Accepted Manuscript

GFT centrality: A new node importance measure for complex networks

Rahul Singh, Abhishek Chakraborty, B.S. Manoj

PII: S0378-4371(17)30659-3
DOI: http://dx.doi.org/10.1016/j.physa.2017.06.018
Reference: PHYSA 18403

To appear in: *Physica A*

Received date : 5 September 2015
Revised date : 19 June 2017

Please cite this article as: R. Singh, A. Chakraborty, B.S. Manoj, GFT centrality: A new node importance measure for complex networks, *Physica A* (2017), http://dx.doi.org/10.1016/j.physa.2017.06.018

GFT Centrality: A New Node Importance Measure for Complex Networks

Rahul Singh

Department of Electrical and Computer Engineering, Iowa State University, Ames, IA 50011 USA

Abhishek Chakraborty, B. S. Manoj*

Department of Avionics, Indian Institute of Space Science and Technology, Thiruvananthapuram, Kerala 695547 India

Abstract

Identifying central nodes is very crucial to design efficient communication networks or to recognize key individuals of a social network. In this paper, we introduce Graph Fourier Transform Centrality (GFT-C), a metric that incorporates local as well as global characteristics of a node, to quantify the importance of a node in a complex network. GFT-C of a reference node in a network is estimated from the GFT coefficients derived from the importance signal of the reference node. Our study reveals the superiority of GFT-C over traditional centralities such as degree centrality, betweenness centrality, closeness centrality, eigenvector centrality, and Google PageRank centrality, in the context of various arbitrary and real-world networks with different degree-degree correlations.

Keywords: Graph Fourier Transform, node importance, GFT-C, centrality measures, graph signal processing, complex networks

*Corresponding author

Email addresses: `rahul.s.in@ieee.org` (Rahul Singh), `abhishek2003slg@ieee.org` (Abhishek Chakraborty), `bsmanoj@ieee.org` (B. S. Manoj)

This work was carried out when the first author was a graduate student at the Indian Institute of Space science and Technology (IIST), Trivandrum, India.

1. Introduction

A wide range of social, biological, and technological phenomena can be described by complex networks [1, 2, 3]. Examples include social networks, world wide web, the Internet, citation networks, collaboration networks, cellular networks, and protein-protein interaction networks [3]. Measure of importance or centrality is fundamental to understand the structural and dynamic properties of complex networks. A few traditional centrality measures such as degree, betweenness, closeness [4], eigenvector [5], and PageRank [6] are defined for unweighted networks, however, the measures have also been extended to weighted networks [6, 7, 8, 9].

Degree Centrality (DC) [9] of a node is defined as the sum of the edge weights incident on that node. Betweenness Centrality (BC) of a node i is measured by counting the presence of node i in the formation of the shortest paths among all source and destination node pairs in a network. In particular, BC of node i in an N node network[3] is defined as $\frac{2}{(N-1)(N-2)}\sum_{i\neq j\neq k}\frac{g_{jk}(i)}{g_{jk}}$, where g_{jk} is the total number of the shortest paths from node j to node k, and $g_{jk}(i)$ is the number of paths between j and k that pass through node i. BC measures gatekeeping and control of information in a network [4]. If a node is not appearing in any of the shortest path formation, BC of that node becomes zero. Closeness Centrality (CC) of a node measures how close the node is to the entire network. The CC value for node i in an N node network is $\frac{N-1}{\sum_j d(i,\ j)}$, where $d(i,\ j)$ is the shortest distance between nodes i and j. A node with high CC score can most efficiently contact other nodes in the network [4]. Evaluation of Eigenvector Centrality (EC) is based on the idea that the most central node is connected to more powerful nodes [10]. EC of a node i is the i^{th} entry of the eigenvector corresponding to the largest eigenvalue of the weighted adjacency matrix. On the other hand, Google PageRank (PR) [6] which ranks websites based on the search engine results is a popular centrality metric to measure relative importance of a node within a node-set. PR assigns a numerical weighting to each link-set of a node recursively based on the number of incoming links and their PR scores.

We introduce a new metric, Graph Fourier Transform Centrality (GFT-C) to quantify how important a particular node is to other nodes in a network. Graph Fourier Transform (GFT) [11], which is an important tool in graph

[3] In this paper, "network" and "graph" are used interchangeably.

signal processing to analyze data residing on graphs, is used for estimating the GFT-C score. The frequency information encoded in the GFT coefficients is the global measure of change in the values of a graph signal as one moves from one node to it's adjacent nodes. GFT-C of i^{th} node is the weighted sum of the GFT coefficients of importance signal corresponding to node i. The importance signal for a reference node i is a graph signal which gives the individual view of the rest of the nodes about the reference node i in terms of the minimum cost to reach i. Smoothness of the importance signal is the key in quantifying the importance of the corresponding node. GFT is used to capture the variations in the importance signal globally, which in turn is utilized to define GFT-C. Thus, GFT-C utilizes not only the local properties, but also the global properties of a network topology. We show the utility of GFT-C by providing experimental results for arbitrary as well as real-world networks. We also compare the performance of GFT-C with DC, BC, CC, EC, and PR in the context of various real-world networks.

Rest of this paper is organized as follows. In Section 2, we present the required preliminaries in the context of GFT. This is followed by Section 3 where we discuss GFT-C in detail. Then we compare our GFT-C performance, in Section 4, with respect to various centrality measures such as degree centrality, betweenness centrality, closeness centrality, eigenvector centrality, and Google PageRank centrality. In Section 5, the behavior of GFT-C, in various networks of differing degree-degree correlations, is thoroughly discussed. We conclude our paper in Section 6.

2. GFT Preliminaries

Graph Signal Processing (GSP) is an emerging field which considers modeling, representation, and processing of signals defined on graphs. In classical signal processing, the structure on which data lies is a regular 1-D grid (e.g., speech signal) or a 2-D grid (e.g., image signal), whereas, graph signal processing deals with signals lying on irregular graphs [11, 12, 13].

A graph $\mathcal{G}$ is represented using the following three parameters: a set of vertices $\mathcal{V}$, a set of edges $\mathcal{E}$, and a weight matrix $\mathbf{W}$. The weight matrix $\mathbf{W}$ is a symmetric matrix (for undirected graphs) whose value at i^{th} row and j^{th} column represents weight corresponding to the edge connecting nodes i and j.

A signal supported on a graph, also known as graph signal, is represented as an N-dimensional vector $\mathbf{f} = [f(1),\ f(2), \dots,\ f(N)]^T$, where $f(i)$ represents value of the graph signal at i^{th} node and $N = |\mathcal{V}|$ is the total number of nodes in the network. To analyze graphs as well as graph signals, graph Laplacian is a useful mathematical tool.

Graph Laplacian [14] is given by $\mathbf{L} = \mathbf{D} - \mathbf{W}$, where $\mathbf{D}$ is the degree matrix and $\mathbf{W}$ is the weight matrix. Degree matrix $\mathbf{D}$ is a diagonal matrix whose non-zero entries represent degree of corresponding nodes. For undirected and connected graphs, the graph Laplacian is a symmetric and positive semi-definite matrix [14]. The eigenvalues of $\mathbf{L}$ in ascending order are $0 = \lambda_0 < \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_{N-1}$. The set of eigenvalues of $\mathbf{L}$ is known as spectrum of the graph which provides the notion of frequency. The graph Laplacian constitutes complete set of orthogonal eigenvectors [14]. We represent these eigenvectors as $[\mathbf{u_0}\ \mathbf{u_1} \ldots \mathbf{u_{N-1}}]$. The eigenvectors corresponding to large eigenvalues change rapidly as one moves from one node to the other, whereas, the eigenvectors corresponding to small eigenvalues constitute very small changes as one moves from a node to the adjacent node. Thus, large eigenvalues correspond to high frequencies and small eigenvalues represent low frequencies.

Analogous to the classical Fourier transform, the Graph Fourier Transform (GFT) [12] is a tool for performing harmonic analysis of graph signals. The eigenvectors of the graph Laplacian act as graph harmonics similar to sinusoids in the classical Fourier transform. Let $\mathbf{U} = [\mathbf{u_0}\ \mathbf{u_1} \ldots \mathbf{u_{N-1}}]$ be the matrix whose columns are the eigenvectors of $\mathbf{L}$, then GFT of the graph signal $\mathbf{f}$ is defined as $\hat{\mathbf{f}} = \mathbf{U}^T\mathbf{f}$. In the transformed vector $\hat{\mathbf{f}} = [\hat{f}(\lambda_0)\ \hat{f}(\lambda_1) \ldots \hat{f}(\lambda_{N-1})]^T$, $\hat{f}(\lambda_\ell) = \langle \mathbf{f}, \mathbf{u}_\ell \rangle$ is the GFT coefficient corresponding to the eigenvalue λ_ℓ. Inverse GFT can be evaluated as $\mathbf{f} = \mathbf{U}\hat{\mathbf{f}}$. We call the set of GFT coefficients of a graph signal as spectrum of the graph signal.

GFT extracts information from the frequency contents, of signals defined on a graph, which cannot be evident in the spatial domain. However, the frequency information encoded in the GFT coefficients is the global measure of change in the values of a graph signal as one moves from one node to the adjacent nodes. Hence, the ability of GFT to capture the global change is the key for accurate identification of influential nodes when GFT-C is concerned.

3. Our Work

In this paper, we propose a spectral approach for assessing the importance of each node in a complex network. We propose Graph Fourier Transform Centrality (GFT-C) which uses GFT coefficients of an importance signal corresponding to the reference node. Our method relies on the global smoothness (or variations) of the carefully defined importance signal corresponding to a reference node. The importance signal for a reference node is the indicator of how remaining nodes in a network are seeing the reference node individually. The importance signal is defined in such a way that the importance information is captured in the global smoothness (or variations) of the signal. Further, GFT coefficients of the importance signal are used to obtain the global view of the reference node.

3.1. The Importance Signal

The importance signal describes the relation of a reference node to the rest of the nodes in a network. The importance signal is characterized by inverse of the cost to reach from an individual node to the reference node. Higher the cost to reach the reference node, lower is the importance of the reference node and vice-versa.

Let the importance signal, a graph signal, corresponding to reference node n on a connected weighted network be $\mathbf{f}_n = [f_n(1)\ f_n(2) \ldots f_n(N)]^T$, where $f_n(i)$ is the inverse of the sum of weights in the shortest path from node i to node n. We normalize the signal such that the sum of signal values, except at the reference node, is unity, i.e., $\sum_{i \neq n} f_n(i) = 1$. We consider the signal value at reference node as unity ($f_n(n) = 1$). Note that any other function which ensures an inversely proportional relationships with the importance signal values, can be used to model GFT-C. However, our selection of the inverse of the path cost is simple and serves the purpose well.

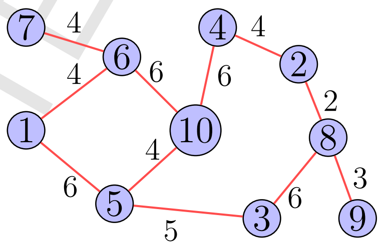


Figure 1: An example weighted graph.

Figure 1 shows a sample weighted graph with 10 nodes. Importance signal along with the intermediate parameters corresponding to node 1 of Figure 1 are listed in Table 1. It is evident that as the cost to reach the reference node decreases, the value of the importance signal increases.

Table 1
Importance Signal for node 1 as Reference

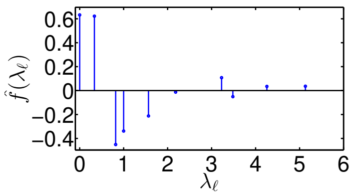
Node	**Shortest Path**	**Cost**	**(Cost)$^{-1}$**	**Importance Signal**
1	-	0	-	1
2	2-8-3-5-1	19	1/19	0.0550
3	3-5-1	11	1/11	0.0950
4	4-10-6-1	16	1/16	0.0653
5	5-1	6	1/6	0.1742
6	6-1	4	1/4	0.2614
7	7-6-1	8	1/8	0.1307
8	8-3-5-1	17	1/17	0.0615
9	9-8-3-5-1	20	1/20	0.0523
10	10-5-1	10	1/10	0.1045

The variations in the importance signal defined above are utilized to score the importance of a reference node. Also note that, for a reference node with high global importance, the importance signal values in the neighborhood of the node is smaller as compared to a reference node with low global importance and the same value of DC. The reason behind this is that the sum of the importance signal values except at the reference node is unity and the importance signal value is inversely related to the cost to reach the reference node. Therefore, for a reference node with high global importance from which the rest of the nodes are not very distant, the distribution of importance signal values is such that the neighboring nodes of the reference node receive small values. On the other hand, when the reference node is of low global importance, the neighboring nodes receive high importance signal values because there are some nodes in the network that are too far from the reference node. These variations in the importance signal, which are captured using GFT, are utilized to quantify the importance of the reference node.

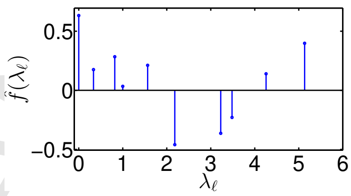
3.2. Graph Fourier Transform Centrality

To define Graph Fourier Transform Centrality (GFT-C) for a reference node, we use GFT coefficients of importance signal defined in Section 3.1. GFT-C measures the importance of a reference node for rest of the network nodes collectively.

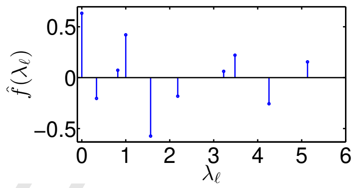
We call GFT of an importance signal as importance spectrum. In Figure 2, we present importance spectra for a few nodes in the network shown in Figure 1. The importance spectra corresponding to node 7 has nominal high frequency components (see Figure 2(a)), whereas, large high frequency components are present in the importance spectra of node 10 (see Figure 2(b)). However, the importance spectra corresponding to node 3 has moderate GFT coefficients corresponding to the high frequencies (see Figure 2(c)). From these observations, we find that the importance information is encoded in the high frequency components of the importance spectrum. The observations from Figure 2 are quite intuitive and follow from the fact that if a node is central to a network, then corresponding importance signal is non-smooth, i.e., variations in the importance signal are high. Based on these observations, we quantify the importance of a reference node using the importance spectrum.



(a) Importance spectrum (node 7).



(b) Importance spectrum (node 10).



(c) Importance spectrum (node 3).

Figure 2: Importance spectra (the network topology is same as Figure 1).

Let GFT values of the importance signal corresponding to reference node n be $\hat{\mathbf{f}}_n = [\hat{f}_n(\lambda_0)\ \hat{f}_n(\lambda_1) \dots \hat{f}_n(\lambda_{N-1})]^T$, which can be calculated as

$$\hat{f}_n(\lambda_\ell) = \sum_{i=1}^{N} f_n(i) u_\ell^*(i), \tag{1}$$

where $f_n(i)$ is the importance of the reference node n with respect to node i. We denote GFT-C of node n as I_n, which is given as

$$I_n = \sum_{\ell=0}^{N-1} w(\lambda_\ell) |\hat{f}_n(\lambda_\ell)|, \tag{2}$$

where $w(\lambda_\ell)$ is the weight assigned to the GFT coefficient corresponding to frequency λ_ℓ. The choice of weights is made by a function which increases exponentially with the frequency (eigenvalues of $\mathbf{L}$), i.e., $w(\lambda_\ell) = e^{(k\lambda_\ell)} - 1$, where $k > 0$. Choosing such weights ensure that: (a) larger weights are assigned to high frequency components of the importance spectrum and smaller weights are assigned to the frequency components corresponding to lower frequencies, and (b) zero weight is assigned to the zero frequency component. It is found experimentally that $k = 0.1$ shows good results. For large values of k, the high frequency GFT coefficients dominate and GFT-C closely resembles EC. This is because of the fact that for large k, the highest frequency component dominates and which is nothing but the eigenvector corresponding to the largest eigenvalue.

Table 2 shows GFT-C scores of the nodes, as found by using Equation (2), in the network shown in Figure 1. We observe that node 10 has the highest score and node 9 has the lowest score (shown in bold font in Table 2).

Table 2
GFT-C for the Network shown in Figure 1

Node	1	2	3	4	5	6	7	8	9	10
GFT-C	0.099	0.093	0.103	0.106	0.121	0.135	0.044	0.118	**0.040**	**0.141**

4. Performance Evaluation of GFT-C

In this section, we estimate GFT-C performance in the context of arbitrary as well as real-world networks and compare with Degree Centrality (DC), Betweenness Centrality (BC), Closeness Centrality (CC), Eigenvector Centrality (EC), and PageRank Centrality (PR). MatlabBGL [15] package is used for conducting the simulation studies. We normalize the sum of GFT-C as well as EC and PR to unity.

4.1. Arbitrary Networks

Here we present the results for three arbitrary networks: a path graph and two unweighted graphs (unweighted graph - I and unweighted graph - II) which contain multiple neighbor nodes with high influence values. The experimental results show the superiority of our GFT-C over other centrality measures.

4.1.1. A Path Graph

First, we consider a path graph (i.e., a string topology network) which is shown in Figure 3. Various centrality scores are listed in Table 3, from which we observe that GFT-C scores for end nodes are low, and as we go towards the middle nodes, the GFT-C scores increase[4]. GFT-C also shows superiority over DC as all the nodes except two end nodes have same DC scores of 2, which is not desirable. Moreover, it can be observed that PR assigns more importance to nodes B and E than nodes C and D. However, BC, CC, and EC follow the same pattern as GFT-C which identifies nodes B and C as the most influential nodes in the network.

Figure 3: A path graph consisting of six nodes.

Table 3
Various Centrality Scores for the Path Graph of Figure 3

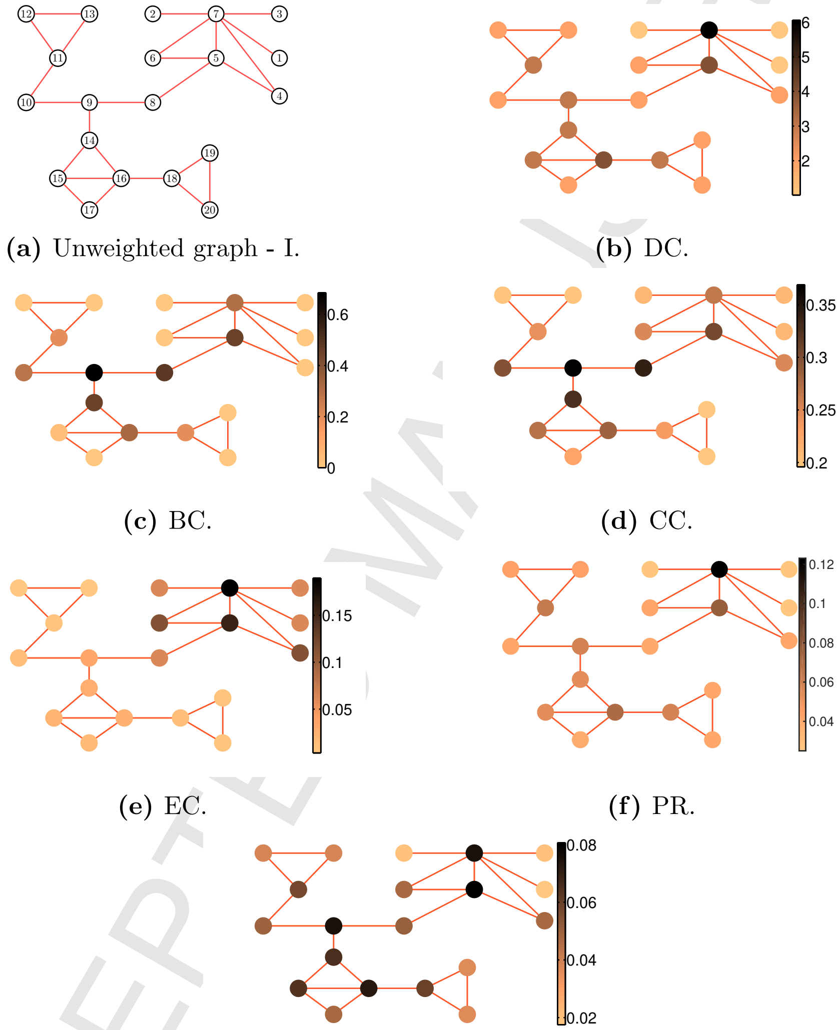
Node	**A**	**B**	**C**	**D**	**E**	**F**
DC	1	**2**	**2**	**2**	**2**	1
BC	0	0.4	**0.6**	**0.6**	0.4	0
CC	0.333	0.454	**0.555**	**0.555**	0.454	0.333
EC	0.099	0.178	**0.223**	**0.223**	0.178	0.099
PR	0.110	**0.199**	0.191	0.191	**0.199**	0.110
GFT-C	0.0936	0.1979	**0.2085**	**0.2085**	0.1979	0.0936

4.1.2. Unweighted Graph - I

Next an unweighted graph, as shown in Figure 4(a), is considered with multiple neighborhood nodes with high influences. Various centrality scores are shown in Figures 4(b)-(g) where the centrality measures are color coded such that high color intensity represents a high value of centrality at the corresponding node. Table 4 lists six centrality scores for all nodes of Figure 4(a).

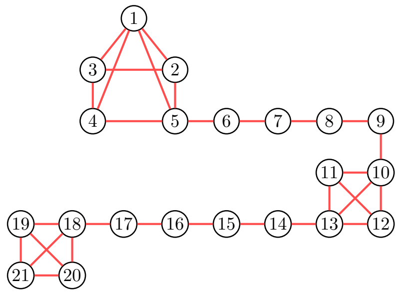
From Table 4, it can be observed that various centrality measures identify different nodes as the important nodes in the graph. For example, nodes 7, 5, and 16 (in descending order) are identified as the most important nodes when

[4]In Table 3, first two highest influential nodes are shown in bold font. The same representation is carried out, to show the first two highest influential nodes, for other networks where individual node influence is tabulated.

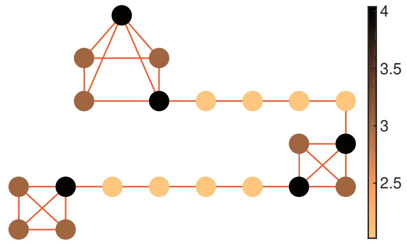


(a) Unweighted graph - I. **(b)** DC. **(c)** BC. **(d)** CC. **(e)** EC. **(f)** PR. **(g)** GFT-C.

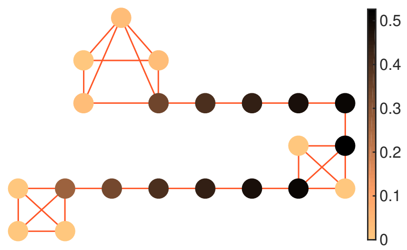
Figure 4: Various centrality measures for the unweighted graph - I.



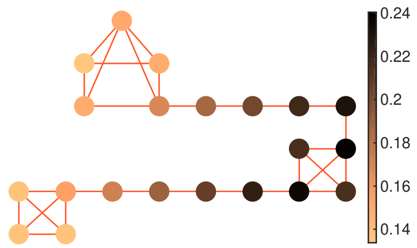
(a) Unweighted graph - II.



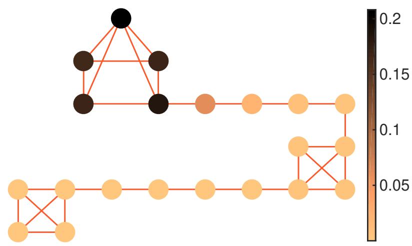
(b) DC.



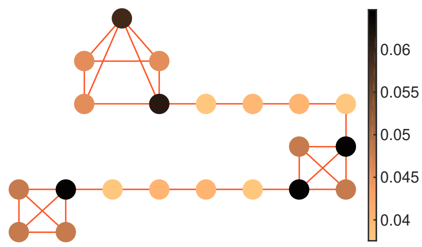
(c) BC.



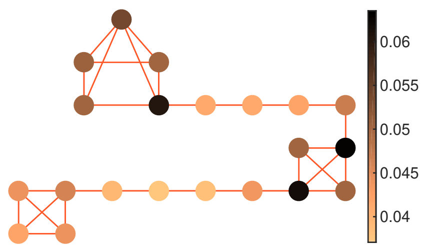
(d) CC.



(e) EC.



(f) PR.



(g) GFT-C.

Figure 5: Various centrality measures for the unweighted graph - II.

DC is concerned (see Figure 4(b)). On the other hand, BC and CC based centrality measures find nodes 9 and 8 as the most influential nodes (see Figures 4(c) and 4(d)). However, as BC and CC capture only global influence, they fail to recognize the importance of nodes 5, 7, and 16 which are influential within their respective neighborhood. EC detects nodes 7 and 5 as the most influential, whereas, due to it's nature of focusing to a particular region or a community (mostly the largest) in a network [16], EC fails to capture importance of other nodes in the network (see Figure 4(e)). As shown in Figure 4(f), PR also detects nodes 7 and 5 as the most influential, however, does not capture the global importance of nodes 9 and 14 in the network.

Our GFT-C, on the contrary, takes care of both local and global influences of all nodes of Figure 4(a) and identifies nodes 5, 7, 9, and 16 as the most important nodes in the unweighted graph (see Figure 4(g)).

Table 4

Various Centrality Scores for the Unweighted Graph - I of Figure 4(a)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
DC	1	1	1	2	**4**	2	**6**	2	3	2	3	2	2	3	3	**4**	2	3	2	2
BC ($\times 10^{-3}$)	0	0	0	0	459	0	301	**491**	**678**	280	199	0	0	456	41	321	0	199	0	0
CC ($\times 10^{-3}$)	211	211	211	250	306	250	264	**339**	**365**	297	244	200	200	328	271	284	226	235	194	194
EC ($\times 10^{-3}$)	62	62	62	114	**158**	114	**188**	63	34	13	6	3	3	27	23	26	16	13	6	6
PR ($\times 10^{-3}$)	25	25	25	41	**77**	41	**122**	40	57	41	62	44	44	54	54	70	38	58	42	42
GFT-C ($\times 10^{-3}$)	18	20	20	48	**80**	48	**75**	51	**75**	50	58	39	39	66	64	**72**	48	60	37	37

4.1.3. Unweighted Graph - II

We consider another arbitrary unweighted network as shown in Figure 5(a). Various centrality scores are also shown in Figures 5(b)-(g) where the centrality measures are color coded such that high color intensity represents a high value of centrality at the corresponding node. Table 5 lists six centrality scores for all nodes of Figure 5(a).

From Table 5, it can be observed that various centrality measures identify different nodes as important nodes when unweighted graph - II is concerned. For example, nodes 1, 5, 10, 13, and 18 are recognized as the most influential nodes when DC is concerned (see Figure 5(b)). On the other hand, BC and CC based centrality measures find nodes 10, 13, and 9 (in descending order) as the most dominant nodes (see Figures 5(c) and 5(d)). Furthermore, BC also discovers other nodes, which form path like topology, also influential. However, both BC and CC fail to recognize the importance of nodes 5 and 18 which are very predominant in the corresponding neighborhoods. On the other hand, due to its characteristic feature of identifying communities, EC wrongly identifies node 1 as the most crucial node and fails to estimate influential nodes in the unweighted graph - II (see Figure 5(e)). As shown in Figure 5(f), PR detects nodes 10, 13, 18, 5, and 1 (in descending order) as the most influential nodes. Although node 10 has more global importance than node 18, PR ranks nodes 10 and 18 equally. However, Our GFT-C takes care of both local and global influences of all nodes in the unweighted network - II and identifies nodes 10, 13, and 5 as the most important nodes (see Figure 5(g)).

Table 5
Various Centrality Scores for the Unweighted Graph - II of Figure 5(a)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
DC	**4**	3	3	3	**4**	2	2	2	2	**4**	3	3	**4**	2	2	2	2	**4**	3	3	3
BC ($\times 10^{-3}$)	32	30	18	30	339	335	442	479	**505**	**521**	0	0	**505**	479	442	395	337	268	0	0	0
CC ($\times 10^{-3}$)	149	148	132	148	168	185	202	217	**230**	**238**	215	215	**235**	222	206	189	171	154	136	136	136
EC ($\times 10^{-3}$)	**206**	**167**	164	**167**	**182**	62	21	7	4	5	4	4	4	1	0	0	0	0	0	0	0
PR ($\times 10^{-3}$)	**59**	46	46	46	**61**	37	40	40	38	**64**	48	48	**64**	38	41	41	38	**64**	48	48	48
GFT-C ($\times 10^{-3}$)	**54**	50	51	50	**61**	41	41	42	47	**63**	50	50	**62**	44	38	37	39	**46**	44	44	42

4.2. Real World Networks

In this paper, we also consider two real world networks: Zachary's karate club network [17] and the network of interlocking directorships in the Netherlands [18], and compare GFT-C performance with other centrality metrics.

4.2.1. Zachary's Karate Club Network

Zachary's karate club network [17] is a social network of friendships between 34 members of a karate club at a US university in the 1970s. Due to certain conflict between the administrator and the instructor of the club, members of the club split into two groups. Graphical representation of the network is depicted in Figure 6. The karate club network is a weighted undirected graph where individuals are represented as nodes and a weighted link between two nodes represents the interaction strength between respective individuals. It can be observed from the figure that, even after splitting, the administrator (node 1 of Figure 6 is shown in *Red colored circle*) and the instructor (node 34 of Figure 6 is also shown in *Red colored circle*) remain central in the network.

The scores of various centralities, in the context of Zachary's karate club network, are listed in Table 6. As expected, GFT-C for node 1 (the administrator) and node 34 (the instructor) receive high scores. However, high value of BC and CC for node 20 is not desirable. High values of GFT-C for nodes 3

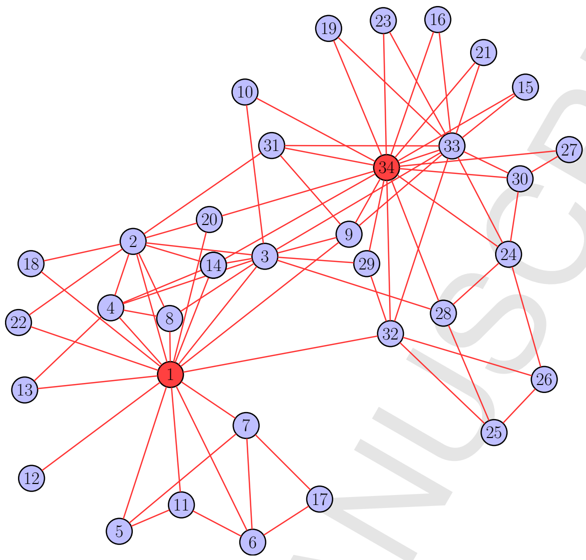


Figure 6: Graphical representation of Zachary's karate club network.

and 33 result from the fact that, along with high DC values, these nodes are connected to the highly influential nodes 1 and 34, respectively. EC and PR follow almost the same trend as GFT-C. Thus, it can be found that only EC, PR, and GFT-C give desirable results when importance of a node in the Zachary's karate club network is concerned.

Table 6
Various Centrality Measures for Zachary's Karate Club Network

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
DC	**42**	29	33	18	8	14	13	13	17	3	8	3	4	17	5	7	6
BC ($\times 10^{-3}$)	**474**	64	69	3	1	29	29	0	25	14	1	0	0	2	0	0	0
CC ($\times 10^{-3}$)	**254**	200	196	177	153	152	154	182	199	191	176	146	205	191	171	138	109
EC ($\times 10^{-3}$)	**67**	65	**77**	43	12	14	14	38	53	11	10	9	9	51	17	24	4
PR($\times 10^{-3}$)	**89**	57	**63**	37	20	34	32	26	33	10	20	10	11	33	13	16	17
GFT-C ($\times 10^{-3}$)	**114**	53	**83**	18	9	13	12	15	38	6	6	6	12	30	9	17	5

	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34
DC	3	3	5	4	4	5	21	7	14	6	13	6	13	11	21	38	**48**
BC ($\times 10^{-3}$)	31	6	**241**	0	0	0	2	64	1	0	12	19	0	6	126	72	**397**
CC ($\times 10^{-3}$)	193	188	**248**	204	177	159	139	158	124	169	156	203	175	174	209	200	**252**
EC ($\times 10^{-3}$)	9	11	16	13	12	17	47	11	26	13	32	18	29	34	45	**71**	**78**
PR ($\times 10^{-3}$)	10	10	13	11	11	13	41	17	29	15	27	15	28	23	42	**76**	**97**
GFT-C ($\times 10^{-3}$)	7	11	14	19	4	10	44	7	24	6	25	6	21	17	50	**118**	**171**

4.2.2. Network of Interlocking Directorships in the Netherlands

We also study different centrality measures for the network of interlocking directorships in the Netherlands [18]. It is a network of interlocking directorships among 16 corporations for the year of 1976. Here, *interlocking directorship* refers to the situation where a member of a board of directors of one corporation also serves as the member in another board of directors. The edge-weight in the network represents the number of interlocks. The weight matrix representing the network is shown in Table 7.

Table 7
Number of Interlocks in the Dutch Network of Directors

Corporation	**Sector**	ABN	AMRO	ENNIA	NS	Buhrmann-T	AGO	AKZO	NB	SHV	FGH	Heineken	Philips	Nat. Ned.	OGEM	RSV	NSU
ABN	Banking	-	0	0	1	2	1	2	1	1	1	2	1	4	0	0	3
AMRO	Banking	0	-	3	2	1	2	1	2	2	0	3	1	2	1	2	0
ENNIA	Insurance	0	3	-	3	1	0	1	0	1	0	0	0	0	1	1	0
NS	Railways	1	2	3	-	0	0	1	1	2	0	0	0	1	0	2	1
Buhrmann-T	Paper	2	1	1	0	-	0	1	0	0	1	0	0	0	1	0	0
AGO	Insurance	1	2	0	0	0	-	0	2	1	0	1	1	0	0	0	3
AKZO	Chemical	2	1	1	1	1	0	-	1	2	1	0	1	1	0	2	0
NB	Nat. Banking	1	2	0	1	0	2	1	-	1	0	1	1	1	0	0	1
SHV	Wholesale	1	2	1	2	0	1	2	1	-	0	0	0	1	0	1	1
FGH	Mortgage	1	0	0	0	1	0	1	0	0	-	0	1	0	1	0	0
Heineken	Beer	2	3	0	0	0	1	0	1	0	0	-	1	0	1	1	0
Philips	Electronic	1	1	0	0	0	1	1	1	0	1	1	-	1	0	1	0
Nat. Ned.	Insurance	4	2	0	1	0	0	1	1	1	0	0	1	-	0	0	2
OGEM	Engineering	0	1	1	0	1	0	0	0	0	1	1	0	0	-	1	0
RSV	Shipbuilding	0	2	1	2	0	0	2	0	1	0	1	1	0	1	-	1
NSU	Shipping	3	0	0	1	0	3	0	1	1	0	0	0	2	0	1	-

Values of different centrality measures for the network are listed in Table 8. The top two corporations based on the GFT-C scores are AMRO and ABN which belong to the banking industry. High values of GFT-C for banking corporations agree with the analysis that the network of interlocks derives its structure mainly from the interlocks of financial corporations and institutions, particularly from those of commercial banks. For this network, GFT-C, PR, EC, and DC provide desirable result. On the contrary, BC and CC scores of Philips is the highest, which shows the failure of BC and CC to identify influential nodes in this particular network.

Table 8
Various Centrality Measures for the Network of Directors

Corporation	DC	BC	CC	EC	PR	GFT-C
ABN	**19**	0.047	0.625	**0.089**	**0.096**	**0.096**
AMRO	**22**	0.015	0.555	**0.104**	**0.109**	**0.112**
ENNIA	11	0.035	0.600	0.062	0.059	0.073
NS	14	0.021	0.577	0.079	0.071	0.064
Buhrmann-T	7	0.013	0.555	0.035	0.042	0.032
AGO	11	0.008	0.555	0.061	0.058	0.069
AKZO	14	0.107	0.682	0.071	0.072	0.077
NB	12	0.068	0.682	0.067	0.062	0.063
SHV	13	0.080	0.652	0.075	0.066	0.067
FGH	5	0.030	0.577	0.020	0.033	0.014
Heineken	10	0.033	0.600	0.057	0.054	0.057
Philips	9	**0.115**	**0.714**	0.046	0.050	0.034
Nat. Ned.	13	0.009	0.577	0.078	0.066	0.086
OGEM	6	0.051	0.577	0.026	0.038	0.020
RSV	12	0.059	0.625	0.063	0.063	0.067
NSU	12	0.010	0.517	0.067	0.062	0.069

5. Observations and Discussion

From the previous section, we observe that GFT-C performs better when identifying node influence in various arbitrary and real-world complex networks are concerned. In other words, our GFT-C metric can efficiently detect more central nodes in a complex network and thus helps effectively comprehending the structure and dynamics of the network.

In this section, we discuss on how our GFT-C behaves when certain statistical measures such as degree-degree correlations, in arbitrary as well as real-world complex networks, is taken into account.

5.1. Observations on Degree-Degree Correlations

Degree-degree correlation [19], which represents the relationship of a connected node-pair, plays major role to determine the structure of a network. Based on the degree-degree correlation, networks can be classified into two regimes: assortative and disassortative networks.

In case of assortative network, a node is keen to attach with another node when they have similar degree. Therefore, in assortative mixing, a hub node [20] which can accommodate unprecedented amount of connections, tend to connect to another hub node. Similarly, nodes with small degree also try to make connection with small degree nodes. On the other hand, disassortative networks show different structural behavior where a node with low degree is inclined to create a link with another node with high degree.

It can be observed that most of the existing centrality measures fail to detect types of degree-degree correlation when assortative as well as disassortative networks are concerned [21, 22]. In order to identify the degree-degree correlation of a network, we used Pearson Correlation Coefficient (PCC) for an undirected network (η) [23]. PCC of an undirected network can be estimated as:

$$\eta = \frac{\sum_{mn} mn(e_{mn} - p_m p_n)}{\sigma_p^2}. \tag{3}$$

In Equation (3), e_{mn} is the fraction of links that connect the nodes with degrees m and n. Moreover, $p_m = \sum_n e_{mn}$ and $p_n = \sum_m e_{mn}$, and σ_p is the standard deviation of the distribution p_n. Note that $-1 \leqslant \eta \leqslant 1$ where $\eta > 0$ indicates assortative mixing and $\eta < 0$ represents disassortative mixing in a network. PCC for all the networks, considered in this paper, are listed in Table 9.

Table 9
Pearson Correlation Coefficients for all networks

Networks	**Pearson Correlation Coefficient** (η)
Path Graph (Figure 3)	-0.2500
Unweighted Graph - I (Figure 4(a))	-0.5013
Unweighted Graph - II (Figure 5(a))	**0.2164**
Zachary's Karate Club Network (Figure 6)	-0.4756
Network of Interlocking Directorships (Table 7)	-0.1570

It can be seen from Table 9 that only unweighted graph - II (Figure5(a)) is an assortative network. All other networks, discussed in this paper, are resembling disassortative mixing.

In the following, we also compare our GFT-C performance with other centrality metrics such as DC, BC, CC, EC, and PR, in the light of Pair-wise Spearman's Rank Correlation Coefficients.

5.2. Observations on Pair-wise Spearman's Rank Correlation Coefficients

To quantify the similarities (or dissimilarities) between GFT-C and other centrality measures, we use Pair-wise Spearman's Rank Correlation Coefficients (PSRCC) [24]. Let $(x_1,\ x_2, \ldots,\ x_n)$ and $(y_1,\ y_2, \ldots,\ y_n)$ be two datasets. If r_{x_i} represents the rank of x_i among i samples, for $i = 1,\ 2, \ldots,\ n$, then PSRCC value between the two datasets is defined as $\rho = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)}$ [24], where $d_i = r_{x_i} - r_{y_i}$. Note that the value of PSRCC lies in the interval $[-1,\ 1]$. A high value of rank correlation coefficient indicates high similarity between the rankings of the two datasets and vice-versa.

Table 10 shows PSRCC scores between GFT-C and other centrality measures for various arbitrary and real-world networks considered in this paper. In case of the path graph (see Figure 3), GFT-C has the same rank value when BC, CC, and EC scores are concerned. However, as PR identifies nodes B and E instead of nodes C and D as influential nodes, as can be seen from Table 3, PSRCC score of PR with respect to GFT-C is low for the path graph. In case of the unweighted graph - I (see Figure 4(a)), GFT-C has high rank similarity with DC and BC, however, shows very low rank similarity with EC (see Table 4). On the other hand, the unweighted graph - II (see Figure 5(a)) shows high rank similarity with only DC, PR, and EC scores.

Table 10
Pair-wise Spearman's Rank Correlation Coefficients between GFT-C and other Centrality Measures

Networks \ PSRCC w.r.t. GFT-C	DC	BC	CC	EC	PR
Path Graph (Figure 3)	0.8281	**0.9562**	**0.9562**	**0.9562**	0.4781
Unweighted Graph - I (Figure 4(a))	**0.9280**	**0.8616**	0.8029	0.2614	**0.8078**
Unweighted Graph - II (Figure 5(a))	**0.8277**	-0.0609	0.0807	**0.4924**	**0.7371**
Zachary's Karate Club Network (Figure 6)	**0.8130**	0.4858	0.3651	**0.8720**	**0.7873**
Network of Interlocking Directorships (Table 7)	**0.8278**	-0.1971	-0.0521	**0.8088**	**0.8206**

In Zachary's karate club network (see Table 6), the coefficient between GFT-C and EC is the highest, indicating high degree of correlation between GFT-C and EC rankings. Moreover, GFT-C correlation with DC and PR are also high in the context of Zachary's karate club network. However, CC value could not properly identify the influential nodes. In case of the network of interlocking directorships, due to high degree of correlation between the pairs, high PSRCC ranking can be seen between GFT-C and DC, PR, and EC (see Table 8). However, with respect to GFT-C, the negative correlation coefficients for CC and BC indicate high dissimilarity between the rankings when network of interlocking directorship is concerned.

Therefore, GFT-C provides better results for the networks, such as arbitrary and real-world, where local as well as global properties are desirable.

6. Conclusion

In this paper, we proposed Graph Fourier Transform Centrality (GFT-C), a new node importance measure for networks. GFT-C considers local (from importance signal) as well as global (from importance spectrum) properties of a node in a complex network. Our observations demonstrated the utility of GFT-C, in identifying influential nodes, for arbitrary as well as real-world networks with different degree-degree correlations. We compared the performance of GFT-C with various centrality measures, e.g., degree, betweenness, closeness, eigenvector, and Google PageRank centralities, and demonstrated the superiority of GFT-C in the context of many real-world networks.

References

[1] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, D.-U. Hwang, Complex networks: Structure and dynamics, Physics Reports 424 (2006) 175–308.

[2] M. E. Newman, The structure and function of complex networks, SIAM Review 45 (2003) 167–256.

[3] R. Albert, A.-L. Barabási, Statistical mechanics of complex networks, Reviews of Modern Physics 74 (2002) 47.

[4] L. C. Freeman, Centrality in social networks conceptual clarification, Social Networks 1 (1979) 215–239.

[5] J. M. Kleinberg, Authoritative sources in a hyperlinked environment, Journal of the ACM 46 (1999) 604–632.

[6] S. Brin, L. Page, The anatomy of a large-scale hypertextual web search engine, Computer Networks and ISDN Systems 30 (1998) 107–117.

[7] M. E. Newman, Analysis of weighted networks, Physical Review E 70 (2004) 056131.

[8] M. E. Newman, Scientific collaboration networks. ii. shortest paths, weighted networks, and centrality, Physical Review E 64 (2001) 016132.

[9] T. Opsahl, F. Agneessens, J. Skvoretz, Node centrality in weighted networks: Generalizing degree and shortest paths, Social Networks 32 (2010) 245–251.

[10] P. Bonacich, Power and centrality: A family of measures, American Journal of Sociology (1987) 1170–1182.

[11] D. K. Hammond, P. Vandergheynst, R. Gribonval, Wavelets on graphs via spectral graph theory, Applied and Computational Harmonic Analysis 30 (2011) 129–150.

[12] D. I. Shuman, S. K. Narang, P. Frossard, A. Ortega, P. Vandergheynst, The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains, IEEE Signal Processing Magazine 30 (2013) 83–98.

[13] S. K. Narang, A. Ortega, Perfect reconstruction two-channel wavelet filter banks for graph structured data, IEEE Transactions on Signal Processing 60 (2012) 2786–2799.

[14] F. R. Chung, Spectral graph theory, volume 92, American Mathematical Society, 1997.

[15] D. Gleich, Matlabbgl [online], `http://www.stanford.edu/dgleich/programs/matlab_bgl/`, (accessed April 23, 2015).

[16] M. U. Ilyas, H. Radha, A klt-inspired node centrality for identifying influential neighborhoods in graphs, in: 44th Annual Conference on Information Sciences and Systems (CISS), IEEE, 2010, pp. 1–7.

[17] W. W. Zachary, An information flow model for conflict and fission in small groups, Journal of Anthropological Research (1977) 452–473.

[18] F. N. Stokman, F. W. Wasseur, D. Elsas, The dutch network: types of interlocks and network structure, Order 501 (1985) 6396.

[19] R. Pastor-Satorras, A. Vázquez, A. Vespignani, Dynamical and correlation properties of the internet, Physical Review Letters 87 (2001) 258701.

[20] A.-L. Barabási, R. Albert, Emergence of scaling in random networks, Science 286 (1999) 509–512.

[21] R. Pastor-Satorras, C. Castellano, Distinct types of eigenvector localization in networks, Scientific Reports 6 (2016).

[22] S. Jalan, A. Kumar, A. Zaikin, J. Kurths, Interplay of degree correlations and cluster synchronization, Physical Review E 94 (2016) 062202.

[23] M. E. Newman, Mixing patterns in networks, Physical Review E 67 (2003) 026126.

[24] Y. Dodge, The concise encyclopedia of statistics, Springer Science & Business Media, 2008.

Research highlights:

- A new centrality measure, Graph Fourier Transform Centrality (GFT-C), is proposed.
- GFT-C utilizes Graph Fourier Transform (GFT) of a carefully defined importance signal on the network.
- GFT-C captures local as well as global influence of a node in a complex network.
- Demonstrated superiority of GFT-C over degree, betweenness, closeness, eigenvector, and Google PageRank centralities for arbitrary as well as real-world networks.