BEYOND ALGEBRAIC CONNECTIVITY OF GRAPHS - EVALUATION OF TOPOLOGY, BASED ON SPECTRAL CLUSTERING

ОТВЪД АЛГЕБРИЧНАТА СВЪРЗАНОСТ НА ГРАФА - СПЕКТРАЛНО КЛЪСТЕРИРАНЕ КАТО ОЦЕНКА НА ТОПОЛОГИЯТА

Dipl.Eng. Mircho Jordanov Mirchev
Faculty of Telecommunications, Technical University of Sofia, 8 Kl. Ohridski Blvd, 1000 Sofia +359 887 412 248, mircho@mirchev.eu

Маг.инж. Мирчо Йорданов Мирчев
Факултет по телекомуникации, Технически университет-София, София 1000, бул."Кл.Охридски" 8
+359 887 412 248, mircho@mirchev.eu

Ключови думи: мрежови топологии, графи, алгебрична свързаност, спектрална теория на графи, вектор на Фидлер

В тази разработка е представен метод за оценка на топологията на графи посредством спектралната теория на графите. Тази теория се занимава с анализ на Айген стойностите и векторите на характеристичните матрици на графите.

Спектралният анализ на един граф, включва изчисление Айген стойностите и Айген векторите на Лапласовата матрица на графа, откъдето се вижда значението на λ_2 *– алгебричната свързаност на графа, както и на вектора на Фидлер, като има съпоставка между неговите елементи и върховете на графа. На база на този анализ, може да се направи предположение за вариант за добавяне на ново ребро в граф, което да дава най-голямо повишение на алгебричната свързаност на графа.*

Въз основа на работата може да се автоматизира процеса на анализ, както и да се направи самообучаващ се алгоритъм за анализ и подобряване на свързаността на графите.

Keywords: network topology, graphs, algebraic connectivity, spectral graph theory, Fiedler vector

This paper presents a method for graph topology evaluation based on the Spectral graph theory, which is based on the analysis of Eigen values and Eigen vectors of the graph matrices.

When doing Spectral graph analysis, the first is to calculate the Eigen values and Eigen vectors of the Laplacian matrix of the graph, which gives the important value λ_2 *–the algebraic connectivity of a graph and the Fiedler vector. This vector has values for each node of the graph. On the base of this analysis, a variant of adding a new edge, which gives the highest gain in the algebraic connectivity, is made.*

Based on this works, a system for automated analysis of graphs and self-learning algorithm for graph analysis and optimization can be made.

INTRODUCTION

Graph theory has always been used to represent in mathematical way real-life scenarios. As each network can be represented as a graph, it has very wide application in the field of telecommunications – to calculate the shortest one or to search for critical edges (bridges) or critical nodes, called articulation points, or evaluate some topological parameters. Until recently most of these tasks were accomplished algorithmically. With the rapid growth of networks, both in size and complexity, the algorithmic methods became too slow to keep up with the network growth. From control perspective, this was addressed with the Software Defined Networks (SDN) ideology, while in evaluation perspective this was addressed by introducing the spectral graph theory, which gives a good evaluation on both local and global topological parameters of the network. The most used measure of "how well the graph is connected" is the algebraic connectivity. It is defined as the second smallest eigenvalue (λ_2) of the Laplacian matrix of the graph[1].

This paper goes beyond the measure of algebraic connectivity and explores the properties of the eigenvector associated with the second smallest eigenvalue (algebraic connectivity). This vector is called *Fiedler vector* and gives a good representation of the components (subgraphs) of the graph, which share similar connectivity properties. The eigenvalues and eigenvectors in linear algebra are defined as follows:

$$(1) \quad A\mathbb{v} = \lambda\mathbb{v}$$

where vector $\mathbb{v}$ is an eigenvector of the matrix A for the eigenvalue λ. From spectral graph theory, the graph spectra of graph G are defined as the spectrum of the adjacency matrix A of G, which is its set of eigenvalues together with their multiplicities. The Laplace spectrum of a finite undirected graph without loops is the spectrum of the Laplace matrix L.

DEFINITIONS

Let $G_c = (V, E)$ be a weighted graph on n vertices in which the edge weights are positive numbers. Denote by $i \sim j$ if the vertices i and j are adjacent and by $w_{i,j}$ the weight of the edge $e_{i,j}$. Let $w_i = \sum_{j \sim i} w_{i,j}$. The Laplacian matrix $L(G_C) = (l_{i,j})$ of G_c is defined as

$$(2) \quad L_{i,j} = \begin{cases} w_i, & \text{if } i = j, \\ -w_{i,j} & \text{if } i \sim j, \\ 0, & \text{otherwise.} \end{cases}$$

Note that G_c is a simple graph whenever $w_{i,j} = 1$ for all $i \sim j$. Denote a simple graph by G.

The $L(G_C)$ is a real symmetric matrix. From this fact and Gershgorin circle theorem, it follows that its eigenvalues are nonnegative real numbers. Moreover, since its rows sum to *0*, *0* is the smallest eigenvalue of $L(G_C)$ and the multiplicity of *0* equals the number of components of G_c. Then we can assume that G_c is a simple connected graph.

Also let T denote the diagonal matrix with *(i,i)*-th entry having w_i. The normalized Laplacian matrix of $G(\mathcal{L})$ is defined to be the matrix:

$$(3) \quad \mathcal{L}_{i,j} = \begin{cases} 1, & \text{if } i = j \text{ and d} \neq 0; \\ -\frac{1}{\sqrt{d_i d_j}}, & \text{if } i \sim j; \\ 0, & \text{otherwise.} \end{cases}$$

We can then write

$$(4) \quad \mathcal{L}_{i,j} = T^{-\frac{1}{2}} L T^{-\frac{1}{2}}$$

with the convention $T^{-1}(i,i) = 0$ for $d_i = 0$

From these definitions we can retrieve the important parameters of graph spectra:

Spectrum of a graph

The spectrum of finite graph G_c is by definition the spectrum of the adjacency matrix *A*, its set of eigenvalues together with their multiplicities. The Laplace spectrum of finite graph G_c is the spectrum of the Laplace matrix L[2].

Since A is real and symmetric, all its eigenvalues are real. Also, for each eigenvalue λ_n, its algebraic multiplicity coincides with its geometric multiplicity. Since *A* has zero diagonal, its trace *tr(A)*, and hence the sum of the eigenvalues is zero.

Similarly, *L* is real and symmetric, so that the Laplace spectrum is real. Moreover, *L* is positive semidefinite and singular, so we can denote the eigenvalues by:

$$(5) \quad \lambda_n \geq \cdots \geq \lambda_2 \geq \lambda_1 = 0$$

The sum of these eigenvalues is *tr(L)*, which is twice the number of edges of G_c. Finally, also $\mathcal{L}$ has real spectrum and nonnegative eigenvalues (but not necessarily singular) and $tr(\mathcal{L})=tr(L)$.

Algebraic connectivity

In [1] the algebraic connectivity $a(G_C)$ of a (connected) graph is defined as the second smallest eigenvalue (λ_2) of the Laplacian matrix of a graph with *n* vertices.

This parameter is used as a generalized measure of "how well is the graph connected" [3]. It has values between *0* and *n* (a fully-connected graph K_n has *n*). This eigenvalue is greater than 0 if and only if G is a connected graph. This is a corollary to the fact that the number of times 0 appears as an eigenvalue in the eigenvector of the Laplacian is the number of connected components in the graph. Therefore, the farther λ_2 is from zero, the more difficult it is to separate a graph into independent components. However, the algebraic connectivity is equal to zero for all disconnected networks. Therefore, as soon as the connectedness is lost, due to failures, for example, this measure becomes less useful as being too coarse.

Fiedler vector

Fiedler vector is the eigenvector associated with the second smallest eigenvalue of the graph G – λ_2. These eigenvectors of the graph Laplacian have been explored extensively recently, mostly in [4–11]. The values of the Fiedler vector provide a good evaluation of the topology and the node significance in terms of global topology, e.g. to find densely connected clusters, to find poorly connected nodes, or to evaluate the global connectivity distribution of a network.[5]

Also, the differences between the values in the Fiedler vector can give an estimation of the distance between the nodes [1], [12].

EXPLORING THE FIEDLER VECTOR

In this chapter, we'll explore a sparsely connected simple graph with 6 nodes show on Figure 1. For this graph we'll calculate the relevant spectral parameters and show how the values in the Fiedler vector identify the worst connected node (it is obvious in this the graph – node 5), the clusters the are formed by the nodes, and also make a proposition where a single new edge should be added to maximize the connectivity and robustness of the graph.

Let's take the graph shown on Figure 1:

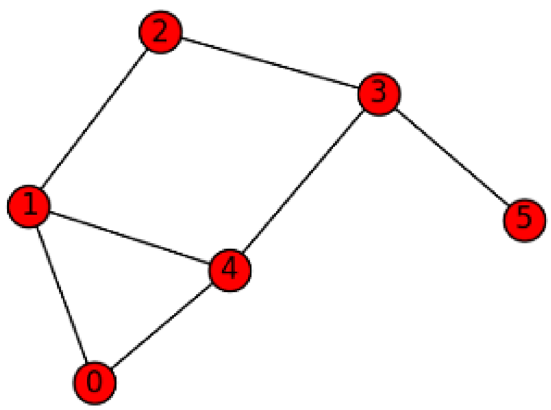

Figure 1 – Example graph G

It has the following matrices:

- Adjacency matrix A:

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix} \quad (6)$$

- Laplacian matrix L:

$$(7)\qquad L=\begin{bmatrix} 2 & -1 & 0 & 0 & -1 & 0\\ -1 & 3 & -1 & 0 & -1 & 0\\ 0 & -1 & 2 & -1 & 0 & 0\\ 0 & 0 & -1 & 3 & -1 & -1\\ -1 & -1 & 0 & -1 & 3 & 0\\ 0 & 0 & 0 & -1 & 0 & 1 \end{bmatrix}$$

For further calculations we'll use the normalized Laplacian matrix [13], [14], which is defined as:

$$(8)\qquad \mathcal{L}_{i,j}=\begin{cases} 1, & \text{if } i=j \text{ and d} \neq 0;\\ -\frac{1}{\sqrt{d_i d_j}}, & \text{if } i\sim j;\\ 0, & \text{otherwise.} \end{cases}$$

but it can be easily derived from the diagonal matrix D by [15]:

$$(9)\qquad \mathcal{L}_{i,j}=D^{-\frac{1}{2}}LD^{-\frac{1}{2}}$$

Eigenvalues and Eigenvectors

For the example graph, we'll use the *Laplacian matrix* from which we'll derive the eigenvalues and the eigenvectors which correspond to them. As we will compare graphs with the same number of nodes (variants of the initial graph), it is easier to use the Laplacian matrix instead of the Normalized Laplacian matrix. The eigenvalues and the corresponding eigenvectors of L and $\mathcal{L}$ are as follows:

- Eigenvalues of Laplacian Matrix

$$(10)\qquad \lambda_n(L)=\begin{vmatrix} 0 & 0.722 & 1.683 & 3 & 3.705 & 4.891 \end{vmatrix}$$

$$(11)\qquad V(L)=\begin{vmatrix} -0.408 & -0.415 & -0.505 & 0.289 & -0.567 & -0.032\\ -0.408 & -0.309 & 0.04 & 0.289 & 0.658 & -0.469\\ -0.408 & -0.069 & 0.759 & 0.289 & -0.205 & 0.356\\ -0.408 & 0.221 & 0.201 & -0.577 & -0.308 & -0.562\\ -0.408 & -0.221 & -0.201 & -0.577 & 0.308 & 0.562\\ -0.408 & 0.794 & -0.294 & 0.289 & 0.114 & 0.144 \end{vmatrix}$$

- Eigenvalues of Normalized Laplacian Matrix

$$(12)\qquad \lambda_n(\mathcal{L})=\begin{vmatrix} 0 & 0.446 & 0.871 & 1.284 & 1.521 & 1.877 \end{vmatrix}$$

$$(13)\qquad V(\mathcal{L})=\begin{vmatrix} -0.378 & -0.441 & 0.317 & 0.353 & -0.661 & -0.013\\ -0.463 & -0.372 & -0.302 & 0.351 & 0.562 & -0.343\\ -0.378 & 0.121 & -0.74 & -0.11 & -0.287 & 0.447\\ -0.463 & 0.537 & 0.068 & -0.274 & -0.195 & -0.616\\ -0.463 & -0.226 & 0.402 & -0.596 & 0.283 & 0.371\\ -0.267 & 0.56 & 0.307 & 0.557 & 0.216 & 0.406 \end{vmatrix}$$

First, we look at the eigenvalue 0 and its eigenvectors. A very elegant result about its multiplicity forms the foundation of spectral clustering. Then we look at the second smallest eigenvalue and the corresponding eigenvector.

Algebraic connectivity

From the calculated eigenvalues and eigenvectors, according to [1], we see that the example graph algebraic connectivity of $\alpha(G) = \lambda_2(L) = 0.722$, respectively normalized algebraic connectivity $\alpha_n(G) = \lambda_2(\mathcal{L}) = 0.446$. This eigenvalue is greater than 0 if and only if G is a connected graph. This is a corollary to the fact that the number of times 0 appears as an eigenvalue in the eigenvector of the Laplacian is the number of connected components in the graph, so we can conclude the example graph G is connected and has one connected component (which is clearly seen in Figure 1).

Fiedler vector

The next interesting thing is the Fiedler vector. According to [12], [16] this vector is the eigenvector associated with λ_2. This vector gives the spectral partitioning of the graph, from which we can derive the subcomponents by their relative connectivity. For the example graph the Fiedler vector is shown on (14) together with the vertices associated with each value:

$$Fv(G) = \begin{vmatrix} -0.415 \\ -0.309 \\ -0.069 \\ 0.221 \\ -0.221 \\ 0.794 \end{vmatrix} \begin{matrix} N_0 \\ N_1 \\ N_2 \\ N_3 \\ N_4 \\ N_5 \end{matrix} \tag{14}$$

Or we can plot it as spectral lines (the Y-axis is only informational and for visual clarity):

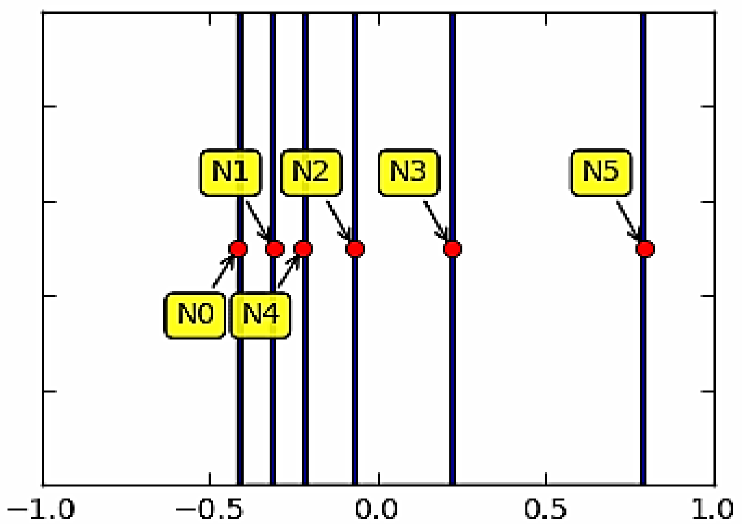

Figure 2 – Laplacian Spectrum of Graph G

Once having the Fiedler vector, we can see if the values, hence the nodes are clustered in spectral space. We apply the DBSCAN scan function from *scikit-learn* package [17] and the result gives the number of clusters: 4, with node/clustering as shown in Table 1

Table 1 – SPECTRAL CLUSTERING OF GRAPH G

Cluster:	Node:	Fiedler vector value
1	N_0	-0.415
1	N_1	-0.309
2	N_2	-0.069
3	N_3	0.221
1	N_4	-0.221
4	N_5	0.794

We clearly see that vertices 0,1,4 form a cluster not only in Laplacian space, Node 2 form a class is its own, Node 3 is little further in the spectrum as it is connected to the least connected node and it is an articulation point itself. Node 5 – the least connected node – is at the edge of the spectrum also [18], [19]. It clearly shows the worse the connectivity of a node – the further from the rest of the nodes in the spectral space [2], [20–22].

ADDING A NEW EDGE

Having done the spectral analysis of the example graph, the next question that arises is about the best place for a new edge to get better connectivity of the graph. For the example graph on Figure 1 all the possible edges that can be added are: (0, 2), (0, 3), (0, 5), (1, 3), (1, 5), (2, 4), (2, 5), (4, 5). For each possible edge, we will compute the Fiedler vector and compare the spectrum of the new graph to the original. In Table 2 are shown the algebraic connectivity together with the Fiedler vector for the original graph, and for each graph, that is constructed with adding one of the possible edges.

Table 2 – SPECTRAL CHARACTERISTICS WHEN ADDING EDGE (X,Y)

Original graph		**Adding Edge (X,Y)**							
		(0, 2)	(0, 3)	(0, 5)	(1, 3)	(1, 5)	(2, 4)	(2, 5)	(4, 5)
Spectral gap $SG_{n,m}$		0.346	0.636	1.209	0.53	1.103	0.152	0.863	1.015
$\boldsymbol{\alpha(G) = \lambda_2(L) =}$	0.722	0.764	0.914	1.586	0.885	1.438	0.731	1	1.382
Avg. SPL	1.667	1.600	1.533	1.467	1.533	1.467	1.600	1.533	1.467
Node:	$\boldsymbol{Fv(G) =}$								
N_0	-0.415	-0.316	-0.307	-0.653	-0.401	-0.702	-0.395	-0.577	-0.602
N_1	-0.309	-0.316	-0.214	-0.271	-0.206	-0.086	-0.313	-0.289	-0.372
N_2	-0.069	-0.195	-0.214	0	-0.097	0.394	-0.189	0.289	0
N_3	0.221	0.195	-0.214	0	0.097	0.308	-0.125	0.289	0.372
N_4	-0.221	-0.195	0.075	0.271	-0.241	-0.308	0.216	-0.289	0
N_5	0.794	0.828	0.873	0.653	0.848	0.394	0.805	0.577	0.602

In [4] is proven that adding a new edge in a graph increases the $\alpha(G)$, more specifically $\alpha(G) \leq \alpha(G') \leq \alpha(G) + 2$. The results in Table 2 show the same.

Also, the results in Table 2 show that the greatest increase of the algebraic connectivity is a result of adding a new edge between the furthermost edges in the spectrum. In the table, this is shown with the value of the spectral gap (or spectral distance), which is defined as the modulus of the difference between the values of Fiedler vector associated with two nodes.

$$\mathrm{SG}_{n,m} = \left|\mathrm{Fv}(G)_{N_n} - \mathrm{Fv}(G)_{N_m}\right| \quad (15)$$

So increasing $\alpha(G)$ via adding one new edge is maximised when the new edge is between the two nodes with the greatest spectral gap, while the average shortest path length is minimized.

CONCLUSION AND FUTURE WORKS

In this paper, the spectral characteristics of graphs are reviewed and through an example, it is shown how these characteristics can be used when taking decisions where to add new edges and how this can affect the overall spectrum of graphs. This can be proven mathematically, parts of the proof can be found in [4], [7], [23–26]. Also, future works can extend the study beyond known sparse graphs to random graphs, described with Erdős–Rényi model [27–30] and also to weighed graphs [31–33].

Based on this work a system for automated analysis of graphs and self-learning algorithm for graph analysis and optimization can be made which can also take into account the clustering properties of nodes.

REFERENCES

[1] M. Fiedler, “Algebraic connectivity of graphs,” *Czechoslovak mathematical journal*, vol. 23, no. 2, pp. 298–305, 1973.

[2] A. E. Brouwer and W. H. Haemers, *Spectra of graphs*. Springer Science & Business Media, 2011.

[3] J. Segovia Silvero and others, “Robustness against large-scale failures in communications networks,” 2011.

[4] N. M. M. de Abreu, “Old and new results on algebraic connectivity of graphs,” *Linear algebra and its applications*, vol. 423, no. 1, pp. 53–73, 2007.

[5] A. Bertrand and M. Moonen, “Distributed computation of the Fiedler vector with application to topology inference in ad hoc networks,” *Signal Processing*, vol. 93, no. 5, pp. 1106–1117, 2013.

[6] R. Byrne, J. Feddema, and C. Abdallah, “Algebraic connectivity and graph robustness,” *Sandia National Laboratories, Albuquerque, New Mexico*, vol. 87185, 2005.

[7] S. Fallat and S. Kirkland, “Extremizing algebraic connectivity subject to graph theoretic constraints,” *Electronic Journal of Linear Algebra*, vol. 3, no. 1, p. 7, 1998.

[8] A. Jamakovic and S. Uhlig, “On the relationship between the algebraic connectivity and graph’s robustness to node and link failures,” in *Next Generation Internet Networks, 3rd EuroNGI Conference on*, 2007, pp. 96–102.

[9] G. Jiang, G. Yu, and J. Cao, “The Least Algebraic Connectivity of Graphs,” *Discrete Dynamics in Nature and Society*, 2015.

[10] S. Kirkland, I. Rocha, and V. Trevisan, “Algebraic connectivity of k-connected graphs,” *Czechoslovak Mathematical Journal*, vol. 65, no. 1, pp. 219–236, 2015.

[11] P. Orponen and S. E. Schaeffer, “Local Clustering of Large Graphs by Approximate Fiedler Vectors [Extended Abstract].”
[12] M. Fiedler, “A property of eigenvectors of nonnegative symmetric matrices and its application to graph theory,” *Czechoslovak Mathematical Journal*, vol. 25, no. 4, pp. 619–633, 1975.
[13] F. R. Chung, *Spectral graph theory*, vol. 92. American Mathematical Soc., 1997.
[14] F. R. Chung, “Eigenvalues of graphs,” in *Proceedings of the International Congress of Mathematicians*, 1995, pp. 1333–1342.
[15] F. J. Hall, “The adjacency matrix, standard Laplacian, and normalized Laplacian, and some eigenvalue interlacing results,” *Department of Mathematics and Statistics, Georgia State University, Atlanta*, 2010.
[16] R. Merris, “Laplacian graph eigenvectors,” *Linear algebra and its applications*, vol. 278, no. 1, pp. 221–236, 1998.
[17] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, and others, “Scikit-learn: Machine learning in Python,” *Journal of Machine Learning Research*, vol. 12, no. Oct, pp. 2825–2830, 2011.
[18] D. Faya, H. Haddadib, S. Uhligc, L. Kilmartind, A. W. Mooreb, J. Kunegise, and M. Iliofotouf, “Network clustering via spectral projections.”
[19] J. Gallier, “Notes on Elementary Spectral Graph Theory. Applications to Graph Clustering Using Normalized Cuts,” *arXiv preprint arXiv:1311.2492*, 2013.
[20] C. J. Alpert, A. B. Kahng, and S.-Z. Yao, “Spectral partitioning with multiple eigenvectors,” *Discrete Applied Mathematics*, vol. 90, no. 1, pp. 3–26, 1999.
[21] S. P. Borgatti and M. G. Everett, “A graph-theoretic perspective on centrality,” *Social networks*, vol. 28, no. 4, pp. 466–484, 2006.
[22] H. Haddadi, M. Rio, G. Iannaccone, A. Moore, and R. Mortier, “Network topologies: inference, modeling, and generation,” *Communications Surveys & Tutorials, IEEE*, vol. 10, no. 2, pp. 48–69, 2008.
[23] K. Ahmat, “Graph Theory and Optimization Problems for Very Large Networks,” *arXiv preprint arXiv:0907.3099*, 2009.
[24] N. Alon, “Eigenvalues and expanders,” *Combinatorica*, vol. 6, no. 2, pp. 83–96, 1986.
[25] A. Berman and X.-D. Zhang, “Lower bounds for the eigenvalues of Laplacian matrices,” *Linear Algebra and its Applications*, vol. 316, no. 1, pp. 13–20, 2000.
[26] L. Donetti, F. Neri, and M. A. Munoz, “Optimal network topologies: Expanders, Cages, Ramanujan graphs, Entangled networks and all that,” *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2006, no. 08, p. P08007, 2006.
[27] F. Chung and M. Radcliffe, “On the spectra of general random graphs,” *the electronic journal of combinatorics*, vol. 18, no. 1, p. P215, 2011.
[28] P. Di Lorenzo and S. Barbarossa, “Distributed Estimation and Control of Algebraic Connectivity Over Random Graphs,” *Signal Processing, IEEE Transactions on*, vol. 62, no. 21, pp. 5615–5628, 2014.
[29] D. Fernholz and V. Ramachandran, “The diameter of sparse random graphs,” *Random Structures and Algorithms*, vol. 31, no. 4, pp. 482–516, 2007.
[30] F. McSherry, “Spectral partitioning of random graphs,” in *Foundations of Computer Science, 2001. Proceedings. 42nd IEEE Symposium on*, 2001, pp. 529–537.
[31] H. Amini and M. Lelarge, “The diameter of weighted random graphs,” *arXiv preprint arXiv:1112.6330*, 2011.
[32] D. Fay, H. Haddadi, A. Thomason, A. W. Moore, R. Mortier, A. Jamakovic, S. Uhlig, and M. Rio, “Weighted spectral distribution for internet topology analysis: theory and applications,” *Networking, IEEE/ACM Transactions on*, vol. 18, no. 1, pp. 164–176, 2010.
[33] D. Fay, H. Haddadi, S. Uhlig, A. W. Moore, R. Mortier, and A. Jamakovic, “Weighted spectral distribution,” *University of Cambridge, Computer Laboratory, Tech. Rep. UCAM-CL-TR-729*, 2008.