Skeletal Graphs from Schrödinger Magnitude and Phase

Francisco Escolano[1(✉)], Edwin R. Hancock[2], and Miguel A. Lozano[3]

[1] Department of Computer Science and AI, University of Alicante, Alicante, Spain
sco@dccia.ua.es
[2] Department of Computer Science, University of York, York, UK
erh@cs.york.ac.uk
[3] Department of Computer Science and AI, University of Alicante, Alicante, Spain
malozano@ua.es

Abstract. Given an adjacency matrix, the problem of finding a simplified version of its associated graph is a challenging one. In this regard, it is desirable to retain the essential connectivity of the original graph in the new representation. To this end, we exploit the magnitude and phase information contained in the Schrödinger Operator of the Laplacian. Recent findings based on continuous-time quantum walks suggest that the long-time averages of both magnitude and phase are sensitive to long-range interactions. In this paper, we depart from this hypothesis to propose a novel representation: skeletal graphs. Using the degree of interaction (or "long-rangedness") as a criterion we propose a structural level set (i.e. a sequence of graphs) from which it emerges the simplified graph. In addition, since the same theory can be applied to weighted graphs, we can analyze the implications of the new representation in the problems of spectral clustering, hashing and ranking from computer vision. In our experiments we will show how coherent transport phenomenon implemented by quantum walks discovers the long-range interactions without breaking the structure of the manifolds on which the graph or its connected components are embedded.

Keywords: Graph simplification · Schrödinger Operator · Quantum walks

1 Introduction

Given a set of points $\mathcal{X} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\} \subset \mathbb{R}^d$ to be clustered, ranked or hashed, and a metric $d : \mathcal{X} \times \mathcal{X} \to \mathbf{R}$, the spectral approach consists of mapping the $\boldsymbol{x}_i$ to the vertices V, with $N = |V|$, of an undirected weighted graph $G(V, E)$ so that the edge weight $W(i, j) = e^{-d^2(\boldsymbol{x}_i, \boldsymbol{x}_j)/t}$ is a similarity measure between $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$.

The analysis of the similarity/affinity matrix $\boldsymbol{W}$ is key to the success of the pattern recognition task. This is usually done through the study of the Laplacian matrix $\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{W}$ or of its normalized counterpart. Then, despite $\boldsymbol{W}$ contains

C.-L. Liu et al. (Eds.): GbRPR 2015, LNCS 9069, pp. 335–344, 2015.
DOI: 10.1007/978-3-319-18224-7_33

pairwise relations, the spectrum and eigenvectors of $\boldsymbol{L}$ contain global information. However, it has been recently pointed out [1] that the limit analysis of a graph Laplacian, for instance setting $t \to 0$ as a consequence of $N \to \infty$, reveals some flaws or degenerate behaviors. For instance, ranking functions, which are usually implemented by Green's functions, diverge when $N \to \infty$. Zhout et al. propose to solve this problem by computing Green's functions of "higher-order" Laplacians, i.e. $\boldsymbol{L}^m$ and $m \geq 0$.

In clustering, the well known approach of normalized cuts method [2] has been improved by introducing topological distances which are consistent with a metric (commute-times) [3]. State-of-the-art methods for image segmentation do not only combine different eigenvectors of the Laplacian (normalized in this case) [4] but also incorporate better dissimilarity measures to make the weight matrix $\boldsymbol{W}$ more discriminative.

Finally, spectral hashing [5][6] consists of fusing the information contained in several eigenvectors of $\boldsymbol{L}$ in order to define uncorrelated and efficient codes. Anchor graphs [7] provide a scalable approach for spectral hashing through the simplification of the original similarity matrix $\boldsymbol{W}$.

Most of the exisiting methods for improving ranking, semi-supervised learning, clustering or hashing rely on the concept of a random walk. For instance, in [8] pixels are labeled in terms of the probability that a random walk will reach them from a given seed. In this regard, the underpinning principle is to minimize the combinatorial Dirichlet integral associated with the weighted Laplacian of a graph. Seeds are assumed to be the boundary conditions of a Dirichlet problem and minimization seeks the values of the unknown labels so that the Dirichlet integral is the smoothest one. This means that a harmonic solution is preferred (the probability of an unknown label for a pixel must be average of the probabilities of neighboring pixels). The method relies on solving a linear system. The resulting method is more robust to noise than normalized cuts.

Recent findings from the study of quantum walks suggest alternative ways of incorporating "high-order" information. For instance, in [9] a quantum version of the Jensen-Shannon divergence is used to compute a graph kernel, whereas in [10] it is exploited to detect both symmetric and anti-symmetric structures in graphs. In [11] and the references therein, it is suggested that quantum walks provide information about long-range interactions. For instance, in dendrimers (trees structured in strata or "generations") a quantum random walk starting at the root reaches (in the limit) nodes lying in the same generation with similar probability.

Our long-term aim is to address the question whether the above ideas (simplification of $\boldsymbol{W}$, incorporation of long-range interactions in $\boldsymbol{L}$, evaluation of the similarity measure) which have contributed to the improvement of spectral methods for clustering, ranking and hashing can be driven from quantum walks. In this paper, we will focus on the improvement of spectral clustering and ranking.

2 Quantum Walks for Analyzing Transport

2.1 Unitary Evolution and the Schrödinger Operator

Let $\mathcal{H} = span\{|j\rangle \mid j = 1, \ldots, N\} = \mathbb{C}^N$ be a $N-$dimensional Hilbert space where $\langle j| = (0 \ldots 1 \ldots 0)$ with a 1 at the $j-th$ position. We use the Dirac braket notation where: $|a\rangle = \boldsymbol{a}$, $\langle a| = \boldsymbol{a}^*$, $\langle a|b\rangle = \boldsymbol{a}^*\boldsymbol{b}$ is the inner product and therefore $\langle j|k\rangle = \boldsymbol{j}^*\boldsymbol{k} = \delta_{jk}$ and $\sum_{j=1}^{N} |j\rangle\langle j| = \mathbf{1}$. Then, a point in the Hilbert space is given by $|\psi\rangle = \sum_{j=1}^{N} c_j|j\rangle$ with $c_j \in \mathbb{C}$ so that $|c_1|^2+|c_2|^2+\ldots+|c_N|^2 = 1$ and $|c_j|^2 = \overline{c}_j c_j$.

The Schrödinger equation describes how the complex state vector $|\psi(t)\rangle \in \mathbb{C}^n$ of a continuous-time quantum walk varies with time:

$$\frac{\partial|\psi(i)\rangle}{\partial t} = -i\boldsymbol{L}|\psi_t\rangle. \tag{1}$$

Given an initial state $|\psi(0)\rangle = \sum_{j=1}^{N} c_j^0|j\rangle$ the latter equation can be solved to give $|\psi(t)\rangle = \Psi(t)|\psi(0)\rangle$, where $\Psi(t) = e^{-i\boldsymbol{L}t}$ is a complex $n \times n$ *unitary matrix*. In this paper we refer to $\Psi(t)$ as the *Schrödinger operator*. In this regard, Stone's theorem [12] establishes a one-to-one correspondence between a time parameterized unitary matrix $\boldsymbol{U}(t)$ and a self-adjoint (Hermitian) operator $\boldsymbol{H} = \boldsymbol{H}^*$ such that there is a unique Hermitian operator satisfying $\boldsymbol{U}(t) = e^{it\boldsymbol{H}}$. Such an operator $\boldsymbol{H}$ is the *Hamiltonian*. In the case of graphs $\boldsymbol{H} = -\boldsymbol{L}$ and then we have that $\Psi(t) = e^{-it\boldsymbol{L}}$ is a unitary matrix for $t \in \mathbb{R}$. Therefore, given an initial state $|\psi(0)\rangle$, the Schrödinger Operator characterizes the evolution of a Continuous-Time Quantum Walk (CTQW). The probability that the quantum walk is at node j is given by $|\langle j|\psi\rangle|^2 = |c_j|^2$. The $|c_j|^2$ are known as the amplitudes of the wave traveling through the graph.

2.2 Long-Time Averages from Magnitude

Different choices of the initial state $|\psi(0)\rangle = \sum_{j=1}^{N} c_j^0|j\rangle$ lead to different ways of probing the graph by exploiting properties of quantum superposition and quantum interference. For instance, in [9], initial amplitudes are set to $c_j^0 = \sqrt{\frac{d_j}{\sum_{k=1}^{N} d_k}}$, in order to compute de quantum version of the Jensen-Shannon divergence. However, in [10], where the focus is on identifying whether the vertices i and j are symmetrically placed in the graph we have that either $c_j^0 = 1/\sqrt{2}$ and $c_k^0 = 1/\sqrt{2}$ (in phase) or $c_j^0 = 1/\sqrt{2}$ and $c_k^0 = -1/\sqrt{2}$ (in antiphase). Actually, the Quantum Jensen-Shannon divergence has a low value when pairs or vertices are located anti-symmetrically and a high value when they are symmetrically placed.

In this paper, we use the classical choice proposed by Farhi and Gutman for studying transport properties of quantum walks in trees [13]. In such approach, states $|j\rangle$ are associated with excitations at the nodes j. Therefore, the evolution of a CTQW commencing at node $|j\rangle$ is given by $|j(t)\rangle = e^{-it\boldsymbol{L}}|j\rangle$. In this regard, the amplitude of a transition between nodes j and k at time t is given

by $c_{jk}(t) = \langle k|j(t)\rangle = \langle k|e^{-it\boldsymbol{L}}|j\rangle$, and the quantum-mechanical probability of such a transition is $\pi_{jk}(t) = |c_{jk}(t)|^2 = \left|\langle k|e^{-it\boldsymbol{L}}|j\rangle\right|^2$.

Since the spectral decomposition of the Laplacian is $\boldsymbol{L} = \Phi\Lambda\Phi^T$, where $\Phi = [\phi_1|\phi_2|\ldots|\phi_n]$ is the $N \times N$ matrix of ordered eigenvectors according to the corresponding eigenvalues $0 = \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$, and $\Lambda = diag(\lambda_1\ \lambda_2\ \ldots\ \lambda_n)$, we have that $e^{-it\boldsymbol{L}} = \Phi e^{-it\Lambda}\Phi^T$ where where $e^{-it\Lambda} = diag(e^{-it\lambda_1}\, e^{-it\lambda_2}\ \ldots\ e^{-it\lambda_n})$. Then, the probability of a transition between j and k at time t is given by

$$\begin{aligned}\pi_{jk}(t) &= \left|\langle k|e^{-it\boldsymbol{L}}|j\rangle\right|^2 = \left|\langle k|\Phi e^{-it\Lambda}\Phi^T|j\rangle\right|^2 \\ &= \left|\langle k|\sum_{u=1}^{N} e^{-it\lambda_u}|\phi_u\rangle\langle\phi_u|j\rangle\right|^2 = \left|\sum_{u=1}^{N} e^{-it\lambda_u}\langle k|\phi_u\rangle\langle\phi_u|j\rangle\right|^2 , \end{aligned} \tag{2}$$

where $\langle k|\phi_u\rangle = \phi_u(k)$ and $\langle\phi_u|j\rangle = \phi_u(j)$ respectively. Therefore we have

$$\pi_{jk}(t) = \sum_{u=1}^{N}\sum_{v=1}^{N} e^{-it(\lambda_u-\lambda_v)} z_u(k,j) z_v(k,j) , \tag{3}$$

where $z_u(k,j) = \phi_u(k)\phi_u(j)$ and $z_v(k,j) = \phi_v(k)\phi_v(j)$ account for the correlations between the $k-$th and $j-$th components of the eigenvectors ϕ_u and ϕ_v. Then, since $e^{-it(\lambda_u-\lambda_v)} = \cos(t(\lambda_u-\lambda_v)) - i\sin(t(\lambda_u-\lambda_v))$ we have that the long-time limit of $\pi_{jk}(t)$ does not exist, whereas the corresponding long-time limit of a classical continuous-time random walk is $1/N$. However, in the quantum mechanical case it is possible to compute the long-time average:

$$\begin{aligned}\chi_{jk} &= \lim_{T\to\infty}\frac{1}{T}\int_0^T \pi_{jk}(t)dt \\ &= \lim_{T\to\infty}\frac{1}{T}\int_0^T \sum_{u=1}^{N}\sum_{v=1}^{N} e^{-it(\lambda_u-\lambda_v)} z_u(k,j) z_v(k,j) \\ &= \lim_{T\to\infty}\sum_{u=1}^{N}\sum_{v=1}^{N} z_u(k,j) z_v(k,j)\frac{1}{T}\int_0^T e^{-it(\lambda_u-\lambda_v)}dt \\ &= \sum_{u=1}^{N}\sum_{v=1}^{N}\delta_{\lambda_u,\lambda_v} z_u(k,j) z_v(k,j) , \end{aligned} \tag{4}$$

where $\delta_{\lambda_u,\lambda_v} = 1$ if $\lambda_u = \lambda_v$ and 0 otherwise. Therefore, the long-time averaged probabilities do not depend directly on the eigenvalues of the Laplacian but on their multiplicity. Mülken and Blumen have recently related the multiplicity to the transport efficiency of CTQWs [11]. More precisely, the averaged return probability $\bar{\pi}(t) = (1/N)\sum_{k=1}^{N}\pi_{k,k}(t)$ decays faster with time than that of a classical continuous-time random walk under conditions of low multiplicity. This means that in the long-time limit more probabilistic mass is allocated to nodes lying far away from the origin of the quantum walk, provided that there is low degeneracy (i.e. multiplicity).

2.3 Long-Time Averages for Phase

The consequences of the above analysis can be summarized as follows. The long-time averages of the amplitude encode transport information and this information is contained in the interactions (correlations) between the eigenvectors of the Laplacian. An asymptotic analysis of the Schrödinger Operator leads to understanding the coherent transport efficiency of the graph. However, the analysis is incomplete if we do not also consider phase in addition to amplitude (magnitude). In order to compute long-time averages of phase we must pay attention to the complex nature of the operator. More precisely, the phase of the transition is $\kappa_{jk}(t) = \arg(c_{jk}(t)) = \arg(\langle k|e^{-it\boldsymbol{L}}|j\rangle)$. From $e^{-it\boldsymbol{L}} = \boldsymbol{\Phi} e^{-it\Lambda}\boldsymbol{\Phi}^T$ we have that $c_{jk}(t) = \sum_{u=1}^{N} e^{-it\lambda_u}\langle k|\phi_u\rangle\langle\phi_u|j\rangle = \sum_{u=1}^{N} e^{-it\lambda_u} z_u(j,k)$. Therefore a transition $c_{jk}(t)$ is a sum of time-dependent phasors. The magnitude of each phasor is the correlation between the $j-$th and $k-$th components of the corresponding eigenvector ϕ_u, and the phase relies on the corresponding eigenvalue λ_u. Actually the values $e^{-it\lambda_u}$ are the eigenvalues of the Schrödinger Operator $\Psi(t) = e^{-it\boldsymbol{L}}$ as well as points in the Argand circle unit radius. We define

$$\begin{aligned}
\xi_{jk} &= \lim_{T\to\infty} \arg\left(\frac{1}{T}\int_0^T c_{jk}(t)dt\right) \\
&= \lim_{T\to\infty} \arg\left(\sum_{u=1}^{N} z_u(j,k)\frac{1}{T}\int_0^T e^{-it\lambda_u}dt\right) \\
&= \lim_{T\to\infty} \arg\left(\sum_{u=2}^{N} z_u(j,k)\frac{(1-e^{-iT\lambda_u})}{iT\lambda_u}\right) \\
&= \lim_{T\to\infty} \arg\left(\frac{1}{T}\left\{-iG(j,k) + \sum_{u=2}^{N} g_u(j,k)(\cos(T\lambda_u) - i\sin(T\lambda_u))\right\}\right) \quad ,(5)
\end{aligned}$$

which accounts for the limiting phase of the transition $c_{jk}(t)$. The limiting phase is obtained from $\arg(\xi_{jk})$, where $\arg(z) = \frac{1}{2i}\log\frac{z}{z^*}$. We have also that $G(j,k) = \sum_{u=2}^{N}\frac{1}{\lambda_u}\phi_u(j)\phi_u(k)$ is the Green's function of the Laplacian and $g_u(j,k) = \frac{1}{\lambda_u}\phi_u(j)\phi_u(k)$. The Green's function is the pseudo-inverse of the Laplacian, that is, $\boldsymbol{G}\boldsymbol{L}^+ = \boldsymbol{1}$. Such function, or more precisely the inverse $(\beta + \boldsymbol{L})^{-1}$ with $\beta > 0$, it is typically used for ranking the nodes in the graph [14], that is to estimate topological distances in the graph. For instance, if we define the vector $\boldsymbol{y} = (y_1 y_2 \ldots y_N)^T$ so that $y_j = 1$ if node j is assumed to be "known", and $y_k = 0$ for $j \neq k$, then the function $\boldsymbol{f} = (f_1 f_2 \ldots f_N)^T = (\beta+\boldsymbol{L})^{-1}\boldsymbol{y}$ yields higher values f_l at nodes l closer to j. Then, the facts that: i) the Green's function appears in the imaginary part of the limit, and ii) the terms of the Green's function appear both in the real and imaginary part, partially explains the structure of the matrix ξ. It is patched with several clusters of components ξ_{jk} in phase, provided that those components link nodes with similar topological distances.

3 Skeletal Similarity Matrices

The simplification of $\boldsymbol{W}$ is a key ingredient of our approach. Such simplification results from computing the long-time averages from magnitude. The diagonal of the symmetric matrix χ contains the long-time probabilities that a CTQW returns to each node. The off-diagonal elements $\chi(j,k)$ are the probabilities that a CTQW commencing at the $j-$th node reaches the $k-$th one in the limit. Then, since $\sum_{k=1}^{T}\chi(j,k)=1$, $\forall j$ we can associate a probability density function (pdf) to each node j. The fraction of off-diagonal probability mass $e(\boldsymbol{W})=\sum_{j=1}^{N}\sum_{k\neq j}^{N}\chi(j,k)/N$ measures the transport efficiency of the weighted

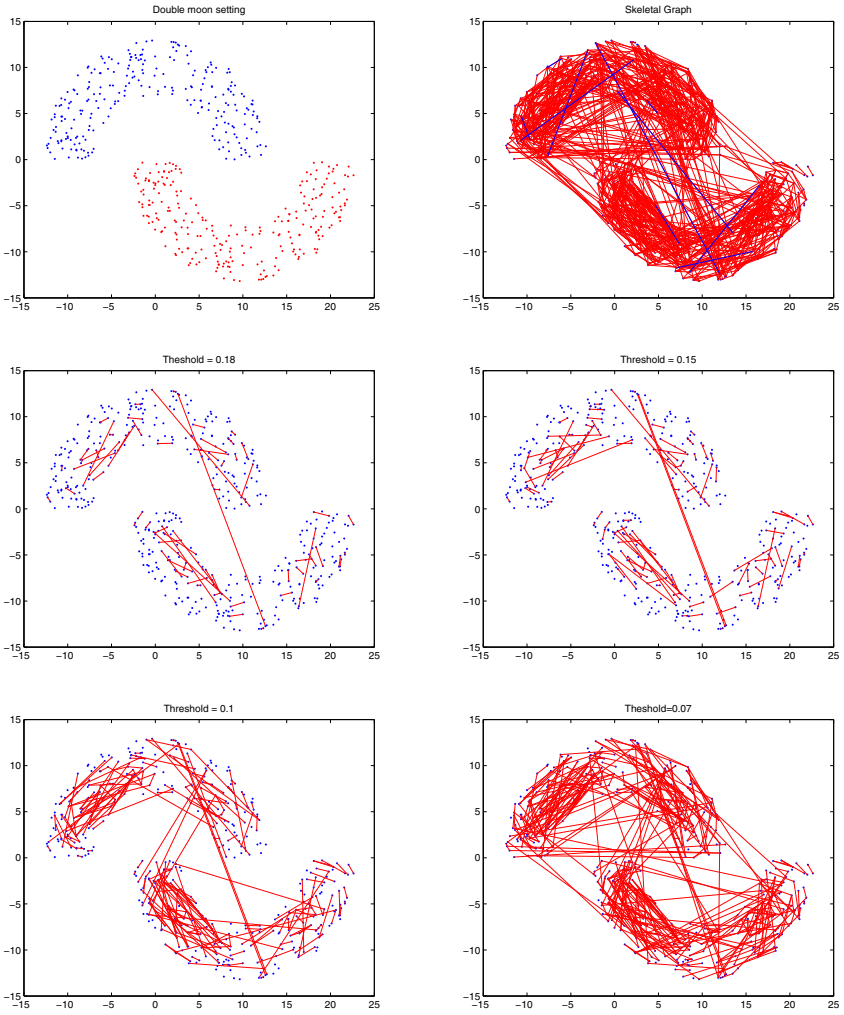

Fig. 1. Graph simplification: Double moon setting (top-left), skeletal graph (top-right), structural level sets for decreasing thresholds

graph. Since the CTQWs have a coherent behavior, $e(\boldsymbol{W})$ increases when there is enough similarity support, provided that the eigenvectors of the Laplacian matrix $\boldsymbol{L}$ are not degenerated. For instance, if $\boldsymbol{W}$ encodes a complete graph we have $\chi = \mathbf{1}$, that is, $e(\boldsymbol{W}) = 0$ since the eigenvectors of the Laplacian are: 0 with multiplicity one, and N, with multiplicity $N - 1$.

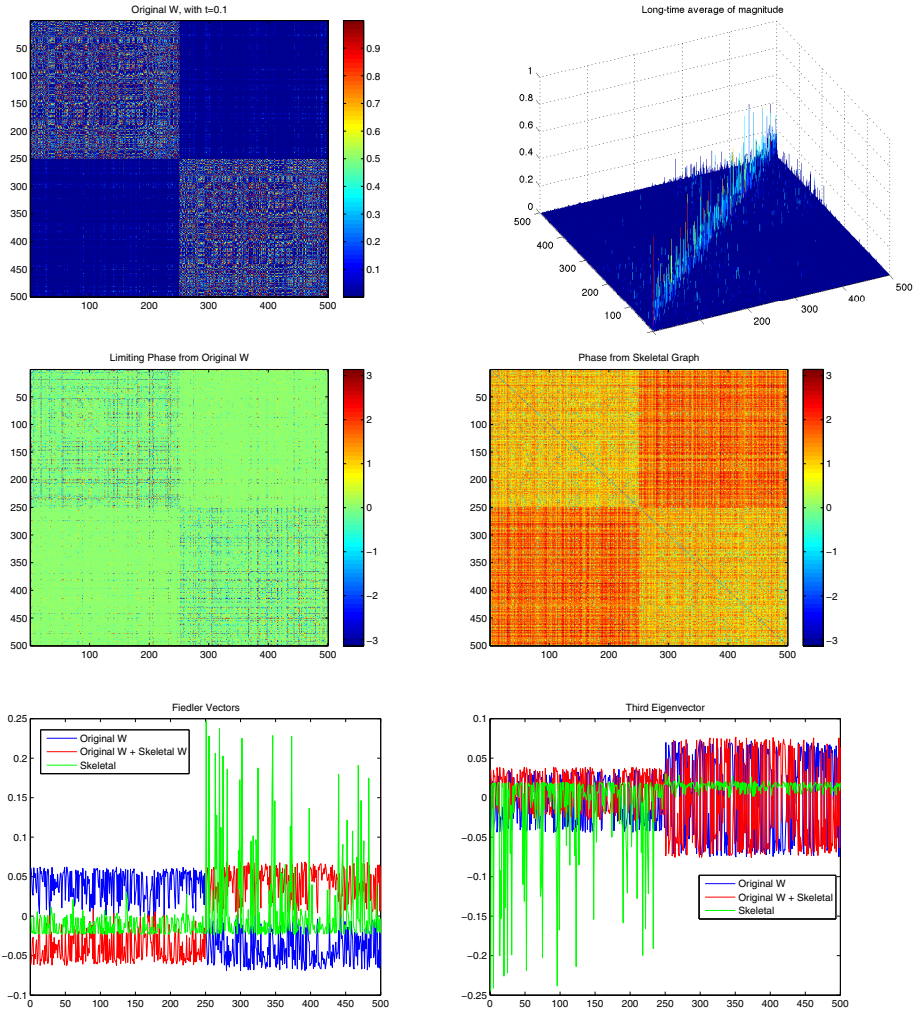

Fig. 2. Graph clustering: Similarity matrix for $t = 0.1$ (top-left), long-time averages of magnitude (top-left), long-time average of phase from the original similarity matrix (second row, left) and from the long-time average of magnitude (second row, right), comparison between the Fiedler vectors (third row, left) and between the third eigenvector (third row,right).

Since such support is not homogeneous we have that χ is sparser than $\boldsymbol{W}$. We also have that off-diagonal probabilities typically correspond to long-range interactions, i.e. to transitive links. If we sort the off-diagonal entries $\chi(j,k)$ in descending order, the sequence of graphs generated by incrementally decreasing the threshold γ used for retaining an edge if $\chi(j,k) \geq \gamma$ leads to a structural level set $\boldsymbol{S}_1, \boldsymbol{S}_2, \ldots, \boldsymbol{S}_r, \ldots$ dominated by long-range interactions. In addition, if we do not stop the sequence until we have $d_j = \sum_{k=1}^{N} S_r(j,k) > 0 \, \forall j$, then $\boldsymbol{S}_r$ is called a *skeletal graph*. In this context, non-zero degree is a mild connectivity criteria but it is enough to form a graph whose links connect points which are topologically close in the manifold, i.e. there is a minimal number of links between isolated clusters (specially when the bandwidth parameter t defining $W(i,j) = e^{-d^2(\boldsymbol{x}_i, \boldsymbol{x}_j)/t}$ is close to zero, as it happens when $N \to \infty$). Actually if our connectivity stopping criterion is stronger (e.g. find a unique connected component) then $\boldsymbol{S}_r$ becomes close to a complete graph. Therefore, the coherence of the CTQWs is exploited to form a sequence of graphs with minimal edge density, whereas if we build the graph from the original similarity information of short-range interactions dominate at the cost of a high density. To exemplify this behavior we have computed the structural level set and skeletal graph of a "double moon" setting (see Fig. 1-top left). The skeletal graph is given at $\gamma = 0.041$ (see Fig. 1-top right, where the last edges added to the structure are colored in blue).

The skeletal graph representation provides graph simplification by retaining only the edges in it. We can also discard nodes if we apply thresholds higher than the one providing the skeletal graph. But the representation is useful insofar the new graph retains good properties of the original $\boldsymbol{W}$, i.e., if it can be used to a given pattern recognition task (e.g. clustering or ranking) with a minimal loss of performance. For instance, when we apply this strategy to spectral clustering, it is better to combine the original affinity matrix with the "filtered" one (for instance by building $\boldsymbol{W}' = \boldsymbol{W} + \alpha\chi$) than to use χ or the weights associated to the skeletal graph alone. For instance, in Fig. 2 we show the results of clustering with and without filtered/sparse information. Fig. 2-top right shows that the CTQWs do not tend to link nodes of different clusters despite such links are significant in $\boldsymbol{W}$ due to the choice of the bandwidth t. The diagonal shows the return probabilities and also shows to what extent coherent transport happens in the graph: we have a high efficiency rate $e(\boldsymbol{W}) = 0.7414$, but most of the off-diagonal probabilities are retained by the links connecting nodes of the same cluster. This is not enough to let the skeletal graphs (or the weights of ξ) obtain a good clustering result. However, if we use $\boldsymbol{W}' = \boldsymbol{W} + \alpha\chi$ with $\alpha = 1$ we obtain a Rand index of 0.8609 which improves the Rand index obtained with the normalized cut method (0.824). The reason is explained by the structure of the Fiedler vectors of the Laplacians associated to each case (Fig. 2-third row, left). The Fiedler vector of $\boldsymbol{W}'$ (red) is very similar (in terms of partitioning information) to that of $\boldsymbol{W}$, but the amplitudes of their oscillations are slightly differ. This effect is better observed in the third eigenvectors (Fig. 2-third row,

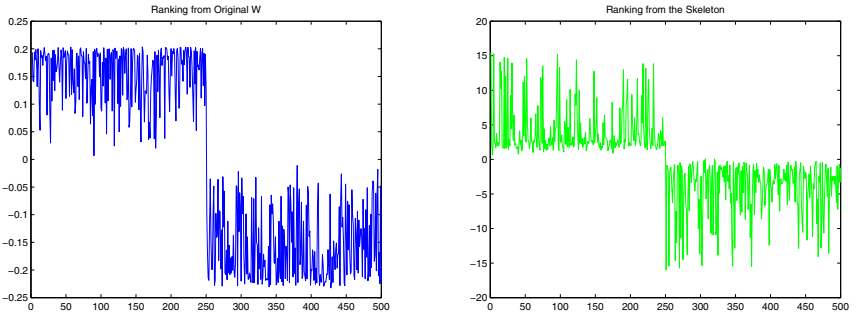

Fig. 3. Graph ranking: Ranking from the pseudo-inverse of the Laplacian. From the original similarity (first column), and from the skeletal similarity (second column)

right). Finally, the eigenvectors associated to the Skeletal graphs are unable to perform the partition.

Finally, in Fig. 3 we show that the skeletal version improves the original one despite having a more sparse pseudo-inverse (it is built from fragmentary evidence).

4 Conclusions

In this paper we have investigated the impact of continuous-time quantum walks in graph simplification, graph-based clustering and graph-based ranking. Future work includes intensive experimentation and the development of an information-theoretic interpretation of coherent transport.

Acknowledgements. Funding. F. Escolano: Project TIN2012-32839 (Spanish Gov.). E. R. Hancock: Royal Society Wolfson Research Merit Award.

References

1. Zhou, X., Belkin, M., Srebro, N.: An iterated graph laplacian approach for ranking on manifolds. In: Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Diego, CA, USA, August 21-24, pp. 877–885 (2011)
2. Shi, J., Malik, J.: Normalized cuts and image segmentation. IEEE Trans. Pattern Anal. Mach. Intell. 22(8), 888–905 (2000)
3. Qiu, H., Hancock, E.R.: Clustering and embedding using commute times. IEEE Trans. Pattern Anal. Mach. Intell. 29(11), 1873–1890 (2007)
4. Arbelaez, P., Maire, M., Fowlkes, C., Malik, J.: Contour detection and hierarchical image segmentation. IEEE Trans. Pattern Anal. Mach. Intell. 33(5), 898–916 (2011)
5. Weiss, Y., Torralba, A., Fergus, R.: Spectral hashing. In: Advances in Neural Information Processing Systems 21, Proceedings of the Twenty-Second Annual Conference on Neural Information Processing Systems, Vancouver, British Columbia, Canada, December 8-11, pp. 1753–1760 (2008)

6. Weiss, Y., Fergus, R., Torralba, A.: Multidimensional spectral hashing. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (eds.) ECCV 2012, Part V. LNCS, vol. 7576, pp. 340–353. Springer, Heidelberg (2012)
7. Liu, W., Wang, J., Kumar, S., Chang, S.: Hashing with graphs. In: Proceedings of the 28th International Conference on Machine Learning, ICML 2011, Bellevue, Washington, USA, June 28 - July 2, pp. 1–8 (2011)
8. Grady, L.: Random walks for image segmentation. TPAMI 28(11), 1768–1783 (2006)
9. Bai, L., Rossi, L., Torsello, A., Hancock, E.R.: A quantum jensen-shannon graph kernel for unattributed graphs. Pattern Recognition 48(2), 344–355 (2015)
10. Rossi, L., Torsello, A., Hancock, E.R., Wilson, R.C.: Characterizing graph symmetries through quantum jensen-shannon divergence. Phys. Rev. E 88, 032806 (2013)
11. Mülken, O., Blumen, A.: Continuous-time quantum walks: Models for coherent transport on complex networks. Physics Reports 502(2-3), 37–87 (2011)
12. Stone, M.: On one-parameter unitary groups in hilbert space. Annals of Mathematics 33(3), 643–648 (1932)
13. Farhi, E., Gutmann, S.: Quantum computation and decision trees. Phys. Rev. A 58, 915–928 (1998)
14. Zhou, D., Weston, J., Gretton, A., Bousquet, O., Schlkopf, B.: Ranking on data manifolds. In: Advances in Neural Information Processing Systems 16. MIT Press (2004)