with Matlab the absorption times from all vertices to a given seed vertex  $j$ , repeated the computation for each  $j \in V$ , and formed a matrix where each column represents the absorption-time vector for the corresponding vertex  $j$ . The columns are ordered so that all absorption-time vectors

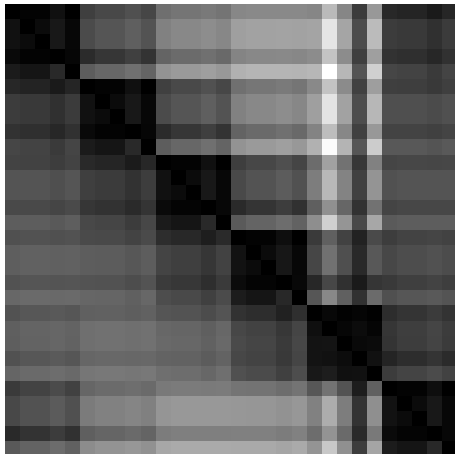


Figure 3: The absorption time matrix composed of 30 absorption-time vectors using each vertex of the caveman graph of Figure 1 in turn as a seed vertex, with white corresponding to the maximum  $m_{i,j}$  thus obtained and black corresponding to the minimum  $m_{i,j}$  and the diagonal zeroes.

associated to a given cave are grouped together, before those of the next cave, and so forth. The matrix is visualised as a gray-scale colour map by placing a tiny *black* square where either  $m_{i,j} = 0$  (that is, along the diagonal) or  $m_{i,j} = 10.6$  (the minimal off-diagonal absorption time observed), a *white* square where  $m_{i,j} = 319.6$  (the maximum observed), and discretising the intermediate values to 254 gray-scale colours correspondingly. The caves can be distinguished as dark five-by-five blocks along the diagonal, although the matrix is somewhat too noisy to be trivially clustered.

Now consider the eigenvalue spectra of matrices  $\hat{\mathbf{P}}$  and  $\mathbf{Q}$ . Matrix  $\hat{\mathbf{P}}$  is still stochastic, so it has largest eigenvalue  $\lambda_0^{(\hat{\mathbf{P}})} = 1$ , and since the chain is absorbing, all the other eigenvalues satisfy  $|\lambda_i^{(\hat{\mathbf{P}})}| < 1$ ,  $i = 1, \dots, n - 1$ .

Denote  $\mathcal{Q} = \mathbf{D}^{\frac{1}{2}} \mathbf{Q} \mathbf{D}^{-\frac{1}{2}}$ , where  $\mathbf{D} = \text{diag}(d_1, \dots, d_{n-1})$ . As  $\mathcal{Q}$  is symmetric (it is obtained by eliminating the last row and column from the symmetric matrix  $\mathcal{P}$ ) and  $\mathbf{Q}$  is similar to  $\mathcal{Q}$ , both have a spectrum of real eigenvalues  $\text{Spec}(\mathbf{Q}) = \{\lambda_1^{(\mathbf{Q})} \geq \dots \geq \lambda_{n-1}^{(\mathbf{Q})}\}$ . This spectrum is properly contained in the interval  $[-1, 1]$ , because for any vertex  $i \neq n$  adjacent to  $n$ ,  $p_{in} > 0$ , and so the  $i^{\text{th}}$  row sum of  $\mathbf{Q}$  is less than 1.

We claim that in fact

$$\text{Spec}(\mathbf{Q}) = \text{Spec}(\hat{\mathbf{P}}) \setminus \{1\}. \quad (8)$$

To prove this claim, let namely  $\lambda \neq 1$  be any non-principal eigenvalue of  $\hat{\mathbf{P}}$  and  $\mathbf{v}$  a corresponding eigenvector, so that  $\hat{\mathbf{P}}\mathbf{v} = \lambda\mathbf{v}$ . Since the  $n^{\text{th}}$  row of  $\hat{\mathbf{P}}$  is zero except for  $\hat{p}_{nn} = 1$ , it follows that  $\lambda v_n = (\hat{\mathbf{P}}\mathbf{v})_n = v_n$ , and since  $\lambda \neq 1$  that

necessarily  $v_n = 0$ . Then for the  $(n - 1)$ -dimensional vector  $\mathbf{v}' = (v_1, \dots, v_{n-1})$  and for any  $i = 1, \dots, n - 1$  it holds that:

$$\begin{aligned}
 (\mathbf{Q}\mathbf{v}')_i &= \sum_{j=1}^{n-1} p_{ij}v'_j = \sum_{j=1}^{n-1} p_{ij}v_j = \sum_{j=1}^n p_{ij}v_j - p_{in}v_n \\
 &= (\hat{\mathbf{P}}\mathbf{v})_i - v_n p_{in} = (\hat{\mathbf{P}}\mathbf{v})_i \\
 &= \lambda v_i = \lambda v'_i.
 \end{aligned} \tag{9}$$

Consequently,  $\mathbf{v}'$  is an eigenvector associated to eigenvalue  $\lambda$  of  $\mathbf{Q}$ . Since  $\lambda$  was chosen arbitrarily from  $\text{Spec}(\hat{\mathbf{P}}) \setminus \{1\}$ , this establishes that  $\text{Spec}(\hat{\mathbf{P}}) \setminus \{1\} \subseteq \text{Spec}(\mathbf{Q})$ . For the converse direction, a similar argument shows that if  $\mathbf{v}' = (v_1, \dots, v_{n-1})$  is an eigenvector associated to an eigenvalue  $\lambda$  of  $\mathbf{Q}$ , then the vector  $\mathbf{v} = (v_1, \dots, v_{n-1}, 0)$  is an eigenvector associated to eigenvalue  $\lambda$  of  $\hat{\mathbf{P}}$ .

## 2 Spectral methods for bipartitioning

### 2.1 Fiedler vectors

Spectral clustering of points in space, often modelled as (complete) weighted graphs, is a widely studied topic [22, 23]. In the context of graphs, the technique is usually applied so that some right eigenvector associated to the smallest nonzero eigenvalue  $\mu_1^{(\mathbf{L})}$  of  $\mathbf{L}$  is used to produce a bipartitioning of the graph such that those vertices that have negative values in the eigenvector form one side of the bipartition  $S$  and the vertices with positive values are the other side  $S \setminus V$ . These eigenvectors are called *Fiedler vectors* following [24, 25], where the technique was first proposed. The corresponding eigenvectors based on  $\mathcal{L}$  are called *normalised Fiedler vectors*. The works on Fiedler-vector based spectral clustering are numerous and go back for decades [26, 27, 28]

For our example graph illustrated in Figure 1, such a bipartition based on  $\mathbf{L}$  puts three of the caves in  $S$  such that it assigns negative values to every other cave along the cycle of six caves. Using the eigenvector of  $\mathcal{L}$ , however, assigns only negative values in the vector and does not yield an intuitive division that preserves the caves. The two vectors are visualised in Figure 4.

If there are only two natural clusters in the graph, such bipartition works nicely. An example is the Zachary karate club network of Figure 2: the corresponding Fiedler vectors are shown in Figure 5. Also, recursively performing bipartitions on the subgraphs induced by  $S$  and  $V \setminus S$  will help cluster the input graph  $G$  in more than two clusters, but a *stopping condition* needs to be imposed to determine when to stop bipartitioning the resulting subgraphs further.

### 2.2 Spectral partitioning as integer program relaxation

The use of Fiedler vectors for graph bipartitioning can be motivated as follows (see for example [22]). Denote a *cut* (bipartition) of a graph  $G = (V, E)$  into



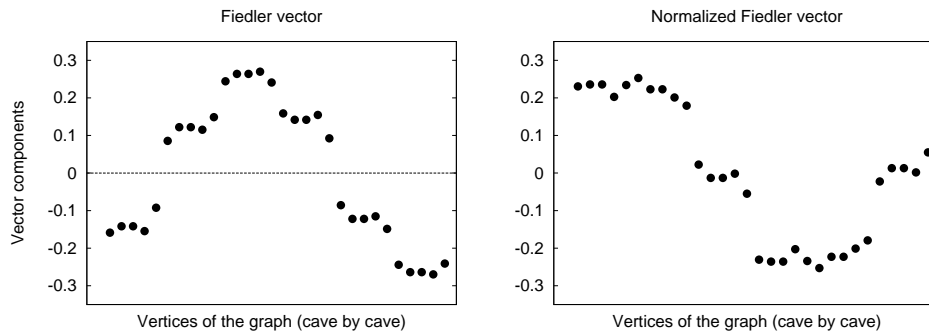


Figure 4: The components of the Fiedler vector (left) and the normalised Fiedler vector (right) for the caveman graph of Figure 1. For the human eye, the six-cluster structure is evident in the Fiedler vectors, whereas in the normalised Fiedler vector the vertices are grouped into four clusters (two of them consisting of two caves).

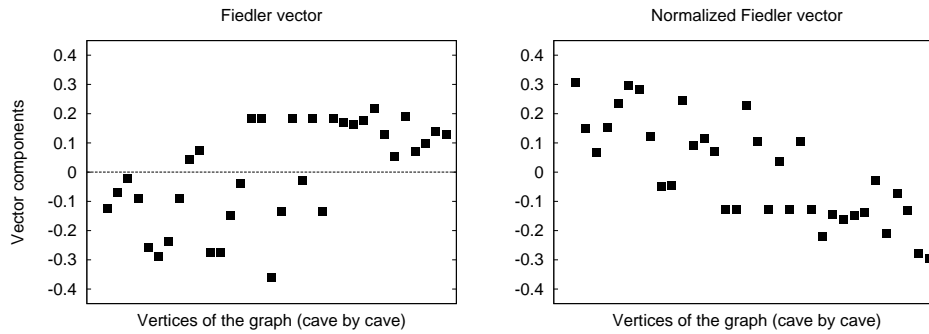


Figure 5: The components of the Fiedler vector (left) and the normalised Fiedler vector (right) for the karate club graph of Figure 2. The vertices can be classified in two groups: those with positive values in the Fiedler vector and those with negative values.

vertex sets  $S$  and  $\bar{S} = V \setminus S$  as  $(S, \bar{S})$ . The *capacity* of a cut  $(S, \bar{S})$  is defined as

$$C(S, \bar{S}) = |\{\{i, j\} \in E : i \in S, j \in \bar{S}\}|. \quad (10)$$

A cut  $(S, \bar{S})$  can be conveniently represented by an indicator vector  $\mathbf{v} \in \{+1, -1\}^n$ , where  $v_i = +1$  if  $i \in S$ , and  $v_i = -1$  if  $i \in \bar{S}$ .

Then

$$C(S, \bar{S}) = \frac{1}{4} \sum_{i \sim j} (v_i - v_j)^2, \quad (11)$$

where the sum is over all the (undirected) edges  $\{i, j\} \in E$ .

For simplicity, assume now that  $|V| = n$  is even, and consider the task of finding an *optimal bisection* of  $G$ , i.e. a cut  $(S, \bar{S})$  that satisfies  $|S| = |\bar{S}| = n/2$  and minimises  $C(S, \bar{S})$  subject to this condition.

This is equivalent to finding an indicator vector  $\mathbf{v} \in \{+1, -1\}^n$  that satisfies  $\sum_i v_i = 0$  and minimises the quadratic form  $\sum_{i \sim j} (v_i - v_j)^2$ , or equivalently (since  $n$  is fixed) minimises the ratio:

$$\begin{aligned} \frac{\frac{1}{4} \sum_{i \sim j} (v_i - v_j)^2}{n/4} &= \frac{\sum_{i \sim j} (v_i - v_j)^2}{n} \\ &= \frac{\sum_{i \sim j} (v_i - v_j)^2}{\sum_i v_i^2}. \end{aligned}$$

Since the all-ones vector  $\mathbf{1}$  is associated to the eigenvalue  $\mu_0^{(\mathbf{L})} = 0$ , we have by the Courant-Fischer characterisation of the smallest nonzero eigenvalue  $\mu_1^{(\mathbf{L})}$ :

$$\mu_1^{(\mathbf{L})} = \min_{\mathbf{v} \perp \mathbf{1}} \frac{\mathbf{v}^T \mathbf{L} \mathbf{v}}{\mathbf{v}^T \mathbf{v}} = \min_{\sum_i v_i = 0} \frac{\sum_{i \sim j} (v_i - v_j)^2}{\sum_i v_i^2}, \quad (12)$$

where the minimum is taken over all vectors  $\mathbf{v} \neq 0$  satisfying the given condition. Since we can without loss of generality also constrain the minimisation to, say, the vectors of norm  $\|\mathbf{v}\|^2 = n$ , we see that the task of finding a Fiedler vector of  $G$  is in fact a fractional relaxation of the combinatorial problem of determining an optimal bisection of  $G$ .

This correspondence motivates the previously indicated spectral approach to bisectioning a connected graph  $G$  [24, 29]:

1. Compute Fiedler vector  $\mathbf{v} \in \mathbb{R}^n$  of  $G$ .
2. Determine cut  $(S, \bar{S})$  by rule:

$$\begin{cases} v_i > \theta & \Rightarrow i \in S, \\ v_i < \theta & \Rightarrow i \in \bar{S}, \end{cases} \quad (13)$$

where  $\theta$  is the median value of the  $v_i$ 's.

The use of *normalised* Fiedler vectors to graph bipartitioning was explored in [30], where it was shown that Fiedler vectors of  $\mathcal{L}$  yield fractionally optimal graph bipartitions according to the *normalised cut capacity* measure:

$$\hat{C}(S, \bar{S}) = \frac{C(S, \bar{S})}{\text{Vol}(S)} + \frac{C(S, \bar{S})}{\text{Vol}(\bar{S})}, \quad (14)$$

where  $\text{Vol}(S) = \sum_{i \in S} d_i$ .

Since  $\mathbf{u}$  is an eigenvector of  $\mathcal{L}$  with eigenvalue  $\lambda$  if and only if  $\mathbf{v} = \mathbf{D}^{-\frac{1}{2}}\mathbf{u}$  is eigenvector of  $\mathbf{D}^{-1}\mathbf{L}$  with eigenvalue  $\lambda$ , the eigenvalue  $\mu_1^{(\mathcal{L})}$  can be characterised in terms of a “degree-adjusted” Rayleigh quotient:

$$\mu_1^{(\mathcal{L})} = \min_{\mathbf{u} \perp \mathbf{D}^{\frac{1}{2}}\mathbf{1}} \frac{\mathbf{u}^T \mathcal{L} \mathbf{u}}{\mathbf{u}^T \mathbf{u}} = \min_{\mathbf{v} \perp \mathbf{D}\mathbf{1}} \frac{\sum_{i \sim j} (v_i - v_j)^2}{\sum_i d_i v_i^2}. \quad (15)$$

Since  $\mathbf{u}$  is an eigenvector of  $\mathcal{L}$  with eigenvalue  $\lambda$  if and only if  $\mathbf{v} = \mathbf{D}^{-\frac{1}{2}}\mathbf{u}$  is eigenvector of  $\mathbf{D}^{-1}\mathbf{L}$  with eigenvalue  $\lambda$ , the eigenvalue  $\mu_1^{(\mathcal{L})}$  can be characterised in terms of a “degree-adjusted” Rayleigh quotient:

$$\mu_1^{(\mathcal{L})} = \min_{\mathbf{u} \perp \mathbf{D}^{\frac{1}{2}}\mathbf{1}} \frac{\mathbf{u}^T \mathcal{L} \mathbf{u}}{\mathbf{u}^T \mathbf{u}} = \min_{\mathbf{v} \perp \mathbf{D}\mathbf{1}} \frac{\sum_{i \sim j} (v_i - v_j)^2}{\sum_i d_i v_i^2}. \quad (16)$$

A natural extension of the spectral clustering idea to the local clustering context is to consider the Laplacian  $\mathbf{L}$  or  $\mathcal{L}$  together with the *Dirichlet boundary condition* that only clustering vectors  $\mathbf{v}$  with the seed vertex  $v_s$  fixed to some particular value are acceptable solutions.

We follow [17, 31] in using the normalised Laplacian  $\mathcal{L}$  and choosing  $v_s = 0$ , or equivalently  $u_s = (\mathbf{D}^{\frac{1}{2}}\mathbf{v})_s = 0$  as the boundary condition. We thus aim to cluster according to the “Dirichlet-Fiedler vector” minimising the constrained Rayleigh quotient:

$$\min_{\mathbf{u}: u_s=0} \frac{\mathbf{u}^T \mathcal{L} \mathbf{u}}{\mathbf{u}^T \mathbf{u}} = \min_{\mathbf{v}: v_s=0} \frac{\sum_{i \sim j} (v_i - v_j)^2}{\sum_i d_i v_i^2}. \quad (17)$$

For notational simplicity, assume again that  $s = n$ , and observe that for every vector  $\mathbf{u} = (u_1, \dots, u_{n-1}, 0)$ , the value of the Rayleigh quotient in equation (17) is the same as the value of the  $(n-1)$ -dimensional quotient with respect to vector  $\mathbf{u}' = (u_1, \dots, u_{n-1})$  and Laplacian  $\mathcal{L}'$  which equals  $\mathcal{L}$  with its  $n^{\text{th}}$  row and column removed. Thus, our clustering vector  $\mathbf{v}$  is, except for the final zero, the one minimising:

$$\min_{\mathbf{u}'} \frac{(\mathbf{u}')^T \mathcal{L}' \mathbf{u}'}{(\mathbf{u}')^T \mathbf{u}'} = \min_{\mathbf{v}'} \frac{\sum_{i \sim j} (v'_i - v'_j)^2}{\sum_i d_i (v'_i)^2}, \quad (18)$$

i.e.  $\mathbf{v}' = \mathbf{D}^{-\frac{1}{2}}\mathbf{u}'$  for the principal eigenvector  $\mathbf{u}'$  of the Laplacian  $\mathcal{L}'$ . Let us denote  $\mathbf{v} = \mathbf{v}^f$  and call this the *local Fiedler vector* associated to graph  $G$  and seed vertex  $s = n$ .

### 3 Local Fiedler vectors and absorption times of random walks

We shall now show that the components of the local Fiedler vector  $\mathbf{v}^f = (v_1, \dots, v_{n-1})$  are in fact approximately proportional to the absorption times  $m_i$  discussed in Section 1.5. The connection between the absorption time provides a natural interpretation to the notion of the local Fiedler vector, and yields further support to the idea of local clustering by constrained spectral techniques. Previously random walks and spectral clustering have been jointly addressed by Meila and Shi [32] and local clustering by PageRank by Andersen, Chung, and Lang [18]. Important papers linking structural properties of graphs to convergence rates of random walks via spectral techniques are [33, 34].

Observe first, from equation (2), that:

$$\mathcal{L}' = \mathbf{I} - \mathbf{D}^{\frac{1}{2}} \mathbf{Q} \mathbf{D}^{-\frac{1}{2}} = \mathbf{I} - \mathcal{Q}, \quad (19)$$

where  $\mathbf{Q}$  is as in Equation (5) and  $\mathbf{D} = \text{diag}(d_1, \dots, d_{n-1})$ .

Since  $\mathcal{Q}$  is similar to  $\mathbf{Q}$ , its spectrum satisfies:

$$\text{Spec}(\mathcal{Q}) = \text{Spec}(\mathbf{Q}) = \text{Spec}(\hat{\mathbf{P}}) \setminus \{1\}. \quad (20)$$

Thus,  $\mu \neq 0$  is an eigenvalue of  $\mathcal{L}'$  if and only if  $\lambda = 1 - \mu \neq 1$  is an eigenvalue of both  $\mathcal{Q}$  and  $\mathbf{Q}$ . Moreover, if  $\mathbf{u}$  is an eigenvector associated to eigenvalue  $\lambda$  in  $\mathcal{Q}$ , then  $\mathbf{v} = \mathbf{D}^{-\frac{1}{2}} \mathbf{u}$  is an eigenvector associated to the same eigenvalue in  $\mathbf{Q}$ .

Let then the eigenvalues of  $\mathcal{Q}$  (or equivalently  $\mathbf{Q}$ ) be  $1 > \lambda_1 \geq \dots \geq \lambda_{n-1} \geq 0$ . Since  $\mathcal{Q}$  is symmetric, it has a corresponding orthonormal system of eigenvectors  $\mathbf{u}_1, \dots, \mathbf{u}_{n-1}$  and a representation:

$$\mathcal{Q} = \sum_{i=1}^{n-1} \lambda_i \mathbf{u}_i \mathbf{u}_i^T. \quad (21)$$

Denoting the component matrices  $\mathbf{U}_i = \mathbf{u}_i \mathbf{u}_i^T$ , we observe that by orthogonality of the eigenvectors we have  $\mathbf{U}_i \mathbf{U}_j = 0$  for  $i \neq j$ , and by normality  $\mathbf{U}_i^2 = \mathbf{U}_i$ . From these two observations it follows that:

$$\mathcal{Q}^t = \sum_{i=1}^{n-1} \lambda_i^t \mathbf{U}_i, \quad \text{for } t = 0, 1, \dots \quad (22)$$

Since  $\mathbf{Q} = \mathbf{D}^{-\frac{1}{2}} \mathcal{Q} \mathbf{D}^{\frac{1}{2}}$ , we obtain from this for  $\mathbf{Q}^t$  the representation:

$$\mathbf{Q}^t = \mathbf{D}^{-\frac{1}{2}} \mathcal{Q}^t \mathbf{D}^{\frac{1}{2}} = \sum_{i=1}^{n-1} \lambda_i^t (\mathbf{D}^{-\frac{1}{2}} \mathbf{u}_i) (\mathbf{u}_i^T \mathbf{D}^{\frac{1}{2}}) = \sum_{i=1}^{n-1} \lambda_i^t \mathbf{v}_i \mathbf{v}_i^T \mathbf{D}, \quad (23)$$

where  $\mathbf{v}_i = \mathbf{D}^{-\frac{1}{2}} \mathbf{u}_i$  is an eigenvector associated to eigenvalue  $\lambda_i$  in  $\mathbf{Q}$ .

Substituting this to Equation (7) and denoting the  $(n - 1)$ -dimensional all-ones vector by  $\mathbf{1}$ , we thus obtain an expression for the vector  $\mathbf{m}$  of absorption times  $m_i$  in terms of the eigenvalues and eigenvectors of  $\mathbf{Q}$ , or equivalently  $\mathcal{Q}$ :

$$\begin{aligned}
\mathbf{m} &= \sum_{t=0}^{\infty} \mathbf{Q}^t \mathbf{1} \\
&= \sum_{t=0}^{\infty} \sum_{i=1}^{n-1} \lambda_i^t \mathbf{v}_i \mathbf{v}_i^T \mathbf{D} \mathbf{1} \\
&= \sum_{t=0}^{\infty} \left( \sum_{i=1}^{n-1} \lambda_i^t \mathbf{v}_i \mathbf{v}_i^T \right) \mathbf{d},
\end{aligned} \tag{24}$$

where  $\mathbf{d} = (d_1, \dots, d_{n-1})^T$ .

Now if the principal eigenvalue  $\lambda_1$  is well-separated from the others, i.e. if the ratio  $|\lambda_i/\lambda_1|$  is small for  $i > 1$ , this yields a good approximation for  $\mathbf{m}$ :

$$\begin{aligned}
\mathbf{m} &= \mathbf{1} + \sum_{t=1}^{\infty} \lambda_1^t \left( \mathbf{v}_1 \mathbf{v}_1^T \mathbf{d} + \underbrace{\sum_{i=2}^{n-1} \left( \frac{\lambda_i}{\lambda_1} \right)^t \mathbf{v}_i \mathbf{v}_i^T \mathbf{d}}_{\text{small-norm "noise"}} \right) \\
&\approx \mathbf{1} + \sum_{t=1}^{\infty} \lambda_1^t \mathbf{v}_1 \mathbf{v}_1^T \mathbf{d} \\
&= \mathbf{1} + \frac{\lambda_1}{1 - \lambda_1} \mathbf{v}_1 \mathbf{v}_1^T \mathbf{d}.
\end{aligned} \tag{25}$$

Even in cases where there is no evident gap in the spectrum and hence near-equality cannot be assumed, we have found in our experiments that the approximations obtained are near-perfectly correlated with the exact absorption times for a variety of graphs.

We study three example graphs to point out the strengths and weaknesses of the proposed approximation. The first example graph is the clustered but highly symmetric caveman graph of Figure 1, where the symmetries present cause problems for the proposed approximation. Our second example is the karate club network shown in Figure 2. The third example graph is a uniform random graph  $\mathcal{G}_{n,p}$ , with  $n = 100$  and  $p = 0.1$  [35], which by definition has no clear cluster structure, and hence the absorption times cannot be expected to have interesting patterns.

In Figure 6, we show comparisons of some approximate and exact spectral computations for three example graphs. In each case, the highest-numbered vertex of the graph has been chosen as the unique seed vertex. It can be noted, from the top row of plots in Figure 6, that the spectra of the graphs'  $\hat{\mathbf{P}}$  matrices do not exhibit large gaps between their second and third largest eigenvalues. Thus, it can not be expected *a priori* that the Fiedler-vector based approximations to the absorption times, from Equation (25), would be even of the same magnitude as the exact ones, as calculated from Equations (6) and (7). (Observe

also how the structure of the caveman graph is reflected in the corresponding  $\hat{\mathbf{P}}$  spectrum: a notable eigenvalue gap occurs after the six largest eigenvalues, each representing the dominant convergence behaviour of one of the clusters.)

Correlations between the approximate and exact absorption times are apparent in the quantile-quantile plots presented in the second row of Figure 6: here the values group diagonally when a linear dependency exists. The correlation is very high in all cases: 0.99863 for the caveman graph, 0.99636 for the karate club network, and 0.99999 for the uniform random graph.

The two lowest rows in Figure 6 present the actual values of the exact and approximate absorption-time vectors, indexed by vertex number. These plots illustrate the usefulness of these quantities for performing a two-classification of the vertices into the local cluster of the seed vertex (low values) versus the other vertices (high values). In fact, for the caveman graph, the full six-cluster structure is visible. In the karate club network it can be seen that two groups are present: one with high values and another one with low values. (Cf. Figure 8, which indicates the “ground truth” clustering of the vertices in this graph.) As expected, the uniform random graph reveals no significant cluster structure, but the vertices near the seed vertex can be identified by their lower values, whereas most of the graph has another, higher value.

In practice, it is not always interesting to compute the absorption times for all vertices, especially in local computation, in which case we may only have approximated some of the components of the Fiedler vector. For these situations, we may write the  $k^{\text{th}}$  component of the result vector as

$$\begin{aligned}
(\mathbf{Q}^t \mathbf{1})_k &= \left( \sum_{i=1}^{n-1} \lambda_i^t \mathbf{v}_i \mathbf{v}_i^T \mathbf{D} \mathbf{1} \right)_k \\
&= \left( \sum_{i=1}^{n-1} \lambda_i^t (\mathbf{v}_i^T \mathbf{d}) \mathbf{v}_i \right)_k \\
&= \sum_{i=1}^{n-1} \lambda_i^t (\mathbf{v}_i)_k \underbrace{\sum_{\ell=1}^{n-1} (\mathbf{v}_i)_\ell (\mathbf{d})_\ell}_{c_i}.
\end{aligned} \tag{26}$$

From this we obtain for the absorption time from vertex  $k$  to vertex  $s$  the expression

$$\begin{aligned}
m_k &= \sum_{t=0}^{\infty} (\mathbf{Q}^t \mathbf{1})_k \\
&= 1 + \sum_{t=1}^{\infty} \lambda_1^t \left( c_1 \cdot (\mathbf{v}_1)_k + \sum_{i=2}^{n-1} \left( \frac{\lambda_i}{\lambda_1} \right)^t c_i \cdot (\mathbf{v}_i)_k \right) \\
&\approx 1 + \sum_{t=1}^{\infty} \lambda_1 \cdot c_1 \cdot (\mathbf{v}_1)_k \\
&= 1 + \underbrace{\frac{\lambda_1}{1 - \lambda_1}}_{c'} \cdot c_1 \cdot (\mathbf{v}_1)_k.
\end{aligned} \tag{27}$$

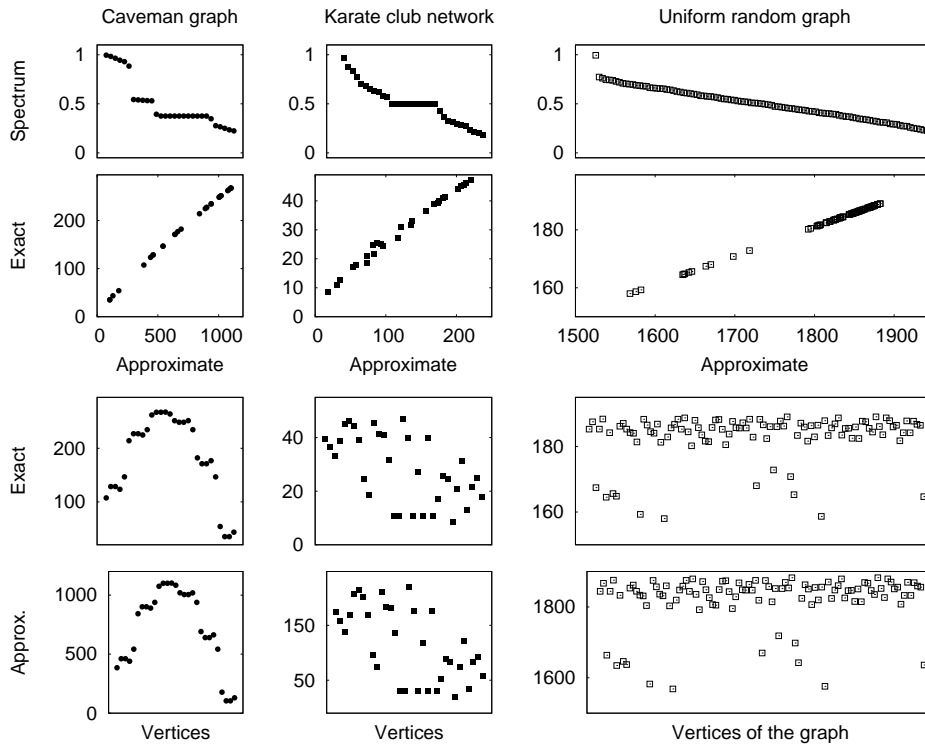


Figure 6: Comparisons of approximate and exact spectral computations for three example graphs: the small graphs of Figures 1 and 2, and a uniform random graph  $\mathcal{G}_{n,p}$ , using a random vertex as the seed vertex. The top row presents the sorted spectra of the  $\hat{\mathbf{P}}$  matrices of the graphs, the second row plots the approximate and exact absorption-time values for the given seed vertex against each other, and the lowest two rows indicate the exact and approximate absorption-time values as ordered by vertex number. The bottom rows can be seen as illustrating the quantities' capability of distinguishing the cluster of the seed vertex (low values) from the other vertices (high values).

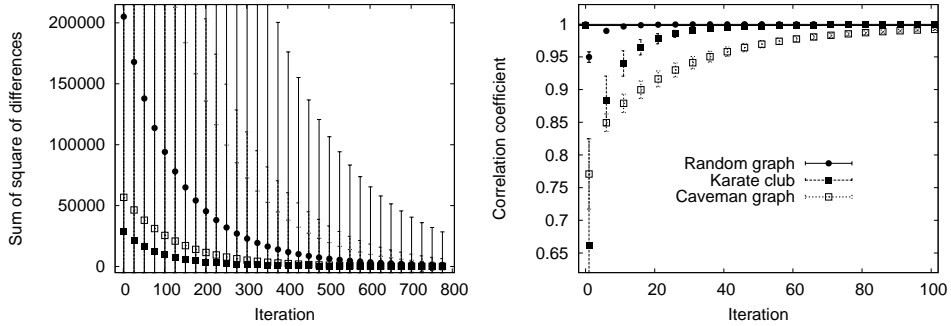


Figure 7: The sum of squares of the difference from the exact absorption-time (on the left) of estimate vectors with different cutoff values for approximating through Equation (24) and Pearson correlation between the exact and the estimate vectors (on the right) for the three example graphs: the small graphs of Figures 1 and 2, and the  $\mathcal{G}_{n,p}$ . The values shown are averaged over the vertex sets of the two small graphs and over a set of 30 vertices selected uniformly at random for the  $\mathcal{G}_{n,p}$  graph. The smallest standard deviation corresponds to the caveman graph and the largest to the uniform random graph. The horizontal lines (all three overlap between 0.980 and 0.997) correspond to the average correlation coefficients between the exact and the approximate absorption times of Equation (25).

Now for a given graph  $G$ ,  $c'$  is a constant and so we obtain the very simple approximate correspondence  $\mathbf{m} \approx \mathbf{1} + c'\mathbf{v}^f$  between the absorption time vector  $\mathbf{m}$  and the local Fiedler vector  $\mathbf{v}^f = \mathbf{v}_1$ .

In order to compare the quality of the approximation as well as to illustrate the computational load in approximating by summing term by term the series of Equation (24), we calculated for each cutoff length the sum of squares of the differences between the partial sums and the exact absorption times, divided by the order of each of the three example graphs: the graph of Figure 1, the Zachary karate club graph of Figure 2, and the uniform random graph  $\mathcal{G}_{n,p}$ . The resulting values over the set of vertices are shown in Figure 7 (on the left) together with the Pearson correlations (on the right) achieved at each iteration. In both plots, mean and standard deviation are shown.

## 4 Local approximation of Fiedler vectors

We take as a starting point the Rayleigh quotient of Equation (18). Since we are free to normalise our eventual Fiedler vector  $\mathbf{v}^f$  to any length we wish, we can constrain the minimisation to vectors  $\mathbf{v}$  that satisfy, say,  $\|\mathbf{v}\|_2^2 = n = |V|$ .



Thus, the task becomes one of finding a vector  $\mathbf{v}$  that satisfies for a given  $s \in V$ :

$$\mathbf{v}^f = \operatorname{argmin} \left\{ \sum_{j \sim k} (v_j - v_k)^2 : v_s = 0, \|\mathbf{v}\|_2^2 = n \right\}. \quad (28)$$

We can solve this task approximately by reformulating the requirement that  $\|\mathbf{v}\|_2^2 = n$  as a ‘‘soft constraint’’ with weight  $c > 0$ , and minimising the objective function

$$f(\mathbf{v}) = \frac{1}{2} \sum_{j \sim k} (v_j - v_k)^2 + \frac{c}{2} \cdot \left( n - \sum_j v_j^2 \right) \quad (29)$$

by gradient descent. Since the partial derivatives of  $f$  have the simple form

$$\frac{\partial f}{\partial v_j} = - \sum_{k \sim j} v_k + (d_j - c) \cdot v_j, \quad (30)$$

the descent step can be computed locally at each vertex at time  $t + 1$ , based on information about the values of the vector  $\mathbf{v}$  at time  $t$ , denoted by  $\tilde{\mathbf{v}}(t)$ , for the vertex itself and its neighbours:

$$\tilde{v}_j(t+1) = \tilde{v}_j(t) + \delta \cdot \left( \sum_{k \sim j} \tilde{v}_k - (d_j - c) \cdot \tilde{v}_j \right), \quad (31)$$

where  $\delta > 0$  is a parameter determining the speed of the descent.

Assuming that the natural cluster of vertex  $s$  is small compared to the order of the graph  $n$ , the normalisation  $\|\mathbf{v}\|_2^2 = n$  entails that most vertices  $j$  in the network will have  $v_j \approx 1$ . Thus the descent iterations (31) can be started from an initial vector  $\tilde{\mathbf{v}}(0)$  that has  $\tilde{v}_s(0) = 0$  for the seed vertex  $s \in V$  and  $\tilde{v}_k(0) = 1$  for all  $k \neq i$ . The estimates need then to be updated at time  $t > 0$  only for those vertices  $j$  that have at least one neighbour  $k$  such that  $\tilde{v}_k(t-1) < 1$ .

Balancing the constraint weight  $c$  against the speed of gradient descent  $\delta$  naturally requires some care. We have obtained reasonably stable results with the following heuristic: given an estimate  $\bar{k}$  for the average degree of the vertices in the network, set  $c = 1/\bar{k}$  and  $\delta = c/10$ . The gradient iterations (31) are then continued until all the changes in the  $v$ -estimates are below  $\varepsilon = \delta/10$ . We leave the calibration of these parameters to future work.

The (approximate) Fiedler values thus obtained represent proximity-values of the vertices in  $V$  to the cluster of vertex  $s$ . Determining a bisection into  $S$  and  $V \setminus S$  is now a one-dimensional two-classification task that can in principle be solved using any of the standard pattern classifiers, such as variations of the basic  $k$ -means algorithm [36].

We illustrate the applicability approximate absorption times for clustering the karate club network (Figure 2). The approximate absorption times shown in Figure 8 are computed directly with Equation (27): the group structure is seen to be strong when the seed vertex is one of the central members of the group,

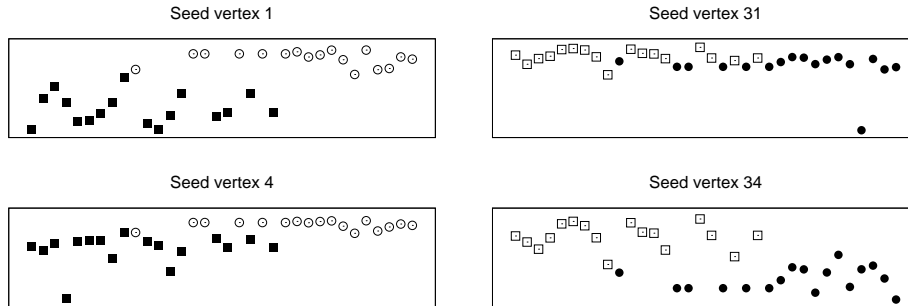


Figure 8: Four examples of two-classifying vertices of the Zachary karate club graph. The examples on the left have the seed vertex among the “rectangles” of Figure 2 and the examples of the right have the seed vertex among the “circles”. The vertices are ordered by their label in Figure 2 and a zero has been inserted to represent the absorption time to the seed vertex itself. The group in which the seed belongs is drawn in black and the other group in white.

whereas the classification task is harder for the “border” vertices, as can be expected. For more extensive examples of clustering with the locally computed approximates, we refer the reader to previous work [11].

## 5 Conclusions and further work

In this work we have derived an expression for the absorption times to a single absorbing vertex  $s$  in a simple random walk in an undirected, unweighted graph in terms of the spectrum of the normalised Laplacian matrix of the graph. We have shown that by only knowing the Fiedler vector corresponding to  $s$  on the boundary and the corresponding eigenvalue provides an approximation of the absorption times if the spectrum of the graph presents a gap after the first eigenvalue. Experimentally we have confirmed that the values given by the approximation are nearly perfectly correlated with the exact absorption times even in the absence of such a gap.

Our motivation is to use the absorption times into a seed vertex  $s$  as a measure of proximity in two-classifying the graph into two partitions: vertices that are “relevant” to the seed vertex and other vertices. Hence, not knowing the exact values but rather another vector of perfectly correlated values is sufficient for separating between the vertices with higher values from those with lower values (which is the classical two-classification task).

Such a two-partition of a graph is known as local clustering. In order for the proposed values to be locally computable, we have also presented a gradient-descent method to approximate the Fiedler vector using only local information in the graph. The method iteratively processes the neighbourhoods of vertices starting from the seed vertex and expanding outwards within the group of po-

tentially “relevant” vertices, without any need to process other parts of the graph. We have illustrated the potential of these vectors in two-classification for local clustering on a classical example graph representing a social network.

In further work, we seek to study further the effects of the presence or absence of a spectral gap in the input graph into the approximation proposed. We also want to calibrate the parameters of the locally computable approximation in such a way that no a priori knowledge of the input graph would be needed, but that the method would rather adapt to the structure of the graph at runtime by dynamic parameter adjustment. Of additional interest are extensions of this work to weighted and directed graphs as well as case studies of applications of local clustering. We also contemplate possible uses for approximate absorption times in resolving other problems of interest that involve complex systems represented as graphs.

## Acknowledgements

The work of Orponen and Schaeffer was supported by the Academy of Finland under grant 206235 (ANNE, 2004–2006). Schaeffer and Avalos received support from the UANL under grant CA1475-07 and from PROMEP under grant 103,5/07/2523. Avalos also thanks CONACYT for support.

A preliminary report on parts of this work was presented as “Local clustering of large graphs by approximate Fiedler vectors” by P. Orponen and S. E. Schaeffer, at the Fourth International Workshop on Efficient and Experimental Algorithms in Santorini, Greece, May 2005. The current work was presented at The Fifteenth Conference of the International Linear Algebra Society (ILAS) in Cancún, Quintana Roo, Mexico, in June 2008.

## References

- [1] Duncan J. Watts and Steven H. Strogatz. Collective dynamics of ‘small world’ networks. *Nature*, 393(6684):440–442, June 1998.
- [2] Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *Science*, 286:509–512, October 1999.
- [3] Michalis Faloutsos, Petros Faloutsos, and Christos Faloutsos. On power-law relationships of the Internet topology. In *Proceedings of the ACM SIGCOMM’99 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communication*, pages 251–262, New York, NY, USA, 1999. ACM Press.
- [4] Sergey N. Dorogovtsev and José Ferreira F. Mendes. *Evolution of Networks: From Biological Nets to the Internet and WWW*. Oxford University Press, Oxford, UK, January 2003.

- [5] Mark E.J. Newman. The structure and function of complex networks. *SIAM Review*, 45(2):167–256, 2003.
- [6] Satu Elisa Virtanen. Properties of nonuniform random graph models. Research Report A77, Helsinki University of Technology, Laboratory for Theoretical Computer Science, Espoo, Finland, May 2003.
- [7] Mark E. J. Newman and Michelle Girvan. Mixing patterns and community structure in networks. In Romualdo Pastor-Satorras, Miguel Rubi, and Albert Diaz-Guilera, editors, *Statistical Mechanics of Complex Networks*, volume 625 of *Lecture Notes in Physics*, pages 66–87, Berlin, Germany, 2003. Springer-Verlag GmbH.
- [8] Jon M. Kleinberg and Steve Lawrence. The structure of the web. *Science*, 294(5548):1849–1850, November 2001.
- [9] Wayne W. Zachary. An information flow model for conflict and fission in small groups. *Journal of Anthropological Research*, 33:452–473, 1977.
- [10] Fang Wu and Bernardo A. Huberman. Finding communities in linear time: a physics approach. *The European Physical Journal B*, 38(2):331–338, 2004.
- [11] Pekka Orponen and Satu Elisa Schaeffer. Local clustering of large graphs by approximate Fiedler vectors. In Sotiris Nikolettseas, editor, *Proceedings of the Fourth International Workshop on Efficient and Experimental Algorithms (WEA '05)*, volume 3505 of *Lecture Notes in Computer Science*, pages 524–533, Berlin/Heidelberg, Germany, 2005. Springer-Verlag GmbH.
- [12] Satu Elisa Schaeffer. Graph clustering. *Computer Science Review*, 1(1):27–64, 2007.
- [13] Duncan J. Watts. *Small Worlds*. Princeton University Press, Princeton, NJ, USA, 1999.
- [14] Jiří Šíma and Satu Elisa Schaeffer. On the NP-completeness of some graph cluster measures. In Jiří Wiedermann, Gerard Tel, Jaroslav Pokorný, Mária Bieliková, and Július Štuller, editors, *Proceedings of the Thirty-second International Conference on Current Trends in Theory and Practice of Computer Science (Sofsem 06)*, volume 3831 of *Lecture Notes in Computer Science*, pages 530–537, Berlin/Heidelberg, Germany, 2006. Springer-Verlag GmbH.
- [15] Satu Elisa Schaeffer. Stochastic local clustering for massive graphs. In T. B. Ho, D. Cheung, and H. Liu, editors, *Proceedings of the Ninth Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD-05)*, volume 3518 of *Lecture Notes in Computer Science*, pages 354–360, Berlin/Heidelberg, Germany, 2005. Springer-Verlag GmbH.
- [16] Mark E.J. Newman and Michelle Girvan. Finding and evaluating community structure in networks. *Physical Review E*, 69:026113, 2004.

- [17] Fan R.K. Chung. *Spectral Graph Theory*. American Mathematical Society, Providence, RI, USA, 1997.
- [18] Reid Andersen, Fan Chung, and Kevin Lang. Local partitioning for directed graphs using PageRank. In *Proceedings of WAW 2007*, pages 166–178, 2007.
- [19] Reid Andersen and Fan Chung. Detecting sharp drops in PageRank and a simplified local partitioning algorithm. In *Proceedings of the Conference on the Theory and Applications of Models of Computation (TAMC)*, pages 1–12, 2007.
- [20] Fan Chung. The heat kernel as the pagerank of a graph. *PNAS*, 105(50):19735–19740, 2007.
- [21] Norman Biggs. *Algebraic Graph Theory*. Cambridge University Press, Cambridge, UK, second edition, 1994.
- [22] Desmond J. Higham, Gabriela Kalna, and Milla Kibble. Spectral clustering and its use in bioinformatics. *Journal of Computational and Applied Mathematics*, 204(1):25–37, July 2007.
- [23] Ravi Kannan, Santosh Vempala, and Adrian Vetta. On clusterings — good, bad and spectral. *Journal of the ACM*, 51(3):497–515, 2004.
- [24] Miroslav Fiedler. Algebraic connectivity of graphs. *Czechoslovak Mathematical Journal*, 23:298–305, 1973.
- [25] Miroslav Fiedler. A property of eigenvectors of nonnegative symmetric matrices and its application to graph theory. *Czechoslovak Mathematical Journal*, 25:619–633, 1975.
- [26] Daniel A. Spielman and Shang-Hua Teng. Spectral partitioning works: planar graphs and finite element meshes. In *Proceedings of the Thirty-seventh IEEE Symposium on Foundations of Computing (FOCS)*, pages 96–105, Los Alamitos, CA, USA, 1996. IEEE Computer Society Press.
- [27] Huaijun Qiu and Edwin R. Hancock. Graph matching and clustering using spectral partitions. *Pattern Recognition*, 39(1):22–34, January 2006.
- [28] Michael Holzrichter and Suely Oliveira. A graph based method for generating the fiedler vector of irregular problems. In *Proceedings of the 11 IPPS/SPDP'99 Workshops Held in Conjunction with the 13th International Parallel Processing Symposium and 10th Symposium on Parallel and Distributed Processing*, volume 1586 of *Lecture Notes In Computer Science*, pages 978–985, London, UK, 1999. Springer-Verlag.
- [29] Wilm E. Donath and A. J. Hoffman. Lower bounds for the partitioning of graphs. *IBM Journal of Research and Development*, 17(5):420–425, September 1973.

- [30] Jianbo Shi and Jitendra Malik. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):888–905, August 2000.
- [31] Fan R.K. Chung and Robert B. Ellis. A chip-firing game and Dirichlet eigenvalues. *Discrete Mathematics*, 257:341–355, 2002.
- [32] Marina Meila and Jianbo Shi. A random walks view of spectral segmentation. In *Proceedings of the Eighth International Conference on Artificial Intelligence and Statistics (AISTATS 2001)*, 2001.
- [33] Noga Alon. Eigenvalues and expanders. *Combinatorica*, 6(2):83–96, 1986.
- [34] Alistair J. Sinclair and Mark R. Jerrum. Approximative counting, uniform generation and rapidly mixing markov chains. *Information and Computation*, 82(1):93–133, July 1989.
- [35] E. N. Gilbert. Random graphs. *Annals of Mathematical Statistics*, 30(4):1141–1144, December 1959.
- [36] John A. Hartigan and Manchek A. Wong. Algorithm AS 136: A  $k$ -means clustering algorithm. *Applied Statistics*, 28:100–108, 1979.