Journal of Complex Networks (2017) **5**, 219–244
doi: 10.1093/comnet/cnw016
Advance Access Publication on 22 June 2016

A graph spectral-based scoring scheme for network comparison

VASUNDHARA GADIYARAM
Indian Institute of Science Mathematics Initiative (IMI), Department of Mathematics and Molecular Biophysics Unit, Indian Institute of Science, Bengaluru 560012, Karnataka, India

SAMBIT GHOSH
Department of Mathematics and Molecular Biophysics Unit, Indian Institute of Science, Bengaluru 560012, Karnataka, India

AND

SARASWATHI VISHVESHWARA†
Molecular Biophysics Unit, Indian Institute of Science, Bengaluru 560012, Karnataka, India
†Corresponding author: Email: sv@mbu.iisc.ernet.in

Edited by: Peter Csermely

[Received on 15 February 2016; accepted on 14 April 2016]

Study of real-world networks is paramount to understand the complex nature of interactions in nature. With a rise in the number of network-based approaches and data availability in various disciplines such as biology, physico-chemical sciences, earth sciences, engineering, economics and social sciences, there is a need for new approaches that are able to capture maximum network features without significant loss of information. The method developed in this article aims to address such requirements and can be applied to study different kinds of networks. Given accurate data points and their connections, it is a challenge to characterize and recognize global patterns. The graph/network spectra are known to hold all the information about the system. A careful analysis of graph spectra should provide the information at any desired level. Here, we have developed a scoring method to characterize and compare network patterns, using the solutions to normalized Laplacian of weighted matrices. The power of this method is demonstrated on real-world networks by comparing protein structure networks, which is an important problem in structural biology and financial stock networks to study dynamic changes. Various components of the scores provide insights at levels ranging from differences in edges to the transmission of these differences to global clustering levels. The scores developed here not only recognize network patterns in comparison with templates but also can serve as clustering technique to group them in a large pool of networks. Further, given accurate data of nodes and edges in a network, the method is applicable to problems in any discipline of interest.

Keywords: Network comparison; spectra of networks; weighted networks; structural analysis of networks; biological and molecular networks; stock networks.

1. Introduction

In recent times, successful interpretations of global features of large and complex systems such as brain, communications, social, biological and financial systems are becoming possible by investigating graphs or networks constructed from local connections. Analysis of such networks frequently requires a comparison between similar networks to find closeness between each other. Identification of changes between the predicted models and experimentally determined structures of macromolecules, tracking changes in

a metabolic network due to gene knockout or mutation, analysing microarray data, development of phylogeny trees and image analysis are some of the examples of network comparison in large complex biological networks.

Comparison of networks can be made either by exact or inexact matching. Exact matching methods require complete or partial node correspondence among the networks [1, 2]. Some approaches used in exact matching are tree search, subgraph isomorphism, maximum common subgraph and clique detection [3–14]. The rigid constraints and exponential worst-case time required for exact matching pose a limitation for real-world network comparison, where the networks are slightly deformed due to noise or non-deterministic elements. On the other hand, inexact matching methods use optimal inexact matching, continuous optimization and graph edit distance.

Another aspect in real-world networks is to compare quantitatively the extent of cluster matching. Spectral approaches are powerful in capturing minute differences between networks [15–27, 29–36]. For example, k-means clustering based on Fiedler vector and dominant modes from principal component analysis are frequently used to compare networks based on clusterings. However, in real-world networks, this reduction may lead to significant loss of information [24]. Also, the methods which consider only eigen values [36–41] cannot distinguish between iso-spectral and isomorphic networks. An earlier effort which includes Hamming distance and Ipsen–Mikhailov score (HIM score [42, 43]) partially addressed this problem, but the degree of change in node clustering is not captured. This is an important issue that deserves attention and we have addressed this in a later section. In this study, we have developed a method of network comparison, which addresses the issue of matching clusters at all levels by considering the spectra and the complete set of eigen vectors of the networks.

In Section 2, we discuss the development of a network score, which captures both local edge difference and global clustering changes in networks. Section 3 demonstrates the sensitivity and robustness of the method. An application of the method on protein structure network comparison, an area of high interest to structural biologist including our group, is presented in detail in Section 4.1. Here, we compare our score with a benchmark scoring technique, namely global distance test (GDT) and discuss the new information captured by our score. Further, in Section 4.2, the wide applicability of this method is demonstrated by analysing fluctuations in financial stock markets. Finally in Section 5, we discuss the key aspects of the Network Similarity Score (NSS) method and briefly highlight the importance of including node clustering in a network comparison method.

2. Method

Network information consisting of nodes and edges can be represented in various matrix forms such as adjacency, Laplacian or normalized Laplacian [25, 44–47]. The spectra (set of eigenvalues) obtained from such matrices provide valuable information about the details of connectivity in the network.

The aim of the study is to obtain a quantitative similarity index between symmetric adjacency matrices A and B, representing two undirected networks of equal size. The normalized Laplacian of a network with adjacency **A** is given by

$$\mathbf{L}^* = \mathbf{D}^{-1/2}\mathbf{L}\mathbf{D}^{-1/2}, \tag{1}$$

where D is the diagonal degree matrix defined as $D_{ii} = \deg(v_i)$, v_i is any vertex in the adjacency matrix A and L is the Laplacian of A, defined as $\mathbf{L} = \mathbf{D}\text{-}\mathbf{A}$.

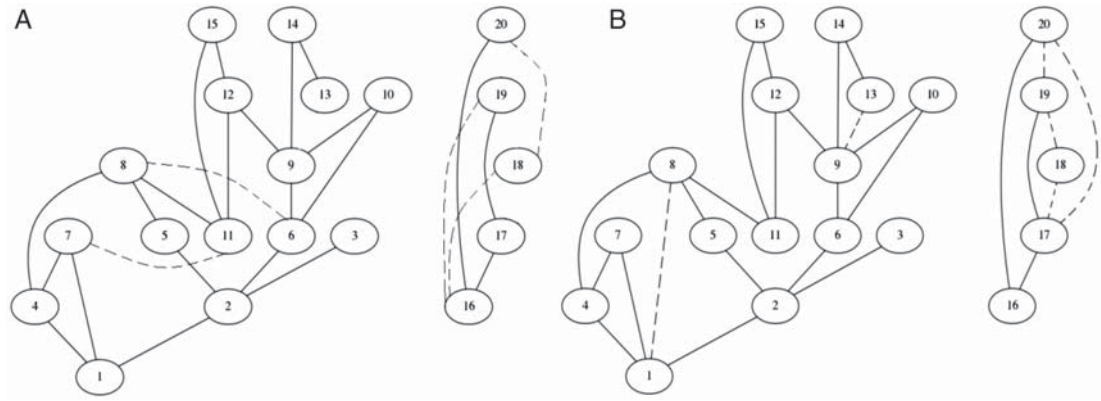

FIG. 1. Sample Network 1 (panel A) and Network 2 (panel B). Dotted edges are unique to the respective network and solid edges are common to Networks 1 and 2.

Once we get the normalized Laplacian matrix, the eigen decomposition of the same is carried out as follows

$$\mathbf{L}^* = \Phi \Lambda \Phi^{\mathrm{T}}, \tag{2}$$

where Λ is the diagonal matrix containing normalized eigen values and Φ contains the normalized eigen vectors. The eigenvalues of normalized Laplacian of a symmetric matrix are real, non-negative and lie between 0 and 2. The eigen vectors corresponding to these eigenvalues represent n (n is the size of the network) orthonormal vectors (mutually perpendicular unit vectors) in an n-dimensional space into which the network is embedded. The set of these eigen vectors acts as a basis to the n-dimensional space and the vector components of each vector give the projections of the nodes on that dimensional vector. Thus, the eigen maps of normalized Laplacian give non-linear embedding of the network onto a unit hypersphere, directly giving embedding of data points.

Two networks are close, if their eigen vectors and their eigenvalues are similar in their respective eigen spaces. A comparison of vertices of two networks from two different eigen decompositions requires alignment of the eigen vectors [48]. The identification of aligned vectors involves matching of node clustering of two networks defined in their projection subspaces, which is done by comparing the cosine similarity between all versus all eigen vectors of the two networks [49]. In the trivial case, where we compare a network with itself, the cosine value for an eigen vector with itself is equal to 1 (since angle is 0 degree) and with any other eigen vector is equal to 0 (since angle is 90 degrees). However, such orthonormality will not exist between the eigenvectors of two different networks and their cosine similarity values will be between 0 and 1 for the corresponding vectors of different networks. Furthermore, this value may turn out to be better (closer to 1) for non-corresponding vectors, resulting with a better alignment with a different vector.

The example of Network 1 and Network 2 of size 20 (represented in Fig. 1) is used to illustrate this point. For each eigen vector of Network 1, the eigen vector of Network 2 for which absolute cosine value is the largest in magnitude is assigned as the aligned eigen vector. (Supplementary appendix 1, Table S2 shows the cosine values between all eigen vectors of Network 1 and all eigen vectors of Network 2.) As we are interested only in the degree of matching and not the direction of alignment of vectors, we take absolute value of the cosines. Aligned eigen vectors presented in Table 1 show that the vectors 3, 4, 5

TABLE 1. *Aligned eigen vectors (eigen values) of Networks 1 and 2 with their absolute cosines and the weightage*

Network 1 Evecs (Evals)	1(0)	2(0)	3(0.158)	4(0.275)	5(0.37)	6(0.5)	7(0.55)	8(0.62)	9(1.074)	10(1.174)
Network 2 Evecs (Evals)	2(0)	1(0)	3(0.09)	4(0.248)	5(0.378)	19(1.608)	6(0.587)	8(0.741)	14(1.412)	13(1.37)
cosine value	0.903	0.969	0.924	0.928	0.83	0.784	0.843	0.726	0.376	0.646
Weightage	1	1	0.766	0.546	0.392	0.304	0.186	0.098	0.031	0.064
Network 1 Evecs (Evals)	11(1.174)	12(1.368)	13(1.488)	14(1.5)	15(1.5)	16(1.5)	17(1.5)	18(1.583)	19(1.82)	20(1.843)
Network 2 Evecs (Evals)	13(1.37)	13(1.37)	12(1.263)	15(1.5)	7(0.726)	7(0.726)	18(1.592)	17(1.56)	16(1.5)	16(1.5)
cosine value	0.646	0.692	0.656	0.57	0.835	0.649	0.633	0.615	0.677	0.766
Weightage	0.064	0.136	0.128	0.25	0.137	0.137	0.296	0.327	0.41	0.422

and 8 of Network 1 align with the vectors of same index of Network 2 (weightage in Table 1 is discussed in Section 2.1.2). However, the best correspondence is not necessarily with the same vector indices. For example, the 6th and 20th eigen vectors of Network 1 exhibit the best possible alignment with 19th and 16th eigen vectors of Network 2, respectively. Additionally, the extent of alignment between the best eigen vectors captured by the cosine values clearly indicates that the alignment is very good (>0.9) with the first four eigen vectors of Network 1 and is the worst (cosine value 0.376) with the 9th eigen vector.

2.1 *Scoring scheme*

From the above discussion, it is clear that a rigorous comparison of two networks can be made at global level by identifying the aligning vectors and the extent of alignment (cosine values) between them. Optimal weightage for different components of the normalized Laplacian matrices and their spectral features can further enhance the refinement of the NSS. Below we define three components of NSS: (1) Correspondence Score (CRS), which represents the extent of eigen vector correspondence, (2) Eigenvalue Weighted Cosine Score (EWCS), in which the cosine value captures the extent of best match between the aligned eigen vectors and the eigenvalues scale the extent of node participation in these vectors and (3) Edge difference Score (EDS), which gives importance for the edge weight differences in the networks.

2.1.1 *Correspondence Score.* When two networks are identical, their eigen vectors and their eigenvalues are identical in their respective eigen spaces. Therefore the eigenvectors of one network align with the eigenvectors of same index in the other network. This is reflected in the diagonal of the cosine matrix containing all 1s in ideal case (Supplementary appendix 1, Fig. S1). Because of orthogonality of the eigen vectors, we also see that the cosines off the diagonal are 0. As the two networks depart from each other, the eigen vectors in the first network tend to associate with different eigen vectors in the second network [50, 51]. Hence the cosine matrix acquires non-zero values off the diagonal, resulting in alignment of eigen vectors with different indices (Supplementary appendix 1, Fig. S2). However, the difference in the networks is reflected in the deviation of the aligned eigen vectors from the ideal case (in other words, deviation of maximum cosine values from the diagonal). To capture the extent of this shift, Spearman's correlation between the indices of aligned eigen values is used.

$$\text{CRS} = 1 - \frac{6 \sum (\text{IndexEvec}_A - \text{IndexEvec}_B)^2}{n(n^2 - 1)}, \tag{3}$$

where, IndexEvec_A and IndexEvec_B are the indices of aligned eigen vectors of Networks 1 and 2 with maximum cosine values (Table 1) and n is the size of the network. For the trivial case that involves comparison of a network with itself (Network 1 with itself), the CRS is 1. For the given sample networks (Networks 1 and 2), CRS is 0.7023. Thus, CRS gives an estimate of the correspondence at all connectivity levels.

2.1.2 *Eigenvalue Weighted Cosine Score.* It has been shown that any estimation using eigen vectors is better by using information about eigenvalues [49]. An interpretation of eigenvalues and eigen vectors of normalized Laplacian matrix is briefly presented here. The eigen vector corresponding to eigenvalue 0 of normalized Laplacian matrix gives information about the network as one single entity, where the multiplicity of 0 indicates the number of connected components in the network. The Fiedler vector corresponding to the second smallest eigenvalue gives information about the next level of clustering. As we move towards the eigenvalue 1 from 0, the vectors give information at each sublevel [34]. The eigenvalue 1 corresponds to nodes having exactly similar neighbourhood, in particular isolated edges [52]. The eigenvalues from 1 to 2 give a picture of branching at each clustering level [53]. Since most of real-world applications require similarity at global level as well as at local level, we give a weightage for the cosine values depending on the distance of the corresponding eigenvalues from 1 (Fig. 2). Such a weighing scheme will give more weightage for dominant eigen vectors. Hence we define cosine score as a weighted average of differences of cosine values from optimal score 1.

$$\text{cosine}(\theta_{ij}) = \frac{(\text{Evec}_i^A . \text{Evec}_j^B)}{\|\text{Evec}_i^A\| \|\text{Evec}_j^B\|} i, j \epsilon N, 1 \leqslant i, j \leqslant n \tag{4}$$

$$\text{EWCS} = \frac{\sum (1 - \text{cosine})^2 |1 - \text{Eval}_A||1 - \text{Eval}_B|}{\sum |1 - \text{Eval}_A||1 - \text{Eval}_\text{B}|}, \tag{5}$$

where EVec_i^A is the ith eigen vector of Network 1 and EVec_j^B is the jth eigen vector of Network 2, which is aligned with EVec_i^A. Eval_A and Eval_B are the eigen values of EVec_i^A and EVec_j^B. For the trivial case that involves comparison of a network with itself (Network 1 with itself), the EWCS is 0. For the given sample networks, EWCS is 0.672.

2.1.3 *Edge Difference Score.* The CRS and EWCS capture the differences at node clustering levels. The source of these differences, however, is due to changes at the edge level. In order to emphasize the difference at individual edge level, we use edge difference score (analogous to Ochiai coefficient), which is the Frobenius norm of the difference in adjacency matrices, normalized by the number of edges in both networks.

$$\boldsymbol{M = A - B} \tag{6}$$

$$\textbf{EDS} = \|\boldsymbol{M}\|_F / \sqrt{(\Sigma \textbf{edge weight}_A \text{ X } \Sigma \textbf{edge weight}_B)}, \tag{7}$$

where A and B are the adjacency matrices of Networks 1 and 2, M is the difference of the two adjacency matrices and $||M||_F$ is the Frobenius Norm of the difference matrix M. For the trivial case of comparing a network with itself, EDS is 0. For the given sample Networks 1 and 2, EDS is 0.0857.

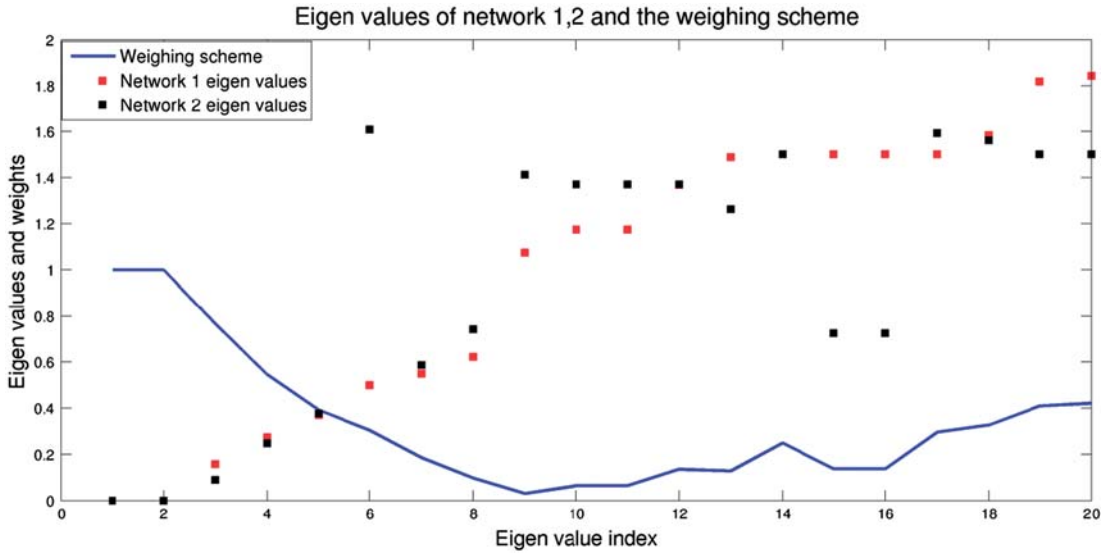

FIG. 2. Eigenvalue weighing scheme for the sample Networks 1 and 2: X axis represents the eigenvalue (1–20) indices and Y axis represents the eigenvalues of 20 aligned eigen vectors (Table 1).

The ideal values of the scores, being similarity of a network with itself are 1, 0 and 0, respectively, for CRS, EWCS and EDS. The closeness of Network 2 with Network 1 can be given by the Euclidian distance of the score components.

$$\mathrm{NSS} = \sqrt{(1 - \mathrm{CRS})^2 + (\mathrm{EWCS})^2 + (\mathrm{EDS})^2} \tag{8}$$

The NSS captures differences at edge level, differences in re-organizations of sub-clusters due to local changes, as well as changes in global connectivity observed in Network 2, as compared with Network 1. For the given sample networks, Network 2 is close to Network 1 with a score of 0.876 (maximum network difference score can be 1.732, which is equal to $\sqrt{1+1+1}$ for the worst case). It should be noted that smaller the score, closer are the networks.

3. Sensitivity and robustness of NSS

To perform sensitivity analysis of NSS, the networks in Fig. 3 are considered, where networks from sets 1, 2 and 3 differ equally from the original network, in terms of Hamming distance. The networks in each of the sets differ in symmetric positions and their spectra are identical as listed in Table 2. However, compared with original network, their node clustering is different. This is also shown by NSSs of these networks against original network (Table 3).

Networks in sets 1 and 2 are isomorphic and hence iso-spectral. These networks cannot be distinguished by comparative methods, which are based only on eigen values. But, as mentioned in Section 1, scores like HIM [42, 43] can distinguish such networks. However, a case arises where in sets 1 and 3, Hamming distance is same within each set. Hence HIM score for 1a versus 1b and 3a versus 3b is same. But as seen from the figure, in set 3, there is a major change in node clustering due to partitioning of the

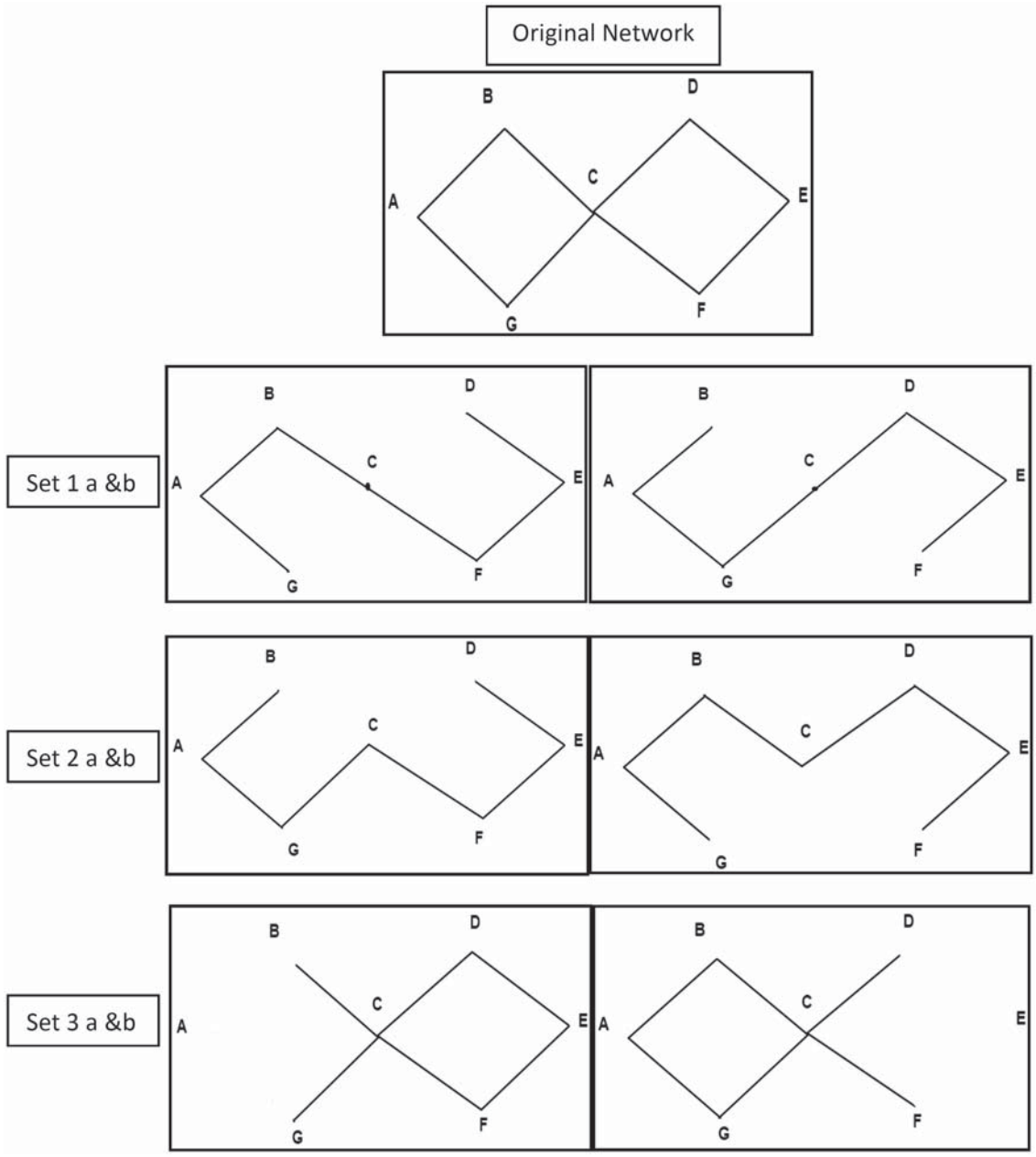

FIG. 3. Original network and sets 1, 2 and 3 contain networks of same Hamming distance from original network for sensitivity analysis. Each set contains networks which differ symmetrically.

TABLE 2. *Table of the adjacency, Laplacian and normalized Laplacian for original network and sets 1, 2 and 3. It can be seen that networks in sets 1 and 2 have identical spectra. This feature is also seen in networks of set 3*

			Adjacency spectra				
Original	−2.4495	−1.4142	0	0	0	1.4142	2.4495
Set 1a	−1.8478	−1.4142	−0.7654	0	0.7654	1.4142	1.8478
Set 1b	−1.8478	−1.4142	−0.7654	0	0.7654	1.4142	1.8478
Set 2a	−1.8478	−1.4142	−0.7654	0	0.7654	1.4142	1.8478
Set 2b	−1.8478	−1.4142	−0.7654	0	0.7654	1.4142	1.8478
Set 3a	−2.2882	−0.874	0	0	0	0.874	2.2882
Set 3b	−2.2882	−0.874	0	0	0	0.874	2.2882
			Laplacian spectra				
Original	0	0.5858	2	2	2.5858	3.4142	5.4142
Set 1a	0	0.1981	0.753	1.555	2.445	3.247	3.8019
Set 1b	0	0.1981	0.753	1.555	2.445	3.247	3.8019
Set 2a	0	0.1981	0.753	1.555	2.445	3.247	3.8019
Set 2b	0	0.1981	0.753	1.555	2.445	3.247	3.8019
Set 3a	0	0	0.7639	1	2	3	5.2361
Set 3b	0	0	0.7639	1	2	3	5.2361
			Normalized Laplacian spectra				
Original	0	0.2929	1	1	1	1.7071	2
Set 1a	0	0.1341	0.5	1	1.5	1.866	2
Set 1b	0	0.1341	0.5	1	1.5	1.866	2
Set 2a	0	0.1341	0.5	1	1.5	1.866	2
Set 2b	0	0.1341	0.5	1	1.5	1.866	2
Set 3a	0	0.5	1	1	1	1.5	2
Set 3b	0	0.5	1	1	1	1.5	2

TABLE 3. *NSSs of original network and sets 1, 2 and 3*

	Original	Set 2a	Set 2b	Set 3a	Set 3b	Set 4a	Set 4b
Original	0	0.1443	0.1443	0.1443	0.1443	0.1453	0.1453
Set 2a		0	0.3477	0.2898	0.2898	0.2369	0.2369
Set 2b			0	0.2898	0.2898	0.2369	0.2369
Set 3a				0	0.3477	0.2369	0.2369
Set 3b					0	0.2369	0.2369
Set 4a						0	0.3812
Set 4b							0

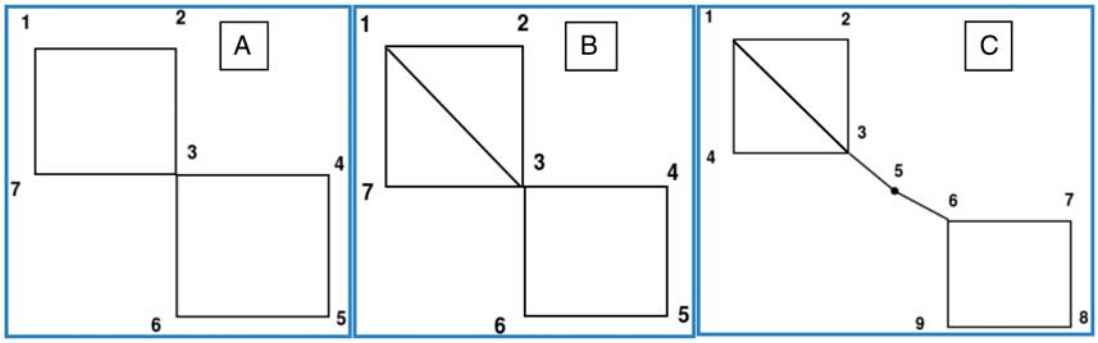

FIG. 4. From left to right panel: Networks A, B and C. Network B contains an additional edge between nodes 1 and 3 as compared to network A. Network C has two additional edges between nodes 3 and 6.

networks. This change is highlighted by higher NSSs between 3a versus 3b (0.3812) than 1a versus 1b (0.3477).

To illustrate the robustness of the method, we consider three control networks (A, B and C) shown in Fig. 4. For each network, each edge is changed from 0 to 1, in steps of 0.1 and the NSS of the changed network with respect to original network is calculated. Tables 5 and 6 consolidate the NSSs of the networks A and B with all their edges changed. The robustness of the method is illustrated by the gradual decrease in the NSSs, as each edge weight changes from 0 to 1.

Network A is a two-fold symmetric network, where we can consider edges 1–2, 1–7, 4–5 and 5–6 as outer edges and 2–3, 3–4, 3–6 and 3–7 as inner edges. Change in any of the outer (inner) edges is symmetric with respect to the original network (Network A). This symmetry is shown by NSSs as well, which are equal for the outer (inner) edges. Network B is same as Network A, except that the symmetry along one axis is removed by adding an edge between nodes 1 and 3. The symmetry along the other axis is still shown by pairs of similar NSSs for symmetric edges (coloured edges in Table 4). Figure 5 plots the NSSs of changed networks of network C, where each edge is changed from 0 to 1. It shows the importance of clustering of a network based on edge weights as given by higher NSSs.

With the help of the above-mentioned control networks, it can be seen that the NSS method is sensitive and robust with respect to the position and weightage of the edges in the network. The effectiveness of the above spectral scoring scheme on real-world applications has been highlighted in the following section, by addressing the problem of protein structure comparison and analysing financial stock correlations.

4. Applications

4.1 *Comparison of protein structure networks*

Protein structure networks have been extensively investigated to understand various aspects of protein structures. These aspects include studies on network behaviour [54–58] of structure-function correlation and pharmacological studies [59], communication within proteins and protein folding [58, 60].Valuable information is generally extracted from network features such as hubs, cliques, communities, shortest paths and so on [61–68]. In recent years, spectral methods have also been used to study different properties of protein structures [69–71].

A key problem in protein science is the analysis of protein structures. Coordinates of the atoms in a protein structure may be generated by experimental methods or by modelling techniques. Further,

TABLE 4. *NSSs for Network A at different edge weights. Given in first row are the edge weights for a particular edge and the NSSs are reported for a certain edge with a given weight. Edges of same colour have same position in the network, which also exhibit identical NSS values for a given edge weight, indicating the robustness of NSS method (The difference in the scores is not visible for higher edge weights in the table, due to lower precision.)*

Edge	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
1–2	0.09455	0.084449	0.074537	0.064771	0.05514	0.045644	0.036274	0.027028	0.017903	0.0089	0
4–5	0.09455	0.084449	0.074537	0.064771	0.05514	0.045644	0.036274	0.027028	0.017903	0.0089	0
5–6	0.09455	0.084449	0.074537	0.064771	0.05514	0.045644	0.036274	0.027028	0.017903	0.0089	0
1–7	0.09455	0.084449	0.074537	0.064771	0.05514	0.045644	0.036274	0.027028	0.017903	0.0089	0
2–3	0.09449	0.084442	0.074536	0.06477	0.05514	0.045644	0.036274	0.027028	0.017903	0.0089	0
3–4	0.09449	0.084442	0.074536	0.06477	0.05514	0.045644	0.036274	0.027028	0.017903	0.0089	0
3–6	0.09449	0.084442	0.074536	0.06477	0.05514	0.045644	0.036274	0.027028	0.017903	0.0089	0
3–7	0.09449	0.084442	0.074536	0.06477	0.05514	0.045644	0.036274	0.027028	0.017903	0.0089	0

TABLE 5. *NSSs for Network B at different edge weights. Given in first row are the edge weights for a particular edge and the NSSs are reported for a certain edge with a given weight. Edges of same colour have same position in the network, which also exhibit identical NSS values for a given edge weight, indicating the robustness of NSS method (The difference in the scores is not visible for higher edge weights in the table, due to lower precision.)*

Edge	**0**	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**	**1**
1–2	0.090013	0.074629	0.065882	0.057279	0.048797	0.040423	0.03215	0.023973	0.015891	0.007901	0
5–4	0.083602	0.074603	0.065871	0.057277	0.048798	0.040423	0.03215	0.023973	0.015891	0.007901	0
6–5	0.083602	0.074603	0.065871	0.057277	0.048798	0.040423	0.03215	0.023973	0.015891	0.007901	0
2–1	0.083535	0.074547	0.065851	0.05727	0.048795	0.040423	0.03215	0.023973	0.015891	0.007901	0
1–7	0.083535	0.074547	0.065851	0.05727	0.048795	0.040423	0.03215	0.023973	0.015891	0.007901	0
2–3	0.083481	0.074577	0.065862	0.057274	0.048796	0.040423	0.03215	0.023973	0.015891	0.007901	0
3–7	0.083481	0.074577	0.065862	0.057274	0.048796	0.040423	0.03215	0.023973	0.015891	0.007901	0
3–4	0.083363	0.074544	0.065851	0.05727	0.048795	0.040423	0.03215	0.023973	0.015891	0.007901	0
3–6	0.083363	0.074544	0.065851	0.05727	0.048795	0.040423	0.03215	0.023973	0.015891	0.007901	0

changes in known structures may be brought about by dynamics or by perturbations such as binding of ligands/mutations/changes in environment and so on. An important issue is to accurately measure the changes, both at the local and at the global level to interpret the effect of perturbation or simulation or to assess the quality of modelled structures. The standard methods available for protein structure comparison are root mean square deviation (RMSD), GDT-Total Score (GDT-TS) [49], Template Modelling-Score (TM-SCORE) [50], Local Distance Difference Test (LDDT) [51], Contact Area Difference-Scores (CAD) [52] and others [53]). Most of them are based on local or global alignment of protein structures, which consider the backbone atoms and only a small portion of the structures [53]. A few measures like Global Distance Calculation-Side chains (GDC-SC), LDDT consider side chain atoms. However, they do not represent side chain interactions in detail. The need for an alignment-free method which also considers

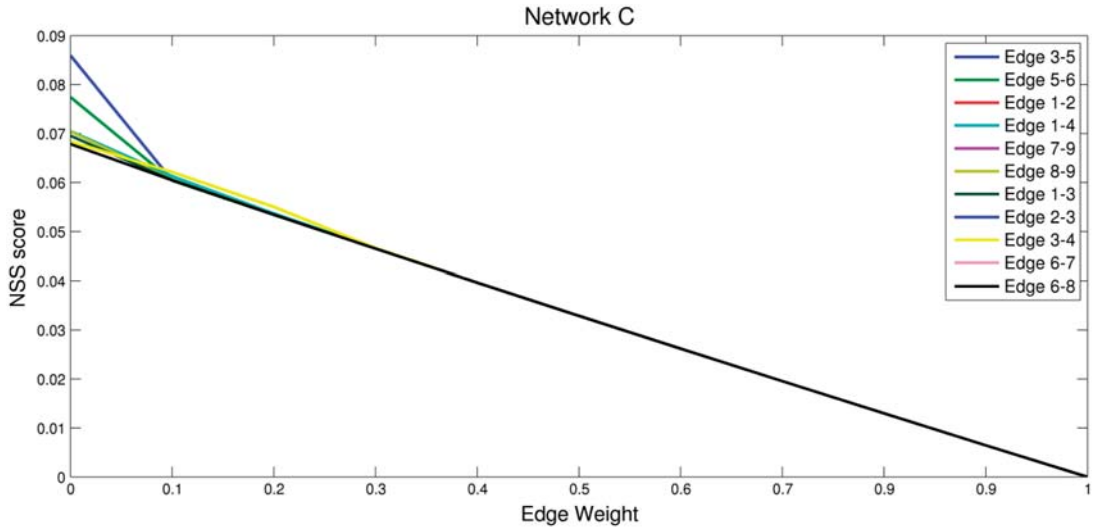

FIG. 5. Plot of NSSs at varying edge weight for network C.

the side chain interactions between the amino acids within the protein and their effect over the complete structure is stressed by Critical Assessment of protein Structure Prediction (CASP) evaluators [53]. The method developed here aims to address this issue by taking a network spectral approach for both backbone and side chain atoms and comparing them at all levels of node clustering.

4.1.1 *Construction of protein backbone and side chain networks.* The nodes and edges are defined in various ways to cast the protein structure in the network format [72]. Basically the level of definition depends on the problem to be addressed. For example, the topology of the protein backbone describes hundreds of different folds observed from experimental structure analysis. On the other hand, the code for a protein to adopt a specific three-dimensional fold/structure is embedded in the one-dimensional sequence of amino acid chain. The detailed non-covalent interactions (edges) between amino acid side chains (nodes) determine the final topology (network). A study of this sequence-structure relationship is the basis for extensive research in the area of protein folding. The characterization of global topology focuses on interactions at the backbone level, whereas the detailed side chain interactions not only contribute to the topology but are also associated with biological function. In such cases, the networks are studied at side chain interaction level. Further, edge weights can be assigned between two amino acids side chains depending on the extent of interactions. The details of constructions of such networks and the chemical aspects of the backbone and the side chain are provided in a recent review [72]. In this study, the backbone network is constructed using the coordinates of Cα atoms and the coordinates of all atoms are used for side chain network. Briefly, a contact-based adjacency matrix is created, which captures the overall topology of the protein backbone. The following criterion is used to generate the pair-wise contacts.

$$A_{\text{backbone}}(i,j) = 1 \text{ if distance } (i\text{th and } j\text{th C}\alpha \text{ atom}) \leqslant 6.5\text{Å} \tag{9}$$

$$A_{\text{backbone}}(i,j) = 0 \text{ if distance } (i\text{th and } j\text{th C}\alpha \text{ atom}) > 6.5\text{Å} \tag{10}$$

where A_{backbone} is the backbone Cα adjacency matrix (of dimensions n *X* n where n is the number of amino acid residues) for a protein and $j \geqslant i+2$. It is evident that the matrix is symmetric along the diagonal and all the diagonal elements are zero. Also, such a backbone matrix is unweighted.

The side chain adjacency matrix is a weighted matrix constructed to transform the uniquely folded geometry of the proteins at the side chain level to a two-dimensional matrix of weights between each contact pair. The algorithm picks up all non-hydrogen atoms of the side chain (i.e. all atoms apart from the backbone C, O, Cα and N) contacts based on the following criteria:

$$A_{\text{Side chain}}\,(i,j) = I_{i,j} \text{ if distance}(i\text{th and } j\text{th side chain atom}) \leqslant 4.5\text{Å}, I_{ij} \epsilon (0,1] \tag{11}$$

$$A_{\text{Side chain}}(i,j) = 0 \text{ if distance } (i\text{th and } j\text{th side chain atom}) > 4.5\text{Å} \tag{12}$$

$$\boldsymbol{I}_{i,j} = \boldsymbol{n}_{i,j}/\boldsymbol{N}_{ij} \tag{13}$$

where $A_{\text{Side chain}}$ is the side-chain-weighted adjacency matrix (of dimensions n *X* n, where n is the number of amino acid residues) for the protein, $j \geqslant i+2$, $I_{i,j}$ is the edge weight given to a pair of amino acids based on the total number of side chain atom–atom contacts between the ith and the jth residue ($n_{i,j}$) and the maximum possible number of contacts that pair can make (N_{ij}) across a database of high-resolution proteins obtained from the Protein Data Bank (http://www.rcsb.org/). Once we get the backbone and the side chain adjacency matrices, we follow the steps discussed in Section 2.

4.1.2 *Model evaluation.* The structures predicted for CASP (http://www.predictioncenter.org/) evaluation, which are being conducted for the past two decades, form an excellent source of modelled structures along with respective structures determined by X-ray crystallography (native structure). To demonstrate the effectiveness of the present spectral method, we discuss the models from TR759 (human periplakin containing 113 residues (Protein Data Bank ID, 4Q28)) under the Refinement category in 11th edition CASP (2014). The method is applied to the backbone and side chain networks separately, while evaluating the modelled structures. As a result, a model protein gives two scores: NSS for the backbone and NSS for the side chain when compared with the corresponding native protein. Figure 6 gives a graphical spread of the two scores for all the TR759 submitted models, keeping the native protein as the reference. Such a spread is very useful to understand the behaviour of a group of structures with respect to a common reference. For example, qualitatively, it is apparent from Fig. 6 that models from TR759 have more variation in their side chain networks over the backbone networks.

To look in-depth into our method performance and discuss some key observations, we select four pairs of models from the TR759 model pool (also highlighted in Fig. 6) based on their similarity at the backbone or side chain scores.

From the scores given in Table 6, it can be seen that the models (TS008_3 and TS008_5) have the same NSS-backbone but differ in their NSS-side chain, with respect to the native. The identity in NSS-backbone is also reflected in scores from other reported methods (Supplementary appendix 1, Table S3). The side chain networks show very minute deviations, which is reflected in small difference in the side chain score (0.037) and the values in Table 6 clearly show that all the different components of the NSS-side chain contribute to this change.

To study such subtle changes between the two models, we take one of the models (TS008_5) as the reference and the other (TS008_3) as the test for comparison. Figure 7 shows the differences in the atom contacts of the two models highlighting the fact that small changes in atom contacts give raise to differences in side chain scores.

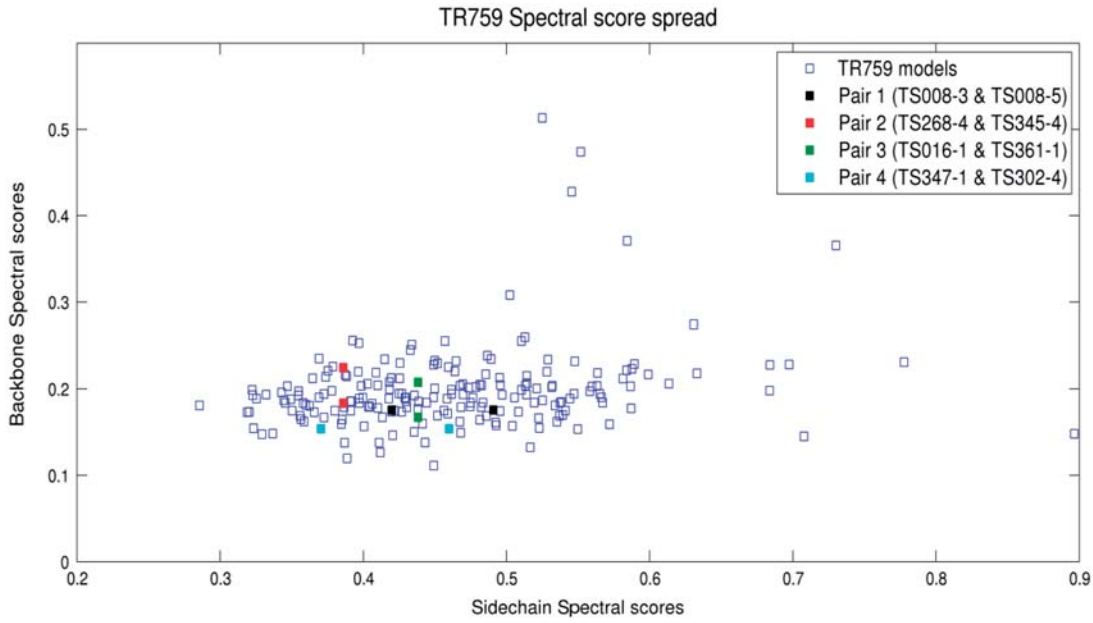

FIG. 6. TR759 backbone versus side chain spectral score spread. Selected pairs 1 and 4 have similar backbone scores and pairs 2 and 3 have similar side chain scores

TABLE 6. *Decomposition of backbone and side chain scores for four pairs of models selected from Fig. 3. The pairs exhibit identical scores with reference to the native structure and different scores between themselves. Scores in boldface highlight the pairs with similar scores*

Pairs of networks		Backbone				Side chain			
Pairs	Model	CRS	EWCS	EDS	NSS	CRS	EWCS	EDS	NSS
1A	Native vs. 008−3	0.953	0.164	0.040	**0.176**	0.656	0.208	0.117	0.419
1B	Native vs. 008-5	0.953	0.164	0.040	**0.176**	0.571	0.209	0.116	0.491
1C	008-5 vs. 008-3	1	0	0	0	0.997	0.034	0.013	0.037
2A	Native vs. 268-4	0.945	0.153	0.038	0.167	0.654	0.241	0.119	**0.438**
2B	Native vs. 345-4	0.929	0.191	0039	0.208	0.679	0.275	0.116	**0.438**
2C	268-4 vs. 345-4	0.997	0.058	0.016	0.060	0.949	0.092	0.059	0.121
3A	Native vs. 016-1	0.919	0.160	0.041	0.183	0.733	0.258	0.106	**0.386**
3B	Native vs. 361-1	0.942	0.213	0.038	0.225	0.729	0.258	0.095	**0.386**
3C	016-1 vs. 361-1	0.918	0.205	0.043	0.225	0.672	0.253	0.0946	0.425
4A	Native vs. 302-4	0.938	0.136	0.040	**0.154**	0.711	0.206	0.106	0.370
4B	Native vs. 347-1	0.957	0.143	0.036	**0.154**	0.610	0.216	0.115	0.460
4C	302-4 vs. 347-1	0.974	0.170	0.028	0.175	0.836	0.153	0.084	0.240

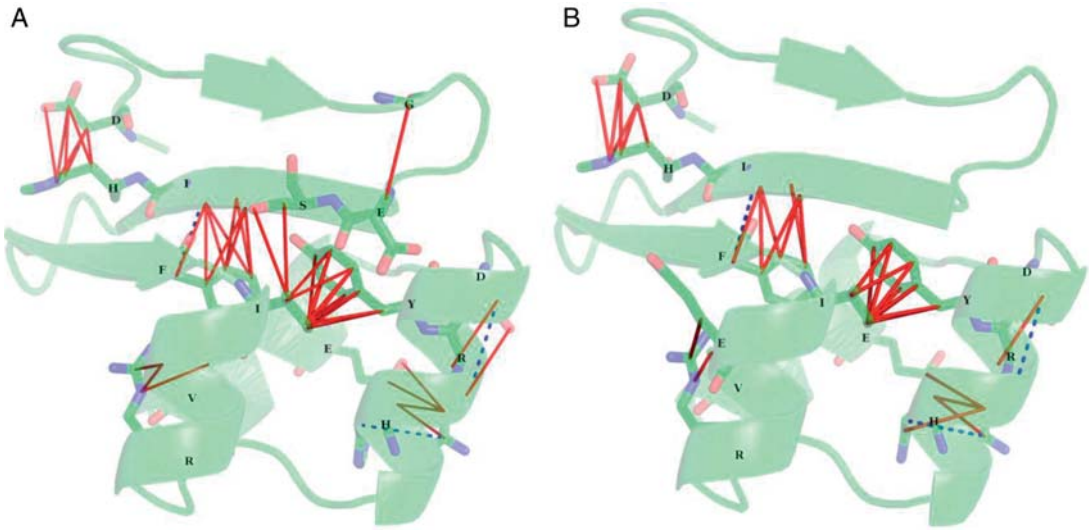

FIG. 7. TS008-5 (A) and TS008-3 (B) atom contact changes with native. Solid lines are additional atom–atom contacts in the respective models when compared to native and dotted lines are contacts unique to native protein, not present in models (although the edge weights in the two models are considerably different from the native, only those connections which differ among themselves are chosen here for depiction).

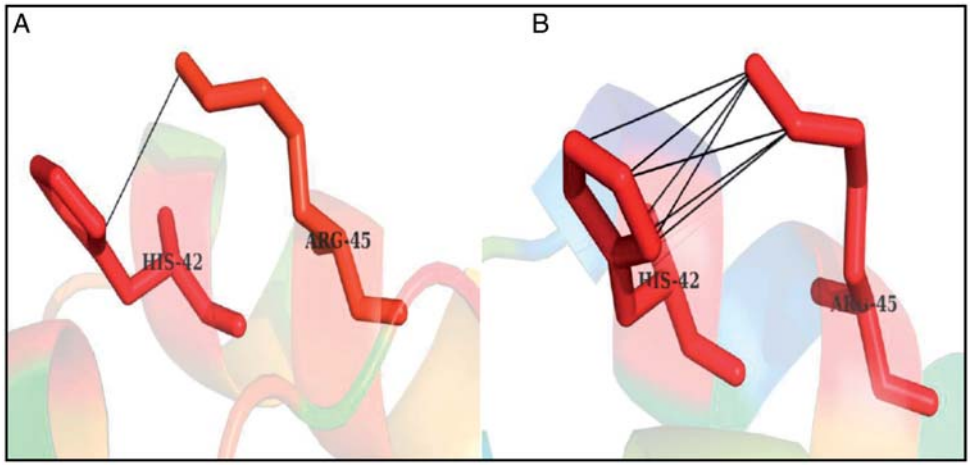

FIG. 8. Example of edge weight difference as a result of change in the atom–atom contacts between TR008_5 (A) and TS008_3 (B).

To probe into the structure-dependent network changes, we first look at the changes at the edge level in the side chain networks, which is captured by EDS. The detailed atomic contacts of the residue (node) pairs with different edge weights are listed in Table 7 and a pictorial representation of side chain atom contacts between residues 42 and 45 in both the models is shown in Fig. 8. Such structural changes result in different side chain edge weights between corresponding residue pairs.

TABLE 7. *Differences in edge weights between the models TS008_3 and TS008_5 due to differences in the number of interacting atom pairs. Boldface atom pairs are exclusive to the respective model and non-boldface atom pairs are common to both models*

	TS008_3		TS008_5		
Residue pair	Atom contacts	Edge weight	Atom contacts	Edge weight	Difference in edge weight
15(SER) 29(ILE)	NONE	0	**CB-CG2,OG-CG2**	0.25	0.25
10(LYS) 14(GLU)	NONE	0	**CA-CB**	0.2	0.2
37(ARG) 60(VAL)	NE-CG1	0.0666	NE-CG1,**NE-CG2,CZ-CG1**	0.2	0.1333
45(ARG) 50(ASP)	CB-CB	0.0435	CB-CB,**CG-OD2,CD-OD2**	0.1304	0.0869
42(HIS) 45(ARG)	**CG-NE**,CG-CZ,**CE1-CZ**,NE2-NE, NE2-CZ,CD2-NE,CD2-CZ	0.25	CG-CZ,NE2-NE,NE2-CZ,CD2-NE,CD2-CZ	0.1786	0.0714
18(ILE) 27(PHE)	CB-CB,CB-CG,CB-CD2,CG1-CE2,CD1-CG,CD1-CD2,CD1-CE2,CG2-CD2	0.4286	CB-CB,CB-CG,CB-CD2,**CB-CE2,CG1-CD2**,CG1-CE2,CD1-CG,CD1-CD2,CD1-CE2,CG2-CD2	0.4762	0.0476
29(ILE) 46(TYR)	CB-CD2,CB-CE2,CG1-CB,CG1-CG,CG1-CD2,CG1-CE2,CD1-CB,CD1-CG,CD1-CD1,CD1-CZ,CD1-CD2,CD1-CE2	0.6191	CB-CD2,CB-CE2,CG1-CB,CG1-CG,**CG1-CD1**,CG1-CD2,CG1-CE2,CD1-CB,CD1-CG,CD1-CD1,**CD1-CE1**,CD1-CZ,CD1-CD2,CD1-CE2	0.6666	0.0476
2(ASP) 19(HIS)	CB-CB,CB-CG,CB-CD2,CG-CG,CG-CD2,OD2-CD2	0.2609	CB-CB,CB-CG,CB-CD2,**CG-CB**,CG-CG,CG-CD2,OD2-CD2	0.3043	0.0435
31(GLU) 37(ARG)	**CB-CZ**	0.3703	NONE	0	0.037
45(ARG) 55(GLU)	NE-CD,NE-OE2,CZ-OE2	0.1111	NE-CD,**NE-OE1**,NE-OE2,CZ-OE2	0.1481	0.037

TABLE 8. *Aligned eigen vectors and their cosine values of the side chain network for TS008_5 and TS008_3*

TS008_5 eigen vectors	1	2	3	4	5	6	7	8	9	10
TS008_3 eigen vectors	1	2	4	3	6	7	8	9	9	10
Max cosine values	0.839	1.000	0.937	0.401	0.994	0.886	0.878	0.742	0.619	0.956
	11	12	13	14	15	16	17	18	19	20
	11	12	13	15	16	17	18	20	19	24
	0.993	0.990	0.998	1.000	0.995	0.993	0.962	0.567	0.985	0.600
	21	22	23	24	25	26	27	28	29	30
	24	22	23	21	25	26	27	28	32	33
	0.720	0.960	0.982	0.560	0.905	1.000	0.947	0.894	0.478	0.365
	31	32	33	34	35	36	37	38	39	40
	33	31	32	34	35	36	37	38	38	40
	0.496	0.890	0.762	0.856	0.893	0.947	1.000	0.912	0.380	0.997
	41	42	43	44	45	46	47	48	49	50
	41	39	39	44	43	45	46	47	49	50
	0.797	0.769	0.551	0.999	0.860	0.952	0.988	0.991	1.000	0.992
	51	52	53	54	55	56	57	58	59	60
	51	52	53	54	56	55	61	57	58	60
	0.988	0.940	0.869	0.929	0.752	0.920	0.728	0.987	0.997	0.841
	61	62								
	62	59								
	0.874	1.000								

The differences in edge weights in the two networks are not only localized to the corresponding edges but also influence the node clustering in their respective eigen spaces. An analysis of the corresponding eigen vectors and their cosine values provides deeper insights into this global level clustering changes. This is demonstrated from the data presented in Table 8. Here, it can be observed that a comparison of 62 versus 62 eigen vectors has resulted in selecting the best aligned vectors and the extent of alignment. Most of the eigen vectors of 008_5 align well with corresponding eigen vectors of 008_3, resulting in very good correlation score (0.997) in Table 8. Also it is observed that most of the cosine values are very close to 1, signifying a good alignment in node clustering along their aligned eigen vectors. A few examples of alignments and the node contributions to the aligned vectors can be seen from Fig 9. It is evident that nodes (residues) from networks of both models have similar vector component magnitude, when the alignment is very good (cosine value close to 1) (Fig. 9A). The eigen vectors along which clustering of residues vary more is given by lower cosine scores and also deviations in the extent of node participation (Fig. 9B), representing the global effect of local edge-weight changes.

To demonstrate local clustering changes due to difference in edge weights, the node clustering of 9th eigen vectors of both models are mapped on their structures (Fig. 10). Cluster A shows residues His42, Arg45 and Glu55, with different atom contacts. It is evident that residue His42 is showing higher cluster participation in 008_3 due to higher edge weight (atom contacts). Similarly clusters B and C show differences in residue participations in accordance with their edge weights (atom contacts).

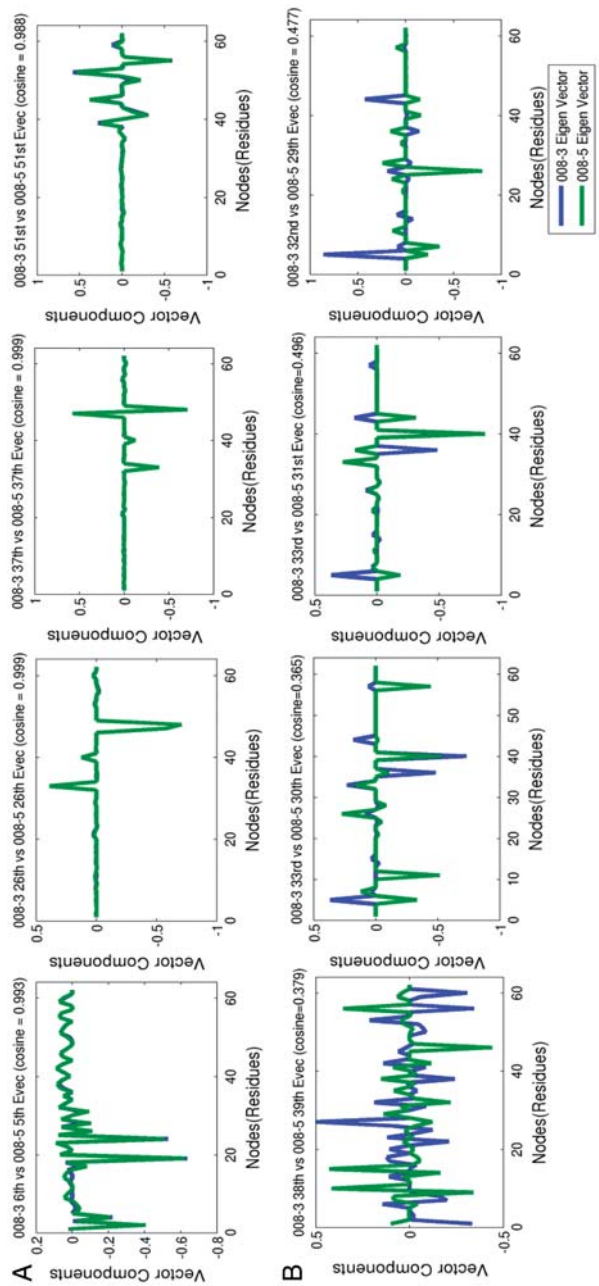

FIG. 9. Vector components of the nodes in aligned eigen vectors from the side chain networks of the two refinement models: Examples of higher cosine scores (panel A) and lower cosine scores (panel B). (Note that in panel A, vector components overlap to a great extent due to near identical values).

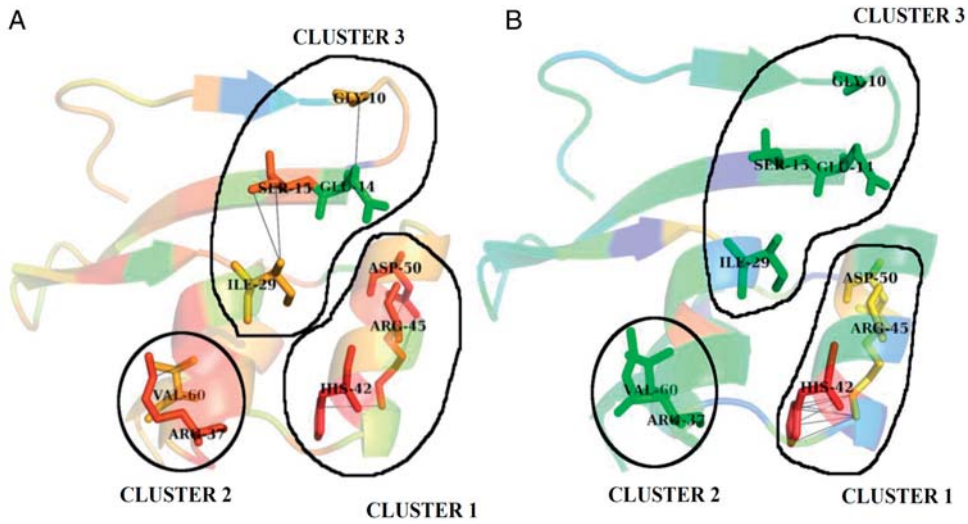

FIG. 10. TS008-5 (A) & TS008-3 (B) clustering changes for the 9th eigen vector. Colours are given according to node eigen vector components. Cluster 1: His42, Arg45 and Asp50; Cluster 2: Val60 and Arg37 and Cluster 3: Gly10, Ser 15, Glu 14 and Ile 29.

Figure 11 shows the model TS008_5 on which changes between TS008_5 and TS008_3 are mapped. The red (thick) lines are the edges where both the models differ in their weights. These local changes result in change in global cluster participation of residues in red (bigger) spheres. It should, however, be noted that the edge weights of these residues are the same. This is an example of global clustering change due to changes in the local connectivity.

4.2 *Stock network variation before and after 2008 crisis*

Financial markets have seen a major crisis during the year 2007–2008. The markets which were rather calm before this period experienced major changes and the correlation between different sectors increased significantly. Because of relative stability of the Indian markets, this crisis did not drastically change the stock returns in India as compared to other markets [73–75]. A study of 36 stocks from various sectors in Indian markets from the period 2004 to 2015 has been performed to analyse the change in the clustering among these sectors. The list of stocks considered is given in Table 9.

Stock network for each year is constructed by considering each stock (36 stocks shown in Table 9) as a node (36 nodes) and edges having weights based on the absolute correlation between returns from the stocks. The variation in these correlations is studied by comparing stock networks from years 2004 to 2015 using NSS. Figure 12 shows the NSSs of 2004 network against other years. The networks in post-crisis period are stabilizing at a farther level compared to 2004.

An all-versus-all NSS comparison of all the networks (from 2004 to 2015) is shown in Fig. 13. It can be seen that networks pertaining to 2004 to 2007 form one group indicating their similarity in stock correlations. From 2008 to 2013, the correlations for these stocks change due to the crisis and they form a different group. This indicates that the networks (stocks) before crisis follow a certain market trend, which changes drastically after the crisis. Additionally, the year 2014 shows less correlation with all the previous years, whereas the correlation with the second box seems to have improved in the year 2015.

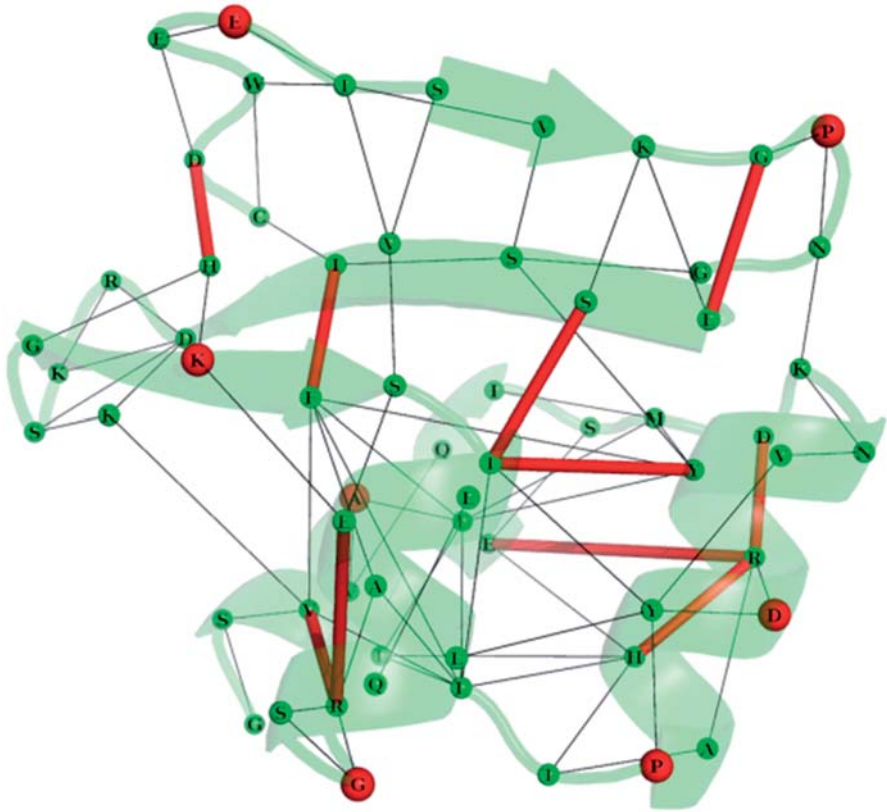

FIG. 11. Global changes in the protein networks of TS008_5 and TS008_3 mapped onto TS008_5. Red (thick) lines are the edges with different weights and black lines are edges having same weights in both the models (Table 7). Red (bigger) spheres are residues having significant change in their cluster participation between the models.

TABLE 9. *Stocks selected from various Indian market sectors*

	Bank		IT		Healthcare		Cement		Capital goods
1	SBI	11	TCS	17	LUPIN	23	ACC	31	BHEL
2	PNB	12	INFOSYS	18	SUNPHARMA	24	AMBUJA	32	L&T
3	BOB	13	WIPRO	19	CIPLA	25	GRASIM		
4	CANBK			20	DRREDDYS	26	ULTRATECH		Private bank
5	UNIONBANK		FMCG					33	AXIS_BANK
6	ANDHRABANK	14	ITC		Metal		Automobile	34	ICICI
7	ALBANK	15	HUL	21	TATASTEEL	27	TATAMOTORS		
8	BANKINDIA	16	GILLETTE	22	HINDALCO	28	M&M		Oil and energy
9	DENABANK					29	MARUTISUZUKI	35	IOC
10	IOB					30	HEROMOTORS	36	ONGC

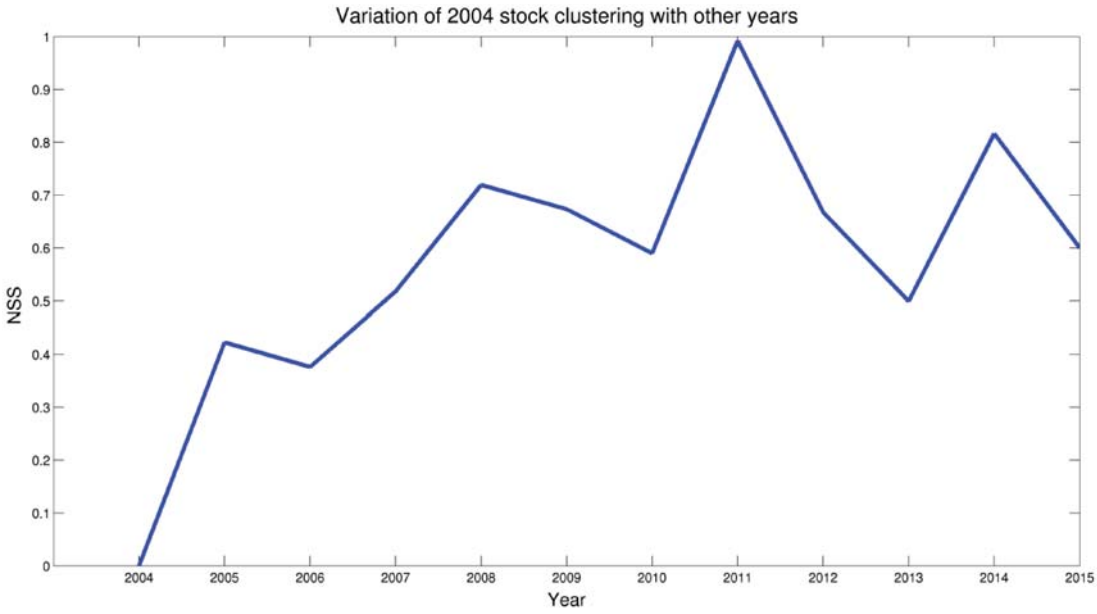

FIG. 12. Variation of stock networks during 2004–2015. The NSS values on Y axis are with reference to the network for the year 2004

It is interesting to note that the transient major fluctuation in 2014 shown from our study coincides with the national election in India.

The average networks of the first group (2004–2007) and the second group (2008–2013) are constructed and the clustering of the stocks in these networks are compared, as shown in Fig. 14. We can clearly see the inter-stock correlations increased for each sector and also between sectors. Also, private banks become more synchronized with public sector banks after the crisis. The healthcare sector does not show a high correlation pre- or post-2008 crisis because the stocks of pharmaceutical companies are relatively stable due to constant demand from consumers. Same is the case with FMCG. The change in the inter-stock correlations leads to change in the network clustering for each year. This dynamic change is captured by NSS and clearly shows two overall clusters in Fig. 13.

5. Discussion

Above we have presented a method to score networks based on the analysis of the spectra of networks. We have also demonstrated the capabilities of the method to extract local and global differences in terms of network clustering. Here, we discuss a few aspects which have not been elaborated in the above section. Although the points discussed below pertain to protein structure networks, they are of general in nature and are applicable to complex networks generated in any context.

Firstly, as mentioned in Section 1, the method is designed for the exact matching case, where the node correspondence is known. However, we wish to point out that it can be used for different sized networks by adding dummy nodes and edges (avoiding isolated nodes) at appropriate positions. For example, networks of two proteins of different sizes can be compared by converting them to the same size. This

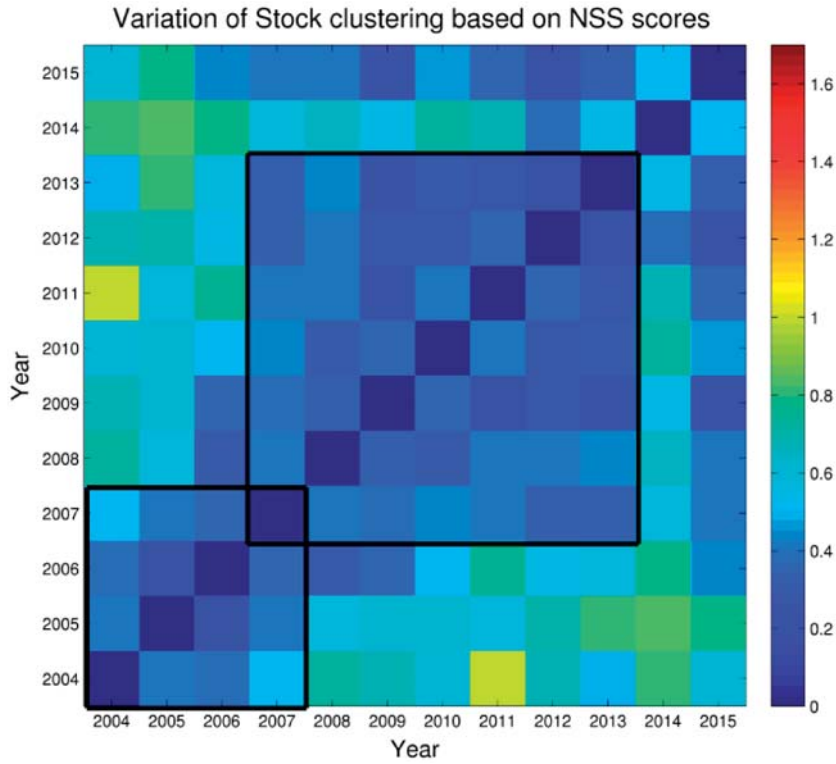

FIG. 13. All-versus-all NSS matrix for 36 stock correlations from 2004 to 2015. The boxes highlight the similarity in the correlations of the stocks till 2007 (small box) and after the 2008 crisis (big box).

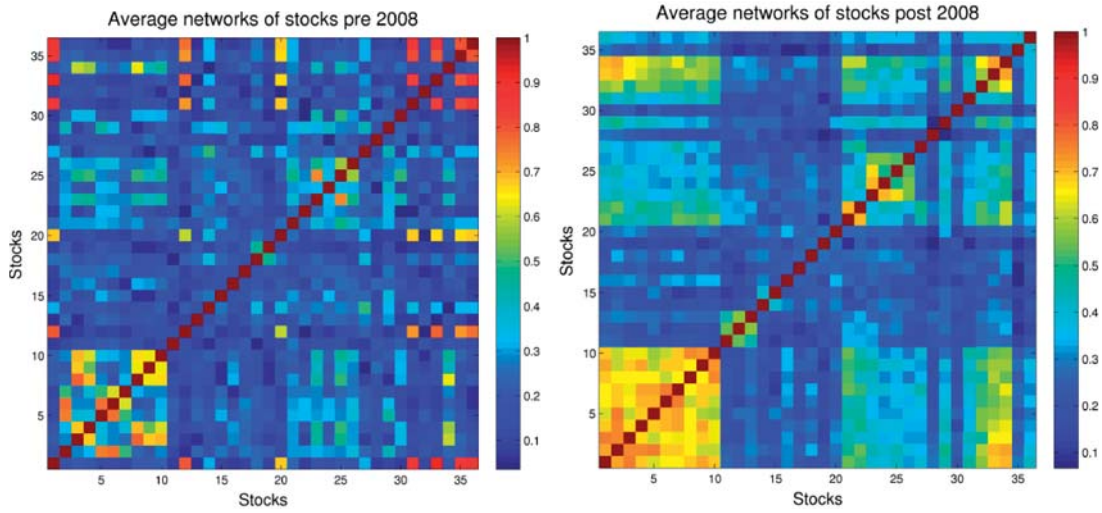

FIG. 14. Average networks of the 36 stocks before and after 2008. The two panels show clearly that the inter-sector connectivity increases post-2008 crisis.

can be achieved by inserting dummy residues at suitable positions, after aligning the proteins either at sequence level or at structural level [76].

Secondly, it should be noted that if two models have similar score with respect to a common template (in this case, the native protein), it does not necessarily need to have similar network. For instance, it is evident from the cases reported in Table 2, where the pairs 2 and 3 have the same side chain score with respect to the native. However, their network differences are highlighted when they are compared against each other. Similarly pairs 1 and 4 have similar backbone score with the native. On comparison with each other, we see that pair 1 has exactly the same backbone network, but pair 4 has different backbone networks. Such occurrences should further be investigated, while analysing large number of networks with a reference. A more discriminating way to do this is to perform an all-versus-all comparison of the networks to find out the underlying network differences. A comparison at this level can become extremely useful in clustering similar networks. This is similar to two-dimensional RMSD in protein structure comparison as opposed to one-dimensional RMSD.

The present network comparison carried out on the normalized Laplacian spectra is well suited to handle weighted networks. Needless to say that investigations based on weighted networks are more rigorous and can provide elegant solutions. Often the edge weights in matrices are converted to binary (0, 1) for the lack of elegant methods to compare large-scale weighted networks. However, the threshold values chosen for binary conversion may be subjective depending on the problem at hand and the total solution to the particular problem may arise only as a composite of solutions at different cut-off values for the edges. Again, in the context of protein side chain networks, a threshold value of high edge weight provides insight into nucleation centres, whereas a low value provides a picture of percolation with global connectivity [77]. The solution of a weighted network efficiently captures all the topological features in a consolidated manner. Further the manner in which a small change in the edge weight is transmitted to the global change has been highlighted by our study on the two models TS008_5 and TS008_3, which differ only in their side chain weights at specific positions.

The different components of the NSS capture different aspects of change in the networks. When the networks are sparse and the node clustering do not vary significantly, as in the case of protein examples in Section 4.1, the EDS component plays the differentiating role (Table 6). For dense networks like financial stocks example in Section 4.2, node clustering changes drastically due to the dynamic changes in the edges. However, these changes might be small when compared to the whole network. In such cases, the CRS and the EWCS components play a major role in differentiating the networks. This is illustrated in Fig. S3 from Supplementary appendix 1.

The performance of the method has been rigorously tested on toy and real-world networks. In Section 3, a controlled change of edge weights has been used to demonstrate the robustness of the method and also the sensitivity of NSS to discriminate iso-spectral networks, which is an added advantage over HIM. Finally, in Section 4.1, we show how NSS is able to give detailed insight about complex protein structure networks which is not possible using the benchmark GDT score. Also, we have provided an example of dynamic change in financial stock market in Section 4.2 which has been discussed in literature using other methods and show that NSS is able to bring out the differences in the networks and is also very useful to interpret the behaviour of the market.

6. Summary

In this article, a scoring method to characterize and compare network patterns, using the solutions to normalized Laplacian of weighted matrices for node clustering has been presented. The novel aspect of this method is that the differences in all possible node clustering levels are used for comparison. The power

of this method is demonstrated by comparing protein structure networks, which is an important problem in structural biology. Various components of the scores provide insights at levels ranging from differences in edges to the transmission of these differences to global clustering levels. The scores developed here not only differentiate networks in comparison to templates but can also serve as a means to cluster networks in a large pool. Although the application is demonstrated in the context of protein structure networks and financial stock networks, the method can be applied to compare any type of complex networks such as those generated from image analyses, micro array data or from biomedical networks.

7. Supplementary data

Supplementary data are available at *COMNET* online.

8. Acknowledgements

SV thanks the National Academy of Sciences, Allahabad, India, for Senior Scientist Fellowship. VG acknowledges University Grants Commission (UGC) for senior research fellowship and SG thanks IISC Mathematics grant. The authors thank Molecular Biophysics Unit (MBU) computation facility. The authors also thank Mr. Ramkumar NSM for insightful discussions regarding stock markets.

References

1. Conte, D., Foggia, P., Sansone, C. & Vento, M. (2004) Thirty years of graph matching in pattern recognition. *Int. J. Pattern Recognit. Artif. Intell.,* **18**, 265–298.
2. Foggia, P., Percannella, G. & Vento, M. (2014) Graph matching and learning in pattern recognition in the last 10 years. *Int. J. Pattern Recognit. Artif. Intell.,* **28**, 1450001.
3. Bron, C. & Kerbosch, J. (1973) Finding all cliques of an undirected graph (algorithm 457). *Commun. ACM,* **16**, 575–576.
4. Cordella, L. P., Foggia, P., Sansone, C., Tortorella, F. & Vento, M. (1998) Graph matching: a fast algorithm and its evaluation. *Proceedings of the Fourteenth International Conference on Pattern Recognition*. IEEE, pp. 1582–1584.
5. Cordella, L. P., Foggia, P., Sansone, C. & Vento, M. (2001) An improved algorithm for matching large graphs. *Third IAPR-TC15 Workshop on Graph-Based Representations in Pattern Recognition*, Brisbane, Australia. Citeseer, pp. 149–159.
6. Ghahraman, D. E., Wong, A. K. & Au, T. (1980) Graph optimal monomorphism algorithms. *IEEE Trans. Syst. Man Cybern.,* **10**, 181–188.
7. Hopcroft, J. E. & Wong, J.-K. (1974) Linear time algorithm for isomorphism of planar graphs (preliminary report). *Proceedings of the Sixth Annual ACM Symposium on Theory of Computing*, Illinois, Chigao, USA. ACM, pp. 172–184.
8. Irniger, C. & Bunke, H. (2001) Graph matching: filtering large databases of graphs using decision trees. *IAPR-TC15 Workshop on Graph-Based Representation in Pattern Recognition*, pp. 239–249.
9. Luks, E. M. (1980) Isomorphism of graphs of bounded valence can be tested in polynomial time. *Twenty-First Annual Symposium on Foundations of Computer Science*. IEEE, pp. 42–49.
10. McGregor, J. J. (1982) Backtrack search algorithms and the maximal common subgraph problem. *Softw. Pract. Exp.,* **12**, 23–34.
11. Messmer, B. T. & Bunke, H. (2000) Efficient subgraph isomorphism detection: a decomposition approach. *IEEE Trans. Knowledge Data Eng.,* **12**, 307–323.
12. Pardalos, P. M., Rappe, J. & Resende, M. G. (1998) An exact parallel algorithm for the maximum clique problem. *High Performance Algorithms and Software in Nonlinear Optimization*, R. De Leone, A. Murli, P. M. Pardalos & G. Toraldo, eds, Springer, pp. 279–300.

13. SHEARER, K., BUNKE, H., VENKATESH, S. & KIERONSKA, D. (1998) *Efficient Graph Matching for Video Indexing*. Vienna: Springer.

14. ULLMANN, J. R. (1976) An algorithm for subgraph isomorphism. *J. ACM,* **23**, 31–42.

15. BRANCA, A., STELLA, E. & DISTANTE, A. (1999) Feature matching by searching maximum clique on high order association graph. *Proceedings of the International Conference on Image Analysis and Processing*. IEEE, pp. 642–658.

16. CORDELLA, L. P., FOGGIA, P., SANSONE, C. & VENTO, M. (1998) *Subgraph Transformations for the Inexact Matching of Attributed Relational Graphs*. Vienna: Springer.

17. FISCHLER, M. A. & ELSCHLAGER, R. A. (1973) The representation and matching of pictorial structures. *IEEE Transactions on Computers*, **22**, 67–92.

18. KITTLER, J. & HANCOCK, E. R. (1989) Combining evidence in probabilistic relaxation. *Int. J. Pattern Recognit. Artif. Intell.,* **3**, 29–51.

19. PELILLO, M. (1995) Relaxation labeling networks for the maximum clique problem. *J. Artif. Neural Netw.,* **2**, 313–328.

20. SANFELIU, A. & FU, K.-S. (1983) A distance measure between attributed relational graphs for pattern recognition. *IEEE Trans. Syst. Man Cybern.*, 353–362.

21. SHAPIRO, L. G. & HARALICK, R. M. (1981) Structural descriptions and inexact matching. *IEEE Trans Pattern Anal. Machine Intell.,* **3**, 504–519.

22. TSAI, W.-H. & FU, K.-S. (1983) Subgraph error-correcting isomorphisms for syntactic pattern recognition. *IEEE Trans. Syst. Man Cybern.*, **13**, 48–62.

23. WONG, A. K., YOU, M. & CHAN, S. (1990) An algorithm for graph optimal monomorphism. *IEEE Trans. Syst. Man Cybern.,* **20**, 628–638.

24. BRAND, M. & HUANG, K. (2003) A unifying theorem for spectral embedding and clustering. *Proceedings of the Ninth International Workshop on Artificial Intelligence and Statistics*, Key West, Florida.

25. BUTLER, S. & CHUNG, F. (2006) Spectral graph theory. *Handbook of Linear Algebra*, L. Hogben, ed, Florida: CRC Press, pp. 47.

26. CARCASSONI, M. & HANCOCK, E. R. (2001) Weighted graph-matching using modal clusters. *Computer Analysis of Images and Patterns*. W. Skarbek, ed, Heidelberg, Berlin: Springer, pp. 142–151.

27. FIEDLER, M. (1973) Algebraic connectivity of graphs. *Czech. Math. J.,* **23**, 298–305.

28. HAGEN, L. & KAHNG, A. B. (1992) New spectral methods for ratio cut partitioning and clustering. *IEEE Trans. Comput. Aided Des. Integr. Circuits Syst.,* **11**, 1074–1085.

29. KOSINOV, S. & CAELLI, T. (2002) Inexact multisubgraph matching using graph eigenspace and clustering models. *Structural, Syntactic, and Statistical Pattern Recognition*, T. Caelli, A. Amin, R. P.W. Duin, D. de Ridder & M. Kamel, eds, Heidelberg, Berlin: Springer, pp. 133–142.

30. LEE, W.-J. & DUIN, R. P. (2008) An inexact graph comparison approach in joint eigenspace. *Structural, Syntactic, and Statistical Pattern Recognition*, N. da Vitoria Lobo, T. Kasparis, F. Roli, J. T. Kwok, M. Georgiopoulos, G. C. Anagnostopoulos & M. Loog, eds, Heidelberg, Berlin: Springer, pp. 35–44.

31. MÉMOLI, F. (2011) A spectral notion of Gromov–Wasserstein distance and related methods. *Appl. Comput. Harmonic Anal.,* **30**, 363–401.

32. SCHAEFFER, S. E. (2007) Graph clustering. *Comput. Sci. Rev.,* **1**, 27–64.

33. SHI, J. & MALIK, J. (2000) Normalized cuts and image segmentation. *IEEE Trans Pattern Anal. Machine Intell.,* **22**, 888–905.

34. VAN LEUKEN, R. H., SYMONOVA, O., VELTKAMP, R. C. & DE AMICIS, R. (2008) Complex Fiedler vectors for shape retrieval. *Structural, Syntactic, and Statistical Pattern Recognition*, N. da Vitoria Lobo, T. Kasparis, F. Roli, J. T. Kwok, M. Georgiopoulos, G. C. Anagnostopoulos & M. Loog, eds, Heidelberg, Berlin: Springer, pp, 167–176.

35. XU, L. & KING, I. (2001) A PCA approach for fast retrieval of structural patterns in attributed graphs. *IEEE Trans. Syst. Man Cybern. B Cybern.,* **31**, 812–817.

36. ZHU, P. & WILSON, R. C. (2005) A study of graph spectra for comparing graphs. proceeding for British Machine Vision Conference, UK. W. Clocksin, A. Fitzgibbon & P. Torr, eds.

37. Banerjee, A. (2012) Structural distance and evolutionary relationship of networks. *BioSystems,* **107**, 186–196.

38. Comellas, F. & Diaz-Lopez, J. (2008) Spectral reconstruction of complex networks. *Phys. A Stat. Mechanics Appl.,* **387**, 6436–6442.

39. Fay, D., Haddadi, H., Moore, A. W., Mortier, R., Uhlig, S. & Jamakovic, A. (2010) A weighted spectrum metric for comparison of internet topologies. *ACM SIGMETRICS Perform. Eval. Rev.,* **37**, 67–72.

40. Ipsen, M. & Mikhailov, A. S. (2002) Evolutionary reconstruction of networks. *Physical Review E*, **66**, 046109.

41. Pincombe, B. (2007) Detecting changes in time series of network graphs using minimum mean squared error and cumulative summation. *ANZIAM J.,* **48**, 450–473.

42. Jurman, G., Visintainer, R., Filosi, M., Riccadonna, S. & Furlanello, C. (2015) The HIM glocal metric and kernel for network comparison and classification. *IEEE International Conference on Data Science and Advanced Analytics (DSAA), Paris. 36678 2015.* IEEE, pp. 1–10.

43. Jurman, G., Visintainer, R. & Furlanello, C. (2010) An introduction to spectral distances in networks (extended version). Preprint arXiv:1005.0103.

44. Butler, S. (2015) Algebraic aspects of the normalized Laplacian. *Recent Trends in Combinatorics.* The IMA Volumes in Mathematics and its Applications, IMA (2016). A. Beveridge, J. R. Griggs, L. Hogben, G. Musiker & P. Tetali, Switzerland: Springer International Publishing. 295–316.

45. Chung, F. R. (1997) *Spectral Graph Theory*, Vol. 92. USA: American Mathematical Society.

46. Von Luxburg, U. (2007) A tutorial on spectral clustering. *Stat. Comput.,* **17**, 395–416.

47. Zumstein, P. (2005) Comparison of spectral methods through the adjacency matrix and the Laplacian of a graph. *Diploma Thesis*, ETH Zürich.

48. Caelli, T. & Caetano, T. (2005) Graphical models for graph matching: approximate models and optimal algorithms. *Pattern Recog. Lett.,* **26**, 339–346.

49. Caelli, T. & Kosinov, S. (2004) An eigenspace projection clustering method for inexact graph matching. *IEEE Trans Pattern Anal. Machine Intell.,* **26**, 515–519.

50. Scott, G. L. & Longuet-Higgins, H. C. (1991) An algorithm for associating the features of two images. *Proc R. Soc. Lond. B Biol. Sci.,* **244**, 21–26.

51. Shapiro, L. S. & Brady, J. M. (1992) Feature-based correspondence: an eigenvector approach. *Image Vision Comput.,* **10**, 283–288.

52. Banerjee, A. & Jost, J. (2008) On the spectrum of the normalized graph Laplacian. *Linear Algebra Appl.,* **428**, 3015–3022.

53. Gould, P. R. (1967) On the geographical interpretation of eigenvalues. *Trans. Inst. Br. Geogr.*, 53–86.

54. Atilgan, A. R., Akan, P. & Baysal, C. (2004) Small-world communication of residues and significance for protein dynamics. *Biophys. J.,* **86**, 85–91.

55. Bagler, G. & Sinha, S. (2005) Network properties of protein structures. *Phys. A Stat. Mechanics Appl.,* **346**, 27–33.

56. Brinda, K. V. & Vishveshwara, S. (2005) A network representation of protein structures: implications for protein stability. *Biophys. J.,* **89**, 4159–4170.

57. Greene, L. H. & Higman, V. A. (2003) Uncovering network systems within protein structures. *J. Mol. Biol.,* **334**, 781–791.

58. Vendruscolo, M., Dokholyan, N. V., Paci, E. & Karplus, M. (2002) Small-world view of the amino acids that play a key role in protein folding. *Physical Review E*, **65**, 061910.

59. Csermely, P., Korcsmáros, T., Kiss, H. J., London, G. & Nussinov, R. (2013) Structure and dynamics of molecular networks: a novel paradigm of drug discovery: a comprehensive review. *Pharmacol. Ther.,* **138**, 333–408.

60. Vendruscolo, M., Paci, E., Dobson, C. M. & Karplus, M. (2001) Three key residues form a critical contact network in a protein folding transition state. *Nature,* **409**, 641–645.

61. Artymiuk, P. J., Spriggs, R. V. & Willett, P. (2005) Graph theoretic methods for the analysis of structural relationships in biological macromolecules. *JASIS,* **56**, 518–528.

62. Bahar, I. (1999) Dynamics of proteins and biomolecular complexes: inferring functional motions from structure. *Rev. Chem. Eng.,* **15**, 319–348.

63. Chatterjee, S., Ghosh, S. & Vishveshwara, S. (2013) Network properties of decoys and CASP predicted models: a comparison with native protein structures. *Mol. Biosyst.,* **9**, 1774–1788.
64. Grindley, H. M., Artymiuk, P. J., Rice, D. W. & Willett, P. (1993) Identification of tertiary structure resemblance in proteins using a maximal common subgraph isomorphism algorithm. *J. Mol. Biol.,* **229**, 707–721.
65. Kannan, N. & Vishveshwara, S. (1999) Identification of side-chain clusters in protein structures by a graph spectral method. *J. Mol. Biol.,* **292**, 441–464.
66. Mitchell, E. M., Artymiuk, P. J., Rice, D. W. & Willett, P. (1990) Use of techniques derived from graph theory to compare secondary structure motifs in proteins. *J. Mol. Biol.,* **212**, 151–166.
67. Przytycka, T., Srinivasan, R. & Rose, G. D. (2002) Recursive domains in proteins. *Protein Sci.,* **11**, 409–417.
68. Thorpe, M., Lei, M., Rader, A., Jacobs, D. J. & Kuhn, L. A. (2001) Protein flexibility and dynamics using constraint theory. *J. Mol. Graphics Model.,* **19**, 60–69.
69. Krishnadev, O., Brinda, K. V. & Vishveshwara, S. (2005) A graph spectral analysis of the structural similarity network of protein chains. *Proteins,* **61**, 152–163.
70. Livi, L., Maiorino, E., Pinna, A., Sadeghian, A., Rizzi, A. & Giuliani, A. (2014) Analysis of heat kernel highlights the strongly modular and heat-preserving structure of proteins. Preprint arXiv:1409.1819.
71. Sistla, R. K., KV, B. & Vishveshwara, S. (2005) Identification of domains and domain interface residues in multidomain proteins from graph spectral method. *Proteins Struct. Funct. Bioinform.,* **59**, 616–626.
72. Bhattacharyya, M., Ghosh, S. & Vishveshwara, S. (2016) Protein structure and function: looking through the network of side-chain interactions. *Curr. Protein Peptide Sci.,* **17**, 4–25.
73. Chakrabarti, G. & Sen, C. (2013) *Momentum Trading on the Indian Stock Market.* India: Springer.
74. Manda, K. (2010) *Stock Market Volatility during the 2008 Financial Crisis*. Citeseer. New York: New York University.
75. Rehan, M. M., Haran, M. H., Chauhan, N. S. & Grover, D. (2013) Visualizing the indian stock market: a complex networks approach. *Int. J. Adv. Eng. Technol.,* **6**, 1348.
76. Bhattacharyya, M., Upadhyay, R. & Vishveshwara, S. (2012) Interaction signatures stabilizing the nad (p)-binding rossmann fold: a structure network approach. *PLoS One,* **7**, e51676.
77. Deb, D., Vishveshwara, S. & Vishveshwara, S. (2009) Understanding protein structure from a percolation perspective. *Biophys. J.* **97**, 1787–1794.