Journal of Engineering and Applied Sciences 14 (2): 526-533, 2019
ISSN: 1816-949X

Investigation on the Use of Graph Signal Processing for an Intelligent Taxis Transportation System to Study Human Activites

[1]Ali Khalaf Nawar Al-Attabi and [2]Jidong Huang
[1]Department of Electrical Engineering, College of Engineering, University of Wasit, Kut, Iraq
[2]Department of Electrical Engineering, College of Engineering and Computer Science,
California State University, Fullerton, California

Abstract: This study demonstrates the benefits of using Graph Signal Processing (GSP) techniques for an intelligent taxis transportation system. Graph signal processing, an application arising to handle multiple source signals on a graph, has developed into an active field of research during the last several years due to its ability to analyze enormous datasets or dynamic data that usually pose a challenge to researchers. We introduce a possible method of using graph signal processing and its operations to analyze signals in a network of taxi stand locations where the taxis can be sensors for human activities. An example is given using real data of taxi's and stand's locations in San Francisco where the number of taxis around these stands is the detected signal. The results showed the effectiveness of using graph Fourier transform to detect the anomalies in the signals which represent unusual transportation activities for human or driver distributions within the taxi network.

Key words: Graph signal processing, analysis huge data, an intelligent taxis transportation system, multiple source, dynamic data, transportation activities

INTRODUCTION

A public transport system plays a significant role in offering comfortable and flexible service in urban cities. One of the commonly used public transport systems is based on the use of taxis. In the past; many taxi companies depended on the drivers' experience to search for passengers because of the inherent randomness in the taxi service system. This method caused many problems. One of them has a significant effect on the taxi service ever since it existed and it happens when taxis are waiting at a vacant stand while customers may be waiting in vain elsewhere (Meng *et al*., 2010). Now a days, taxi companies equip the taxis with GPS (Global Position System) receivers to track taxis via. wireless communication between the taxis and a control center. Thus, a network of taxis on the road can be tracked and located. This technique will locate the taxis to each of the taxi requests received to reduce the waiting time problem (Wang *et al*., 2014). To provide good services, researchers have been focused on the taxi's mobility patterns that are essential in traffic modeling and forecasting (Leutzbach, 1988). Also, analysis of the collective mobility through taxi use data serves as a tool to detect peoples activities (Liao *et al*., 2010; Candia *et al*., 2008). Taxis data analysis including their locations and occupancies over a period of days, months or years is a challenge, especially for big cities. The widely-gathered data will be able to show the distribution of the taxis, the mobility of humans and their activities. The analyzed information can be used to improve the service quality in reducing pick-up time and finding the best allocation for the taxis per customer's requests.

In the last few years, the emerging field of processing signals on graphs has made important advances in the analysis and processing of the large data where the graph refers to a combinatorial structure of vertices and edges (Shuman *et al*., 2013). The most significant development had been made by deriving a spectral framework for analyzing signals; it is equivalent to how the Fourier transform allows one to decompose complex signals in terms of their fundamental frequencies. The spectral transform defines graph signals depending on their relation to the spectral properties of the underlying graph. This method has led to many new algorithms such as the graph-based filtering and denoising methods (Zhang and Hancock, 2008; Susnjara *et al*., 2015). Now a days, processing a signal on a graph is prominently used in the field of sensor networks due to continuous streams of data from sensors with very high spatial and temporal resolution (Loukas, 2015).

Corresponding Author: Ali Khalaf Nawar Al-Attabi, Department of Electrical Engineering, College of Engineering, University of Wasit, Kut, Iraq

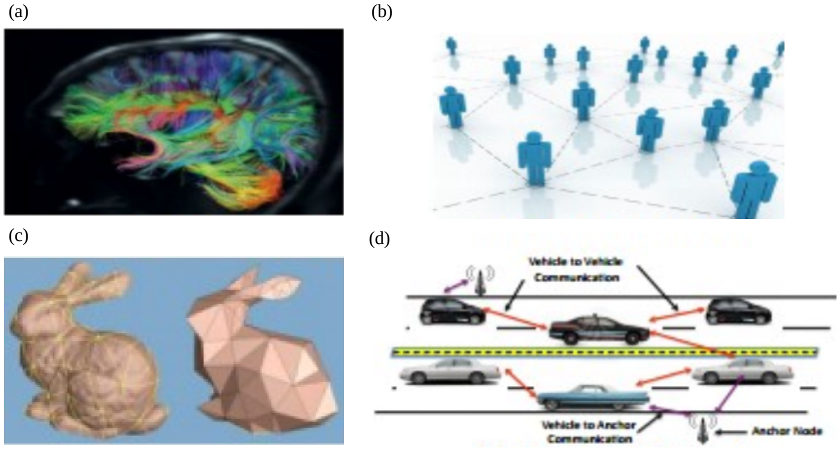
(a)
(b)
(c)
(d)
Vehicle to Vehicle Communication
Vehicle to Anchor Communication
Anchor Node

Fig. 1: An example of some real-world graph structured data. The values on the nodes will be the signals: a) Neuronal networks; b) Social networks; c) Computer graphics and d) Vehicular networks

MATERIALS AND METHODS

Graph theory: The graph is a mathematical structure that demonstrates a set of objects which are related to each other. The objects are shown by vertices or nodes and the relations are represented by edges or links that interconnect the vertices. Therefore, graphs are useful for describing the geometric structures of data domains in many applications such as social network, energy, sensor and neuronal networks. A graph can be undirected, if for each pair of connected nodes, there is no origin node and destiny node (the edge connects the nodes bidirectionally) or the graph can be directed: edges go from one node to another node. In many applications, it is suitable to use undirected graphs. However, there are certain applications that inherently require a directed graph representation. The vertices and the edges vary depending on the application of interest. For instance in social network graphs, the vertices correspond to the users and links are present between users if they share a social or a relationship (Venkatesan and Kannan, 2013). Figure 1 demonstrates some examples of processing data on graphs.

Formally, a graph, G of size N is represented by G = {V, E, W} and defined as a graph, consists of a finite set of vertices V = {v_i}, a finite set of edges E = {e_{ij}} and an unweighted or a weighted adjacency matrix W. In an unweighted graph, the W_{ij} = 1 if there is a relationship between two vertices i and j, otherwise, W_{ij} = 0. In a weighted graph, the related weight with each of the edges measures the strength of the relation between the corresponding nodes. If there is an edge e_{ij} connecting two vertices i and j, the W_{ij} represents the weight of this edge; otherwise, W_{ij} = 0 (6). For example in sensor network graphs, the weights are inversely proportional to the physical distance between the sensor nodes and reflect the correlation between the sensor signals at those nodes. The weight matrix of a weighted graph is a N×N matrix. For the rest of this study, only undirected weighted graphs are considered. There is also, another matrix of interest in GSP called the Laplacian matrix L. L defined in Eq. 1 where D is the diagonal degree matrix. The vector of degrees is denoted by d since, it is the diagonal elements of D; also, each component of d is the degree of the corresponding vertex $d_j = \Sigma_{i\neq j} W_{ij}$ as shown in Fig. 2.

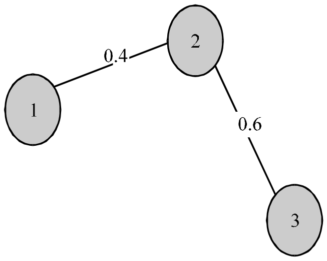

Fig. 2: An example of a graph consists of 3 nodes with two edges. It is represented by the weighted matrix and the degree matrix

$$W = \begin{bmatrix} 0 & 0.4 & 0 \\ 0.4 & 0 & 0.6 \\ 0 & 0.6 & 0 \end{bmatrix} D = \begin{bmatrix} 0.4 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0.6 \end{bmatrix} \quad (1)$$

As the graph Laplacian (L) is a real symmealic matrix, it has a complete set of orthonormal Eigenvectors. This matrix will be very useful for the spectral analysis of the graph as shown in the next section. From the Laplacian matrix, the normalized Laplacian matrix, L_{norm} obtained using the degree matrix D as shown in Eq. 2. Using the normalized Laplacian matrix will make the frequency analysis easier by normalizing the frequency range from 0-2. L and L_{norm} are not similar matrices, so, their Eigenvectors and Eigenvalues are different. Furthermore, selection of a suitable Laplacian matrix for a particular graph-based problem will depend on the application:

$$L_{norm} = D^{-\frac{1}{p}} L D^{-\frac{1}{p}} \quad (2)$$

For any an undirected graph, the Laplacian matrix is symmetric and positive definite. This means the Eigenvectors will be orthogonal and the Eigenvalues will be real and non-negative by applying a singular value decomposition to the matrix as shown in Eq. 3:

$$L = U\Lambda U^T \quad (3)$$

where, Λ is a diagonal matrix of non-negative real Eigenvalues. The columns of U are the Eigenvectors {u0, u2, ..., u_{N-1}} corresponding to the ordered Eigenvalues $0 = \lambda_0 \leq \lambda_1 \leq \lambda_2 \leq \lambda_{N-1}$, constitute an orthonormal basis for $\mathbb{R}^N$ (Shuman *et al*., 2016).

Graph fourier transform: Mathematically, the Fourier transform with respect to a set of operators is the expansion of a signal into a basis of the operator's eigenfunctions. In classical signal processing, Discrete Fourier Transform (DFT) has been one of the most important tools for analyzing signals. The DFT is an alternative basis representation of signals in the time-domain. The set of basis vectors, which decompose the given signal is the Fourier basis. It has many interesting properties that can be exploited for signal analysis. Many of the existing signal processing techniques for both the time and the image signals depend on the DFT representation of the signal. This motivates researchers to derive a set of basis vectors similar to the Fourier basis for graph signals. Having basis vectors that are analogous to the Fourier basis will capture notions of high and low frequencies on graphs similar to sinusoids in the time domain. A low-frequency graph signal would be one that varies very slowly with respect to its neighbors. A high-frequency signal would be one that varies significantly with respect to the adjacent nodes. Also, the basis vectors should be invariant to the node ordering. The research has focused particularly on the properties of Laplacian matrix and its Eigenvectors. Interestingly, these Eigenvectors are analogous to sinusoids in the time domain in that they have a natural signal-frequency interpretation. The spectral decomposition theorem guarantees the existence of an orthonormal matrix U that diagonalizes L (Venkatesan and Kannan, 2013).

The definition of the graph fourier transform as shown in Eq. 4 is mentioned in many research papers such as Sandryhaila and Moura (2014a, b), Shuman *et al.* (2013). The Eigenvectors of the graph Laplacian are used to find the Graph Fourier Transform (GFT) for the vector f∈$\mathbb{R}^N$ that is showing the observed signals at each vertex on the graph. The GFT is analogous to the classical Fourier transform, given by Eq. 5, for the signals in the time domain. Furthermore, the Inverse Graph Fourier Transform (IGFT) is defined by Eq. 6:

$$\hat{f}(\lambda_i) = \langle f, U_i \rangle = \sum_{j=1}^{N} f(j) U_i^*(j) \quad (4)$$

$$\hat{f}(\xi) = \langle f, e^{2\pi i\xi t} \rangle = \int_{\mathbb{R}} f(t) e^{-2\pi i\xi t} dt \quad (5)$$

$$f(j) = \sum_{j=1}^{N} \hat{f}(\lambda_i) U_i(j) \quad (6)$$

Given a node ordering, each element of an Eigenvector can be associated with a corresponding node of the graph. For connected graphs, the Laplacian eigenvector u_0 associated with the Eigenvalue λ_0 is constant and equal to 1/√N at each vertex. Thus, u_0 does not change its value across nodes and hence, it is like a DC signal on a graph. The graph Laplacian Eigenvectors that are associated with low frequencies λ_i will cause the slow fluctuation across the graph. For example, having two connected vertices by an edge with a small weight shows that the values of the signal at those locations are likely to be not similar. On the other hand, the Eigenvectors that are associated with larger Eigenvalues vary more rapidly where the connected vertices by an edge with high weight are more likely to have similar values. Therefore, research has been conducted to study the effectiveness of using graph Fourier transform in the detection of anomalies on measuring signals within a receiver network (Mahyari and Aviyente, 2014; Huang and Siliang, 2016).

In summary, the Eigenvalues and the Eigenvectors carry the notion of frequency whereas λ_0 shows constant values in our signal on the graph and λ_{N-1} represents maximum variations in the signal on the graph.

Smoothness on graph: One of the important properties in analyzing a signal is the smoothness with respect to the intrinsic structure of the data domain which is represented by the weight matrix in graph signal processing. For continuous signals, differential geometry provides tools

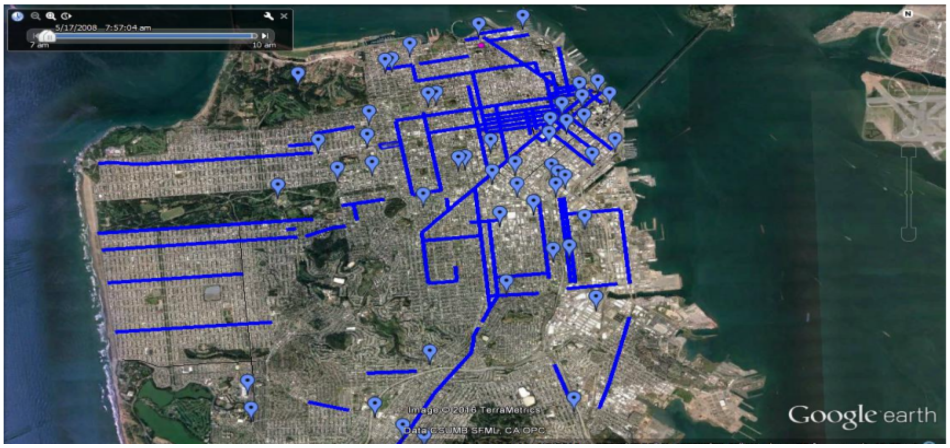

Fig. 3: The fFifty taxi cabs stand locations in San Francisco (Anonymous, 2016)

to incorporate the geometric structure of the underlying manifold into the analysis of the signal on differentiable manifolds. Discrete calculus provides "a set of definitions and differential operators that make it possible to operate the machinery of multivariate calculus on a finite, discrete space (16, p.1)". The smoothness of the graph may be determined to study the signal variations on the graph. There are two kinds of the smoothness. The first type is the local smoothness that is given by Eq. 7:

$$\|V_i f\|_2 = \left[\sum_{j\in Nj\neq i} W_{i,j}\left[f(j)\text{-}f(i)^2\right]^2\right]^{1/2} \tag{7}$$

Equation 7 provides a measure of the local smoothness of signal f around vertex i. The local smoothness is a small number when f has similar values at i and all its neighbors. The second type is the global smoothness that is given by Eq. 8:

$$S_2(f) = \frac{1}{2}\sum_{i\in V}\sum_{j\in Nj\neq i}\left[f(j)\text{-}f(i)\right]^2 = f^T Lf \tag{8}$$

The global smoothness is also known as the graph Laplacian quadratic and it is equal to zero if and only if f is constant across all vertices. In general, $S_2(f)$ is small when the f has similar values at adjacent vertices connected by edges with a large weight. In summary, the connections among the nodes in any graph are encoded in the Laplacian graph matrix which is used to define both the graph Fourier transform and different notions of the smoothness.

In addition to the above operations on the graph, there are other operations of processing signals on the graph such as down-sampling, signal translation, different methods of filtering the signal and signal denoising, etc.

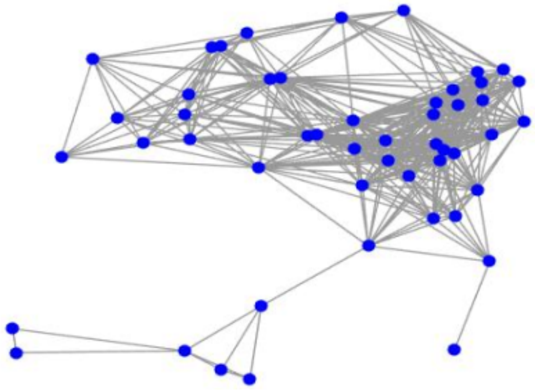

Fig. 4: Graph structure of 50 stands, blue circles represent vertices and gray links represent edges

Experimental: Based on the taxi cab stand locations in San Francisco which are shown in Fig. 3, the graph was constructed using their coordinates to analyze the performance of the GSP method in this framework.

A weighted undirected graph with a set of 50 vertices and a weight function; W: 50×50→ℝ was considered. Each entry of the weight matrix contains the weight of the edge connecting the corresponding vertices; $W_{i,j}$: $W(v_i v_j)$ and it is created proportionally to the inverse of the distances between them The distances vary from around 0.14 km to about 11.29 km, the closer two locations or vertices, the higher spatial correlation. A threshold was set up to remove those connections with smaller weights (Huang and Siliang, 2016; Anonymous, 2016). In this research, the connections having distances longer than 3 km, among the taxi stand pairs were removed. If there is no edge between two vertices, the weight is set to zero. The constructed graph is shown in Fig. 4. W is a symmetric matrix. The °d(i) is defined as the sum of the weights of incident edges:

$$d(i) = \sum_{j=1}^{50} W_{i,j}$$

And the matrix $D_{ii} = d(i)$. The graph Laplacian L is defined as L = D-W which is always symmetric and positive semi-definite.

The next step after constructing the graph is to calculate and process signals observed on each node in the graph. It is appropriate to consider a signal f as a vector of size (50×1), since, the ith component represents the signal value at the ith vertex. To generate the signal for this study, mobility traces of taxi cabs in San Francisco, the USA in May 2008 are used (Grady and Polimeni, 2010; Piorkowski *et al*., 2009). This data show cab locations, occupancy and time. More than 500 taxi cabs are provided with GPS devices to gather this information, approximately every 30 or 60 sec. The locations of 50 cabs were interpolated by synchronizing their data for specific times. The distances of vehicles to the stand locations were calculated and the count of vehicles (f) within a 1 km radius of the taxi stand for a period of two days was used for the data processing.

RESULTS AND DISCUSSION

After creating the graph and the observing signal on each node in the vertex domain, the graph Fourier transform was calculated using Eq. 4 that is equivalent to $\hat{f} = U^{\square} * f$ where U is the eigenvector of the Laplacian matrix. The cumulative spectral plot is given in Fig. 5.

The first spectral components as shown in the first row of Fig. 5 has the highest value compared to other components of the spectrum. The first row of the matrix U^*, u_0 associated with the eigenvalue λ_0 has the same positive magnitudes and equals to $1/\sqrt{50}$ at each vertex. Thus, the first spectral or DC component equally depends on the values of observed signals on all the vertices. Using 50 taxi cabs will increase the magnitude of the DC component when these taxi cabs are located between two or more of the taxi stand locations. In few words, stands, which are located in the middle of the city have maximum overlap in calculating the number of cabs around their locations due to the short distance between them. For instance**,** Fig. 6 and 7 shows signals on the vertices when the DC component has the highest value at the 33th time step. Therefore, many cabs are located in the downtown on Sat, 17 May 2008 23:15:00 GMT and the adjacent nodes almost shared the same magnitude of the signal.

Also, Fig. 5 shows the variation of the different spectral components over the selected time. For example, the 7th row has important negative spectral components within the time interval (75, 93) and (160, 190). The seventh spectral component is calculated by using the

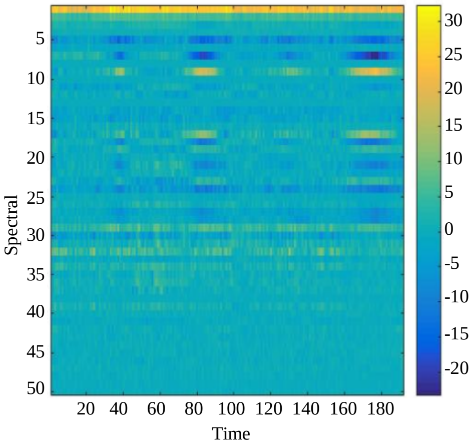

Fig. 5: The cumulative spectral plot for real data from Sat, 17 May 2008 15:00:00 GMT to Mon, 19 May 2008 15:00:00 GMT and 15 min as a time step

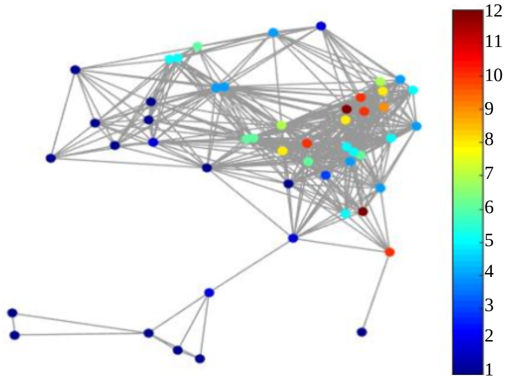

Fig. 6: The observed signals in the vertex domain when the DC component has the highest value

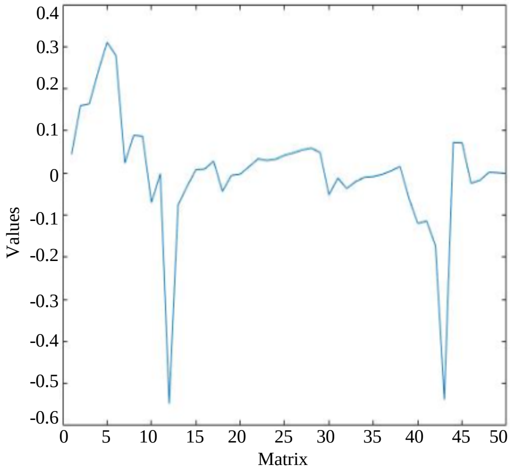

Fig. 7: The 7th Eigenvector of the Laplacian matrix L

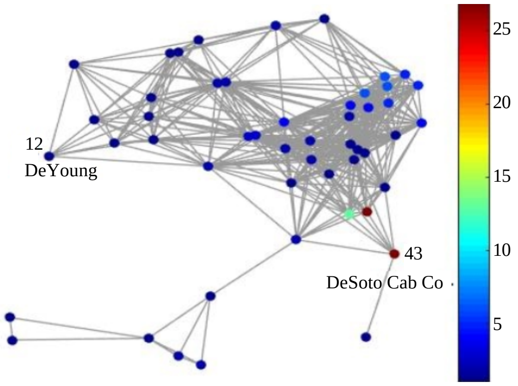

Fig. 8: The signal in vertex domain shows that most of the cabs are located around the node (43) at the 180th time step (Mon, 19 May 2008 12:00:00 GMT)

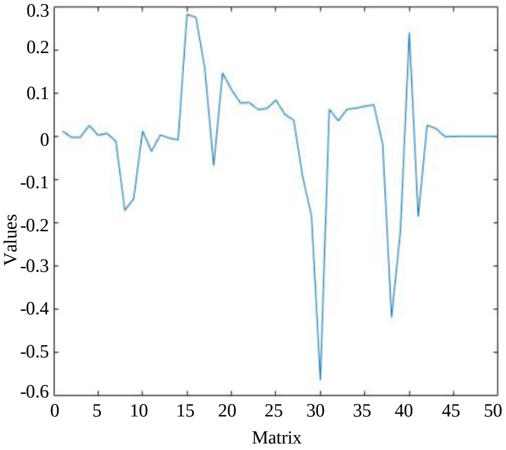

Fig. 9: The 29th eigenvector of the Laplacian matrix L

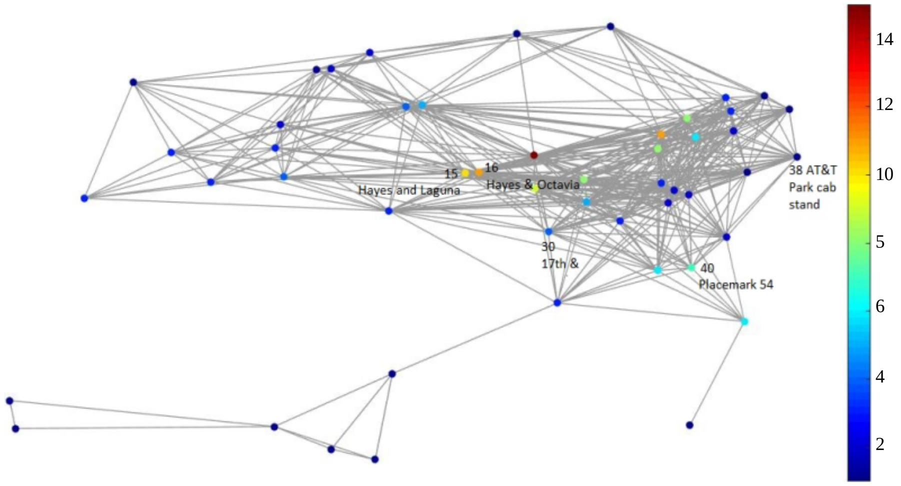

Fig. 10: The signal in vertex domain when the 29th spectral component is the highest at the 48th time step (Sun, 18 May 2008 03:00:00 GMT)

seventh row of the matrix U' which has significant negative components corresponding to the 12th and 43th stands as shown in Fig. 7.

Having cabs around one of these stands or both of them will highly affect the 7th spectral component as shown in Fig. 8. By taking another example, the 29th row has positive spectral components within the whole-time interval. The 29th spectral component is calculated by using the 29th row of the matrix U' which has positive components corresponding to the 15th, 16th and 40th stands and significant negative components corresponding to the 30th and 38th stands as shown in Fig. 9.

Therefore, the number of cabs around the 15th, 16th and 40th stands and are more than the number of cabs around the 30th and 38th stands as shown in Fig. 10. The Eigenvectors of the Laplacian matrix carry information about the structure of graphs, so, the above analyses illustrated that the anomaly of the spectral components reflects the relationship between the measured signals on the vertices (stands) whether or not they were adjacent.

The global smoothness that is found using Eq. 8 is shown in Fig. 11. This research is done using the graph signal processing toolbox, GSPBox (Perraudin *et al*., 2014). In general, having a variation in the measured signal on the nodes is represented by a big value of the global smoothness, especially when these nodes have large weights or they are adjacent. Figure 12 demonstrates the variation in the signal on the adjacent nodes when the

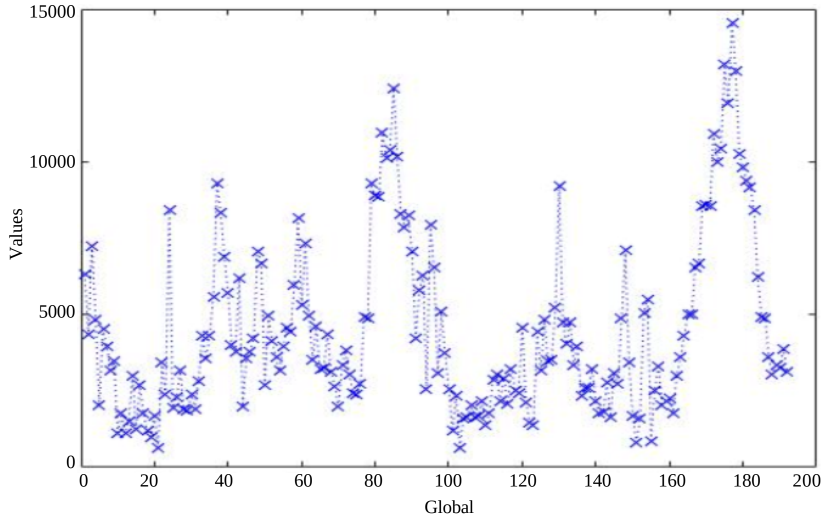

Fig. 11: The global smoothness of signals on the graph of 50 nodes

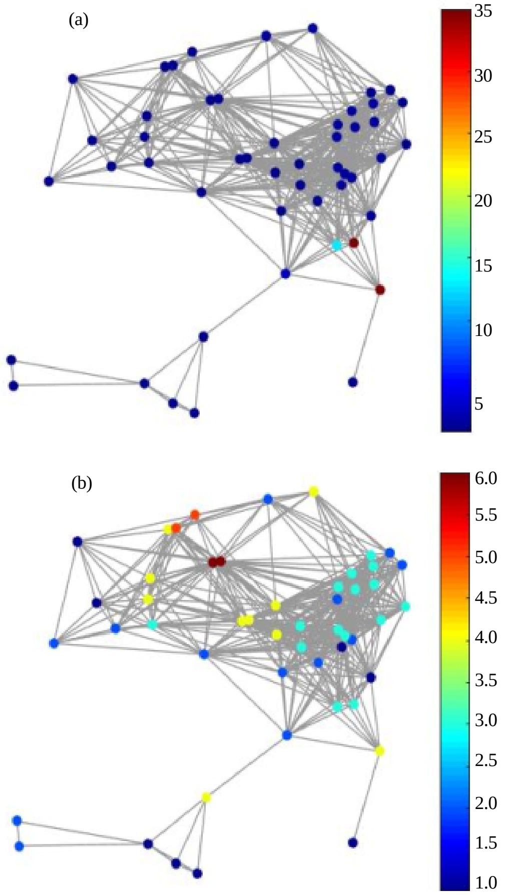

Fig. 12: The signals in the vertex domain compared to the global smoothness values: a) The value of the global smoothness is the highest and b) The value of the global smoothness is the lowest

highest and lowest values of the global smoothness were observed at the 177th and 21th time steps, respectively.

CONCLUSION

Having a graph of a sensor network allows the capability to handle signals from multiple sensors and distinguishes the anomalies in signals using the graph Fourier transform that depends on the underlying structure of the graphs. The obtained results can be used to provide good insight into the sensor's status in a network or they can be used for monitoring purposes, especially for a lot of stands, a huge number of cabs and considering the real range of the communication. For instance, it may be useful to apply GSP to study the driver's activity patterns, human mobility in term of the taxi occupancy, reducing the waiting time of the passengers and improving service quality. Moreover, comparing different regions over time could be a problem due to the size of the data for classical methods; while using the concept of the graph Fourier transform will make this research easier.

RECOMMENDATIONS

This framework may be effective for further Research Such as for roadside Units (RSUs) which offer connectivity support to passing vehicles in vehicular and hoc networks or for processing data on the number of cars passing through specific points in the city.

REFERENCES

Anonymous, 2016. Best places/streets to catch a taxi cab in San Francisco. San Francisco, California, USA.https://www.google.com/maps/d/viewer?mid=18sMrsEUD_X1-x5yn0MlEPhj0Uso&hl=en_US&ll=-3.81666561775622e-14%2C119.513491270321O8&z=1

Candia, J., M.C. Gonzalez, P. Wang, T. Schoenharl and G. Madey *et al*., 2008. Uncovering individual and collective human dynamics from mobile phone records. J. Phys. Math. Theor., 41: 1-16.

Grady, L.J. and J.R. Polimeni, 2010. Discrete Calculus: Applied Analysis on Graphs for Computational Science. Springer Science & Business Media, Berlin, Germany, ISBN:978-1-84996-290-2, Pages: 366.

Huang, J. and W. Siliang, 2016. A study on the use of graph signal processing techniques for satellite-based navigation systems. Proceedings of the 2016 International Confrence on Technical Meeting of the Institute of Navigation, January 25-28, 2016, Hyatt Regency Monterey, California, USA., pp: 448-455.

Leutzbach, W., 1988. Introduction to the Theory of Traffic Flow. 1st Edn., Springer, Berlin, Germany, ISBN:978-3-642-61353-1, Pages: 204.

Liao, Z., S. Yang and J. Liang, 2010. Detection of abnormal crowd distribution. Proceedings of the Joint 2010 IEEE/ACM International Conference on Green Computing and Communications and Cyber, Physical and Social Computing, December 18-20, 2010, IEEE, Hangzhou, China, ISBN:978-0-7695-4331-4, pp: 600-604.

Loukas, A., 2015. Distributed graph filters. Ph.D Thesis, Delft University of Technology, Delft, Netherlands.

Mahyari, A.G. and S. Aviyente, 2014. Fourier transform for signals on dynamic graphs. Proceedings of the 48th Asilomar Conference on Signals, Systems and Computers, November 2-5, 2014, IEEE, Pacific Grove, California, USA., ISBN:978-1-4799-8295-0, pp: 2001-2004.

Meng, Q., S. Mabu, L. Yu and K. Hirasawa, 2010. A novel taxi dispatch system integrating a multi-customer strategy and genetic network programming. J. Adv. Comput. Intell. Inf., 14: 442-452.

Perraudin, N., J. Paratte, D. Shuman, L. Martin and V. Kalofolias *et al*., 2014. GSPBOX: A toolbox for signal processing on graphs. Comput. Sci., 2: 1-8.

Piorkowski, M., N. Sarafijanovic-Djukic and M. Grossglauser, 2009. CRAWDAD dataset epfl/mobility. Cajun Crawdad's Inc., Byhalia, Mississippi, USA.

Sandryhaila, A. and J.M. Moura, 2014a. Big data analysis with signal processing on graphs: Representation and processing of massive data sets with irregular structure. IEEE. Signal Process. Mag., 31: 80-90.

Sandryhaila, A. and J.M. Moura, 2014b. Discrete signal processing on graphs: Frequency analysis. IEEE. Trans. Signal Process., 62: 3042-3054.

Shuman, D.I., B. Ricaud and P. Vandergheynst, 2016. Vertex-frequency analysis on graphs. Appl. Comput. Harmon. Anal., 40: 260-291.

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

Susnjara, A., N. Perraudin, D. Kressner and P. Vandergheynst, 2015. Accelerated filtering on graphs using Lanczos method. J. Signal Process. Graphs, 1: 1-11.

Venkatesan, N.E. and R. Kannan, 2013. Graph structured data viewed through a fourier lens. Ph.D Thesis, Berkeley University of California, California, USA.

Wang, H., R.L. Cheu and D.H. Lee, 2014. Intelligent taxi dispatch system for advance reservations. J. Pub. Transp., 17: 115-128.

Zhang, F. and E.R. Hancock, 2008. Graph spectral image smoothing using the heat kernel. Pattern Recogn., 41: 3328-3342.