EasyChair Preprint

№ 687

Geo and Graph Analytics for Dynamic Cellular Transactions Insights, Improving Quality of Service and Business Decisions: "Quality X Map"

Dahj Muwawa Jean Nestor and Kingsley A. Ogudo

December 19, 2018

Geo and Graph Analytics for Dynamic Cellular Transactions Insights, Improving Quality of Service and Business Decisions: "Quality X Map"

***Abstract*— Everyday, millions of users on their phone generate huge amount of traffic; intelligent industrial machines and IoT devices also generate monumental traffic, both affecting Network performance due to the shared communication medium used. With the large amount of human and machine generated data, Communications Service Providers (CSPs) face the challenge of finding values and addressing Quality of Experience (QoE). The advance of Data and Predictive Analytics methods makes it not easier but possible to dive deep into Network transactions to build intelligent insights, helping with business decisions. In this paper, we use Geo-Analytics and Intelligent Localization algorithms, combined with Graph Processing to provide a dynamic insight of Network data, to sustain business decisions and improve users and devices' QoS and QoE.**

Keywords—Communication Service Providers (CSPs), Geo-Analytics, Graph Analytics, Cellular Transactions, Machine Learning, Quality of Service (QoS), Quality of Experience (QoE), Business Decisions, Internet of Things (IoT).

I. INTRODUCTION

As Data Analytics is taking its toll across all the sectors of businesses, so it is in the Telecommunications environment. With massive generation of data records by subscribers and IoT devices, Communication Service providers (CSPs) face challenges in transforming the data to actionable values for low cost effective business decision making.

Hence, big investments on Data Analytics, Customer Experience Management (CEM) and Service Quality Management (SQM). It is therefore, not objective to only take advantage of the fast processing capabilities of the Data Analytics systems, but to also provide visualization in a simplistically manner, supporting and enhancing business decisions. Data visualization, with the advancement of computing power, has taken a different perspective [1]. Subscribers are constantly using their devices, meaning the volume size of network traffic or transactions keep growing, accelerating the need for fast, efficient and user-friendly data analytics systems to improve decision making process.

Estimating the number of transactions that can be generated in a period by telephones users, we introduce an effective, realistic and simple way to visualize network transactions information using Geographic Information System (GIS) and graph processing algorithm. The objective of the paper is to go beyond the popular data visualization schemes, such as charts (pies, bars, lines), maps and tables for data insights. Graph algorithms have been in use for a long time in the history of computer science and Data Analytics [2]. Figure 1 illustrates the concept of the study which explores the advantage of spatial Analytics, Geo databases functionalities, flexibility and agility of graph processing.

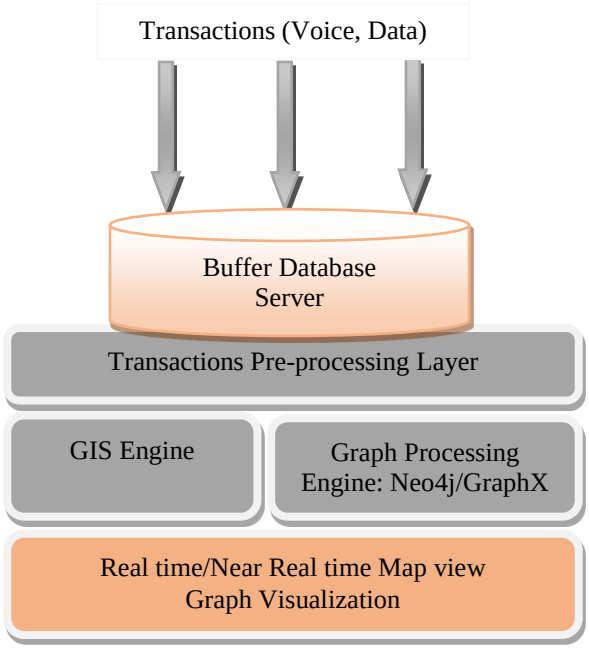

Figure 1 Study Concept and Model

The focus of this paper is more on the application layer (visualization) because the Back-End platform can be customized depending on the needs, ranging from Structured to unstructured data. The **Quality Exploration Map (QXM)** is the denomination used in this paper to refer to the application visualization layer.

II. STUDY REVIEW

The use of Graph Processing in the past has been remarkable in areas such as Social Networks with big companies such as Facebook, LinkedIn, and Twitter adopting the technology and modeling of diverse web traffic [3]. Many researches also are being conducted around the study field of graph processing, with the development of technologies such as Spark GraphX [4].

Jakob Smedegaard Andersen and Olaf Zukunft used GraphX to evaluate the scaling of Graph-Algorithms for Big Data [5], in which they analyze social media graphs using semi-clustering method. The study explores relationship and interactions

between users, the addition of friends, channel subscription, message exchanges etc. Relationships have certain weight depending on the interactions and messages exchanges. In this model, the Vertices connote users and the Edges connote relationships between users. The objective of Jakob and Olaf's study of social behavior was to define a certain number of user groups defining a certain criterion such as strong relationships with another category of user group. The relationship equation is given by the equation (1), given a cluster of users c, a set of relationships or properties S:

$$Sc = \frac{I_C - f_B B_C}{V_C(V_C - 1)/2} \quad (1)$$

Where Ic is the sum of the weights of all relationships inside the cluster c, Bc the sum of weights of all relationships that go outside the cluster, f_B a factor to Bc, which is assumed to be in the interval [0, 1] and Vc is the number of vertices, here representing users in the cluster. The output of a semi-clustering is illustrated in Fig. 2.

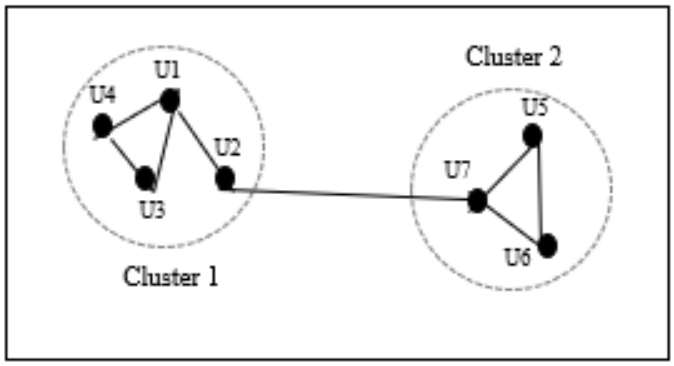

Fig. 2 Illustration of a semi-clustering Graph Algorithm Output

The semi-clustering of graph modeling has been studied by G. Malewicz et al. in their Pregel model [6]. However, the model is undirected. Graph processing used in GraphX or Neo4j are mostly directed.
Using the same approach, transactions from Network data users are analyzed based on a set of criteria to provide the Communication Service Providers, a deep insight of customer's transaction information, network device modules information and a ground to improve Quality of Service (QoS) which is good for business.

On the other side, the use of Geo Location has been widespread to give a different insight of spatial data; such examples include the use of heatmaps, density maps, 3D Maps in different areas such as Network Cells planning and Optimization, mobile phones Geolocation and many more applications. Exploiting the power of Geo Analytics, facilitates the understanding of traffic distribution. Michelle Angelico *et la.* used Smart Data Localization, based on Geo Analytics of environmental and territorial data to improve quality of life by promoting the development of Business Support platforms [7]. Jianghua Zhao *et al.* go even further using GIS system to predict traffic noise, segmenting the roads which at sensitive points, contributes with the highest noise level [8].

- The Continuous QoS and QoE song: Communication Service Providers are always taking quality as a serious topic. It is a song that will continuously be played because Network Infrastructures are continuously changing. Good network quality gives competitive edge to CSPs [9].

III. Contribution of the Study

The use case aims to provide an efficient visualization of network traffic displaying, at the same time the quality indicators. The use case model or approach provides the CSPs with a visualization that facilitates the detection of abnormalities in the Network. Using this method, CSPs not only saves on CAPEX investment, they leverage on the openness and advancement of technologies such as Big Data, Prediction Analysis, Geo Analytics etc. Adopting a graph and visualizations as approached in this paper also gives CSPs a multi-department use case platform for good and cost effective business decision making process. The area of Business Intelligence and Data Visualization is an evolving area and will always leave gaps for improvements.

IV. Methodology and Design

When referring to Analytics a certain number of processes and models are overlooked. From data collection to data visualization, complex methods and algorithms can be considered. However, in this paper, data preprocessing is considered done as part of the Back-End functionalities. Fig. 3 shows the steps used for this research starting directly with Exploratory Data Analysis.

- Exploratory Data Analysis: to understand the dataset and the meaning of variables and exploration of indicators.
- Data Query Aggregation: to create and facilitate the modeling of vertices and edges of the dataset. This is the fundamental use case step to support decisions in the model.
- Data Visualization: data presentation layer, mostly based on a Map or/and graphical processing view. For this paper, Google Map, Spark GraphX and Neo4j are used with R to illustrates the model.

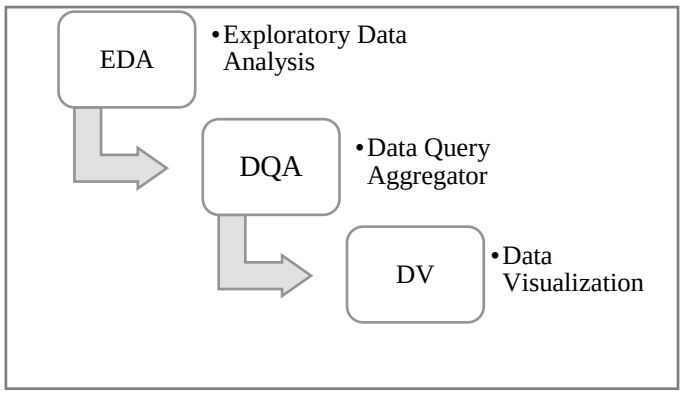

Fig. 3 Design Process Used in this Paper

A. Quality X Map Model

The visualization Model is built on Graph processing approach overlaid with GIS view. "X" refers to the indicators or aggregation of indicators which can be used to visualize the quality level. X can be indicator, score, aggregator, count, mark depending on the applications.

Given a Structured Dataset D with variables k, u, v and c, the dataset can be represented as:

$$D(u, v, c) \rightarrow k^i|u^i|v^i|c^i \qquad (2)$$

Where k and u are the variable keys of the dataset, v is the variable Indicators (KPIs, Calculated score, count) and c is the spatial variable. Explicitly, the dataset can be shown as:

Table 1 Transactional Dataset Model

k	u	v	c
k^1	u^1	v^1	c^1
k^2	u^2	v^2	c^2
.	.	.	.
.	.	.	.
.	.	.	.
k^i	u^i	v^i	c^i

The Quality X Map allows us to represent the transactions in the dataset in holistic manner, giving at glance the geographical points of interest as illustrated in Fig. 4. The Figure shows in real time, near-real time or historical manner, the transactions analysis. On the Graph processing, k and u are vertices of the graphs systems and v represents the edges of the graphs as shown in Fig. 5. The model can be used for streaming applications where the edges are directed as files sent. To represent data on the GIS, the dataset must have spatial information, either as coordinates or administrative regions. Therefore, for Geo Intelligence, c must be present.

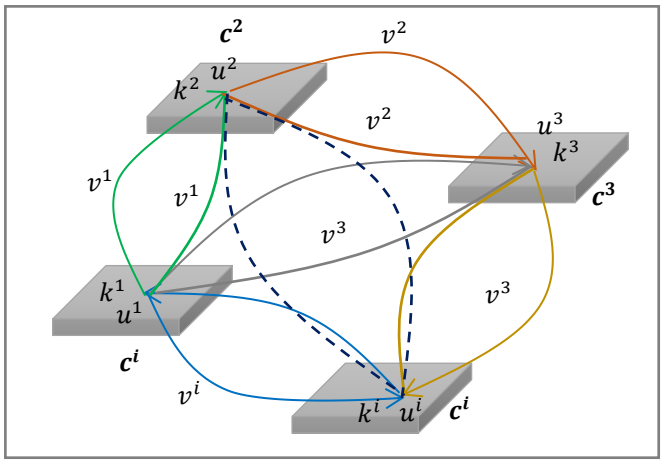

Fig. 4 Quality X Map model for GIS view

v^i is the Quality variable that represents a relationship or interaction between two dataset keys k, v by geographical region or coordinates c^i. Typical use case, number of data sessions between Gauteng and Western Cape servers; active subscribers YouTube streaming sessions towards a specific Gateway server.

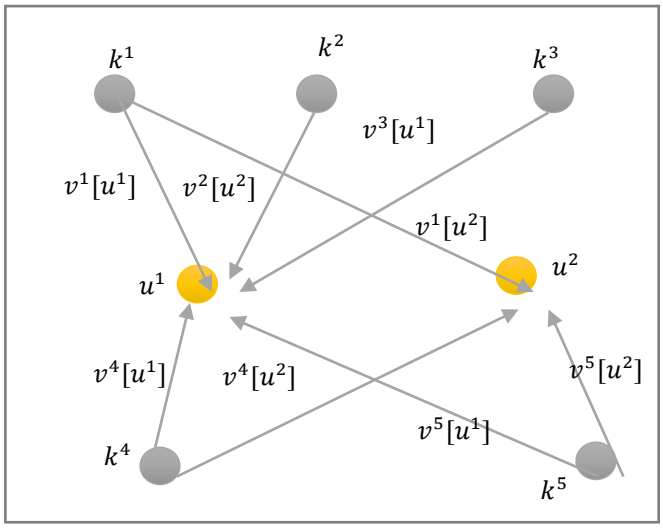

Fig. 5 Quality X Map Model Graphical view

$v[u]$ is the quality index towards the node u and vertices whose weight determines the value of X. And k is the vertices and start nodes for transactions. Example, analyzing the transactions towards a server to check for possible congested links and failures. In the Graphical view, the parameter c is not required unless it is taken as vertices to enable a graphical view aggregated on c.

B. Quality X for Mobile Data Transactions

In this section, we use public Telecoms data to apply the Quality Map model in a real dataset. We use quantitative QoS (Quality

of Service) [10]. Data from an Internet Service Provider is analyzed to provide the overall performance of Services provided to customers. The summary of the data is shown in table 2.

Table 2 Mobile Data Dataset Summary for Analysis

Quality variable	Unit	Data type	Role
AccessID	None	Varchar	Vertex
InternetGateway	None	Varchar	Vertex
ProtocolApp	None	Varchar	Vertex
ActiveDataSessions	None	Bigint	Edge
DataUsage	Gbytes	Bigint	Edge
RtxPacket	Gbytes	Bigint	Edge
packetloss	%	Decimal	Edge
Ul_latency	msec	Decimal	Edge
Dl_latency	msec	Decimal	Edge
Dlthroughput	Mbits/sec	Decimal	Edge
Ulthroughput	Mbits/sec	Decimal	Edge
Region	None	Varchar	Can be used as Vertex.
Latitude	None	Decimal	Can be used as Vertex
Longitude	None	Decimal	Can be used as Vertex

Region (administrative), Latitude and Longitude (Geographical Coordinates) are spatial variables which can also be used as vertices for regional aggregated values.
Using GraphX, the below use cases are evaluated using Quality Map model.

1) Real time Network Transactions and Congestion Analysis using Quality Map Model.

CSPs and ISPs (Internet Service Providers) have to keep a close look on Server's load to avoid possible server crashes and overload. While many mechanisms and algorithms of congestion control are being studied [11], one of the challenges is to oversee possible congestion in links. Figure 6 illustrates the Geo representation of the Quality X Map Model for traffic links analysis. Based on table 2, the ISP has 2 Gateway servers for national internet connection. The model provides the simultaneous active data sessions per server, giving the opportunity to view any other Indicator related to traffic Analysis. Figure 6 dynamically shows the amount of sessions originating from different access sites towards the gateway servers. Due to load balancing, Access sites can connect to both servers. Fig. 7 shows the same use case based on Graph processing using iGraph with R. Representing the same data on a chart could be difficult to read (~ 100 access points).

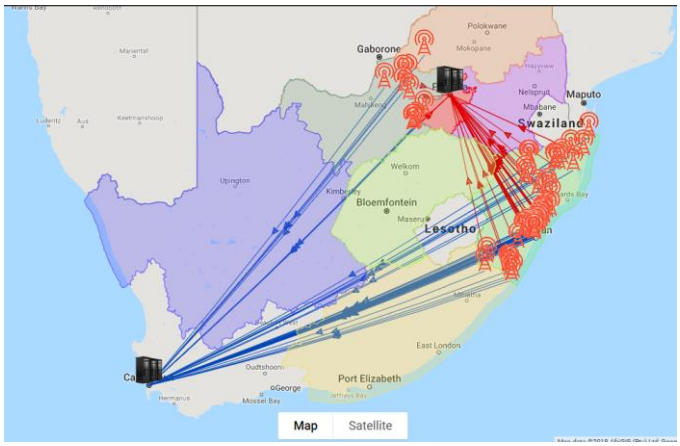

Fig. 6 Real time Geo Analytics of Server transactions for traffic load

The two servers of the CSP are shown in Figure 6 with one big gateway server in Pretoria and the other one in Cape Town. The tower represents the access sites from the users. From the Map, provide the ISP with the ability to pin-point the server having excessive traffic loads. The Model displays the traffic flows as it happens, in near real time.

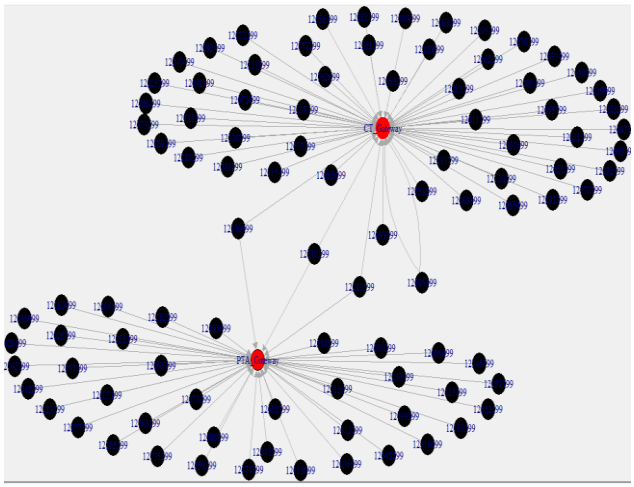

Fig. 7 Simultaneous Active Data sessions per Gateway Server

Fig. 7 displays the same analysis in a Graphical way, reducing the possibility to overcrowd the data space and ease recommendation processes.

2) NetFlix Traffic Analysis using Quality X Map Model

The objective is to use the model to provide NetFlix (video-streaming) traffic Analysis per Region. Applications for which may be linked to sales, marketing and Operations.

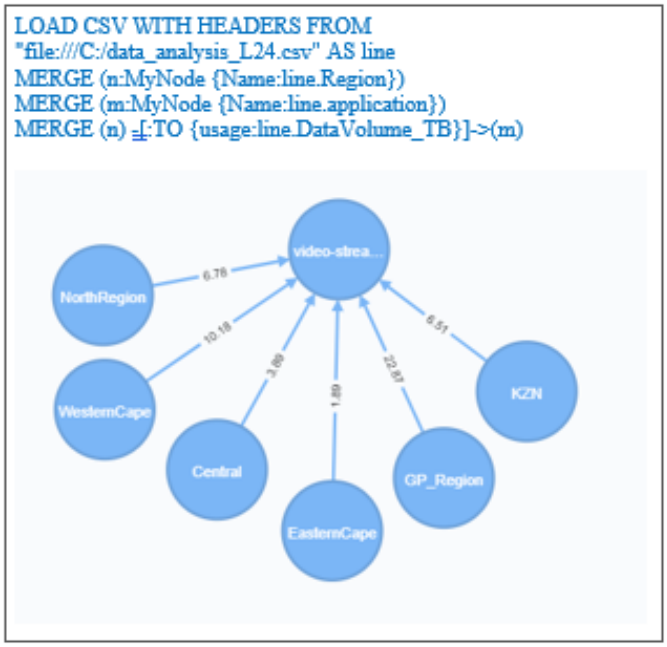

Fig. 8 NetFlix Traffic Analysis Per Region

Fig. 8 displays the Graph Model of Analyzing traffic distribution of specific traffic using Neo4j platform. The edges display the total number of data used in Terabytes. The edges represent the interaction between different nodes or vertices. Any other metric could ne used in place of data volume.

3) Customer Call Analysis using Quality X Map Model

The use case uses the Quality X Map to plot the user call activities in a dynamic way. The application of which could be linked to customer experience and fraud detection. Fig. 9 illustrates the Analysis of group call trace.

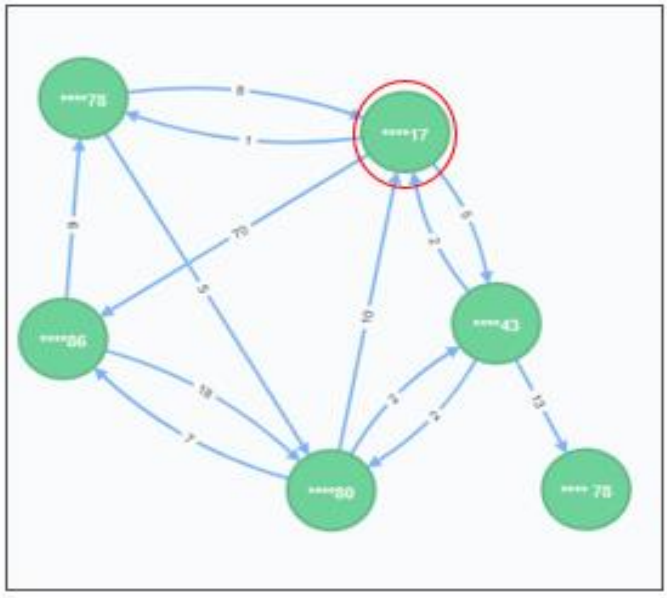

Fig. 9 Network Graph of Customer Call Analysis

The target subscriber is ***17 (Subscriber anonymized for privacy purposes). The analysis exposes the call transactions from and towards the same group of people with user ****17 as reference. The users are vertices and the number of calls is edges, showing the interactions between vertices (subscribers). ****17 made 5 calls to ***43 who made 2 calls to ****80. Subscriber ****80 then called ****17 back 10 times.

V. Conclusion

Information visualization is a capital part of Data Analytics process, because it is through visualized reports and dashboards that business decisions are taken. Overlooking beyond the traditional forms of data presentations, the Quality X Map model allows an instantaneous view of Network status and performance in an easy and understanding way for all. Graph processing is becoming more and more an efficient way to analyze and visualize data. Our study combines the benefits of Graph processing and GIS to provide a different business Insight to Data transaction characteristics. The model can be adjusted to fit any other area beside Telecoms. With the increase in cellular usage, a certain level of dynamism should be adopted in handling transaction data [12]. Geo Analytics in combination with Graph processing as described in this paper, provides accessibility to critical network information at any location in real time and near-real time [13].

The Quality X Map provides the following benefits:

- Flexible data visualization: every network transaction can be represented as a graph or as GIS if it contains spatial information.
- Very easy to expend the data by adding vertices.
- Easy to use model for Real time monitoring.

Telecoms transactions are of a big magnitude because of the number of events executed by Mobile devices, therefore, require Geo-Analytics and Graph Processing for low cost effective smart reporting, prediction and recommendation capabilities [14]. Looking at Celular Network planning, the use of Geo Analytics has been of capital use [15], [16]. That is to insinuates that the communication environment applies Geographic data in diverse operations including Line of Sight Analysis for Microwave Transmission signals, users and event geolocations [17]. The model pushes the use of Geo Intelligence and Graph processing towards Analytics and Customer Experience analysis.

VI. Future Work

The model has been applied to a small set of datasets, with less emphasis on Data pre-processing. The future work will focus on the Analysis of roaming traffic, which we believe would be very suitable for the Quality X Map model of Visualization. We will include large network dataset, taking advantage of Big data processing mechanisms such as Hadoop, Spark in-memory

computing [18] to speed up the processing of real time and batch big data. That also means that the computing power of the current server will be improved (RAM, Storages) to allow the processing of large datasets. We will also look at using some clustered algorithms to enhance the model and the span of application. We have used two different Graph visualization platforms to demonstrate the model (Neo4j and iGraph with R); However, as data increases, we will use the Spark GraphX to handle more data. The future work will include the Analysis of various domains of Network as shown in Fig. 10, including network performance scores, Quality Score Map.

Fig. 10 Future work' s scope for Quality X Map Model

REFERENCES

[1] A. Azzalini and B. Scarpa, "New problems and new opportunities", in *Data Analysis & Data Mining, and Introduction*, Ed. By Oxford University Press, 2012, pp 2-8.

[2] R. Diestel, Graph Theory, 4th ed. Heidelberg: Springer, 2010.

[3] A. Silberstein, A. Machanavajjhala, and R. Ramakrishnan, "Feed following: The big data challenge in social applications," in Databases and Social Networks, ser. DBSocial '11. New York, NY, USA: ACM, 2011, pp. 1–6. [Online]. Available: http://doi.acm.org/10.1145/1996413.1996414

[4] J.E. Gonzalez, R.S. Xin, A. Dave, D. Crankshaw, M.J. Franklin and I. Stoica, "GraphX: graph processing in a distributed dataflow framework", OSDI'14 Proceedings of the 11th USENIX conference on Operating Systems Design and Implementation, Broomfield, Pg: 599-613, 2014.

[5] J.K. Andersen and O. Zunkuft, "Evaluating the Scale of Graph-Algorithms for Big Data using GraphX", IEEE 2nd International Conference on Open and Big Data, 2016.

[6] G. Malewicz, M. H. Austern, A. J. Bik, J. C. Dehnert, I. Horn, N. Leiser, and G. Czajkowski, "Pregel: A system for large-scale graph processing," in Proc. of the 2010 ACM SIGMOD Intl. Conf. on Management of Data, ser. SIGMOD'10. New York, NY, USA: ACM, 2010, pp. 135–146.

[7] M. Angelaccio, B. Buttarazi, A. Basili and W. Liguori, "Using Geo Business Intelligence to Improve Quality of Life", IEEE International, 2012.

[8] J. Zhao, Q. Qin, C. Xie, J. Wang, and Q. Meng, "AN EFFICIENT METHOD OF PREDICTING TRAFFIC NOISE USING GIS", Institute of Remote Sensing and GIS, Peking University 10087 1, Beijing, China, 2013 IEEE.

[9] Power, J. D., and Associates (2003), "Wireless network quality assessment study™". Retrieved November 11 th, 2004, from http://www.jdpower.com/studies/pressrelease.asp?StudyDD=891.

[10] R. Rodríguez, D. Fernández, H. Montes, S. Hierrezuelo, G. Gómez, "Quality of Service Mechansms" in *End-to-End Quality of Service over Cellular Networks*, Ed. John Wiley & Sons, 2005, pp 103-137.

[11] S. Fang, C.H. Foh, K.MM.Aung, "Prompt Congestion Reaction Scheme for Data Center Network using Multiple Congestion Points", IEEE ICC Next Generation Networking Sympsium, pp. 2679-2673, 2012

[12] Brockmann, T., Stieglitz, S., Kmieciak, J., and Diederich, S., 2012. "User Acceptance of Mobile Business Intelligence Services". 2012 15th International Conference on Network Based Information Systems, vol. 23, pp. 861-866.

[13] P. Gray, "Business Intelligence: a new name or the Future of DSS?" in *DSS in the Uncertainty of the Internet Age*, Eds Bui T., Sroka H., Stanek S. Goluchowski J., Publisher of the Karol Adamiecki University of Economics, Katowice, 2011.

[14] F. Facchini, R. Giuliano, F. Mazzenga, "UWB Detect and Avoid Procedure for WiMAX victims", IET Commun., Vol.3, Is.2, Feb. 2009, p.268-278.

[15] R. Giuliano, F. Mazzenga, "Exponential Effective SINR Approximations for OFDM/OFDMA-based Cellular System Planning", IEEE Trans. Wir. Com. (TWC2009), Vol.8, No.9, Sep. 2009, p.4434-4439.

[16] R. Giuliano, P. Loreti, F. Mazzenga, G. Santella, "Fractional Frequency Reuse Planning for WiMAX over Frequency Selective Channels", IEEE Int. Wirel. Comm. Mob. Comp. Conf. (IWCMC 2008), Chania Greece, Aug. 2008, p.666-671

[17] J. Lempiäinen, M. Manninen, Radio Interface System Planning for GSM/GPRS/UMTS, Kluwer Academic Publisher, 2001.

[18] J. Betser, D. Belanger, "Architecting the Entreprise via Big Data Analytics" in *Big Data and Business Analytics*, Ed. by Jay Liebowitz, CRC Press: Taylor & Francis Group, 2013, pp 2-50.