

# RANKING NODES IN COMPLEX NETWORKS: A CASE STUDY OF THE GAUBUS

T. Moyo<sup>1\*</sup>, W. Musakwa<sup>2</sup>

<sup>1</sup> Department of Operations and Quality Management, University of Johannesburg. Corner Siemert & Beit Streets, Doornfontein 0184 Johannesburg, South Africa- thembanijoe@gmail.com

<sup>2</sup> Department of Town and Regional Planning, University of Johannesburg, Corner Siemert & Beit Streets, Doornfontein 0184 Johannesburg, South Africa - wmusakwa@uj.ac.za

**KEY WORDS:** Mobility, Centrality, Strava Data, Public Transport, Nodes

## ABSTRACT:

Connecting points of interest through a well-planned, inter-connected network provides manifold benefits to commuters and service providers. In the South African context, traffic congestion has become of great concern. Given how the South Africa community is slowly developing towards the use of multi-modes of mobility, the Gautrain network can be used to promote the use of multi-modes of mobility, as the Gautrain has been identified as the backbone of mobility within the Gauteng province. Currently commuters have the option to board the Gaibus (a form of Bus Rapid Transit) at their origin points which will take them to the Gautrain station to board the Gautrain. The problem to be solved arises when a commuter wishes to traverse from any bus stop to the Gautrain station, currently he/she only has one option and if the bus network has a shutdown at any point in the network the commuter's journey will not be possible. In solving this problem, we consider the problem of graph robustness (that is creating new alternative routes to increase node/bus stop connectivity). We initial use Strava data, to identify locations were cyclist prefer to cycle and at what time of day. In graph theory, the nodes with most spreading ability are called influential nodes. Identification of most influential nodes and ranking them based on their spreading ability is of vital importance. Closeness centrality and betweenness are one of the most commonly used methods to identify influential nodes in complex networks. Using the Gaibus network we identify the influential nodes/ bus stops, using the betweenness centrality measure. The results reveal the influential nodes with the highest connectivity as these have cross-connections in the network. Identification of the influential nodes presents an important implication for future planning, accessibility, and, more generally, quality of life.

## 1. INTRODUCTION

Over the years there has been a growing body of academic works for solving problems relating identifying influential nodes in networks. Generally in large networks, the nodes with most spreading ability are called influential nodes (Bang-Jensen et al., 1999). Identification of these influential nodes and ranking them based on their spreading ability is of vital importance, as this will ensure high connectivity in the network. Common techniques of exploring the problem of identifying the influential nodes in networks are by measuring either the degree centrality, betweenness centrality or closeness centrality of nodes. (Sabidussi, 1966; Freeman, 1978; Freeman, 1980).

With regards to this research the betweenness centrality algorithm can be used to identify influential nodes (bus stops) along the Gaibus route network. The score deduced from the algorithm will reflect an estimated time of travel for commuters from their points of origin to their destination through the shortest path along the network. However, Frank (1992) articles before one utilises the centrality to determine identify influential nodes, it should be noted that through highly efficient in determining the influential nodes, betweenness centrality due to the computational complexity involved in calculations, is not easily applicable in large-scale networks, hence for this study we analyse only a section of the Gaibus network.

## 2. RELATED WORK

### 2.1 Public Transportation

Public transportation systems in modern cities across the world are becoming more complex, as the movement patterns of citizens has evolved into a complex social case of interest. This scenario has left city planners seeking at address the numerous challenges that come with managing such a complex system.

Globally the move towards developing smart mobility systems has proven a viable endeavour in tacking this complex mobility problem. With the notable milestones in research assessing commuter movement patterns from mined data from Web 2.0 platforms (O'Reilly, 2007; Chen et al, 2011; Hasan & Ukkusuri, 2014), Internet of Things (Ashton, 2009; Zanella, 2014; Moyo & Musakwa, 2016) and Big data (Mcfee et al, 2012; Chen, Chiang & Storey, 2012; McNulty, 2014; Wu et al, 2014).

The term Web 2.0 emerged in early 2004, when researchers outlined how a new generation of the World Wide Web had emerged, which brought with it new opportunities that were previously unachievable (O'Reilly, 2007; Moyo & Musakwa, 2016). In exploring the potential of Web 2.0, Chen et al (2011) utilised check-in data from Foursquare, Facebook and Gowalla to determine factors influencing citizen movement patterns. Whilst Hasan & Ukkusuri (2014) mined data from various Web 2.0 platforms as a means to visualise movement patterns. Through this studies, reveal the potential of mining data from Web 2.0, as it can help city planners monitor and predict mobility trends in the city. However a notable draw-back with this data is that it can only be visualised if the text, images, videos on the Web 2.0 platforms are geo-tagged.

\* Corresponding author

This internet of things (IoT) which is now closely linked with the Fourth Industrial Revolution (4IR) has made it possible for people and devices to stay connected and build relationships over large distances, from meetings being moved from the traditional boardrooms to being held via skype or directions being given via tom-tom instead of paper maps. Riggins & Wamba (2015, p1), has express how “the emerging IoT allows for the tracking and tracing of any tagged mobile object as it moves through its surrounding environment or a stationary device that monitors its changing surroundings.” The term IoT, has also been used to refer to how technology has advanced in the past decade, with various devices now being able to share information with each other through the use of embedded sensors, examples include mobile phones, vehicles, traffic sensors which are linked to various communication systems (Ashton, 2009; Zanella, 2014; Moyo & Musakwa, 2016)

The rapid growth in information flow, storage size and type of data has led to a new term being coined being big data. IBM (2012) has described big data as “datasets whose size is beyond the ability of commonly used software tools to capture, manage and process the data within a tolerable elapsed time.” Although this definition explains big data, it still fails to clearly define big data as it focuses on size aspect of big data. Hence to truly define big data we seek out a comparison with ‘normal data’. This has led to what is now commonly termed the V’s of big data that is volume, velocity, veracity, variability, variety these which will lead to value (Mcfee et al, 2012; Chen, Chiang & Storey, 2012; McNulty, 2014; Wu et al, 2014).

Research into these new data sources hence outlines how intertwined the problem of ensuring reliable public transportation is with balancing the available public transportation infrastructure. Consequently the path towards ensuring smart mobility relies on understanding the existing mobility network and identifying which hubs, bus stops, train stops, have the most influence in the network. These influential nodes can be further invested into to enhance the mobility network experience. A possible means to identify these influential nodes is through centrality measures.

## 2.2 Centrality

Over the past years centrality measures have been used to assess various networks, in order to rank nodes by the level of importance (Agryzkov, et al, 2014). This ranking has assisted researchers to understand mobility networks, social media networks, electronic networks and computer networks (Freeman, 1979; Crucitti, Latora & Porta, 2006).

Understanding the centrality of a network is essential for optimization and solving graph theory problems. Freeman (1979, 225) outlines how “a point is central to the degree that the distances associated with all its geodesics are minimum. Short distances mean fewer message transmissions, shorter times and lower costs.”Consequently numerous centrality measures have been proposed over the years from the degree centrality which ranks nodes based on the number of neighbours (Freeman, 1978) closeness centrality which ranks nodes by the reciprocal of the sum of the length of the shortest paths between the node and all other nodes in the graph

(Sabidussi, 1966; Freeman, 1978). Hence through using the closeness centrality we are able to identify the node which is closest to all other nodes.

The betweenness centrality on the other hand is refers to the centrality measure which detects the amount of influence a particular node has over the flow of data across the network. For connected networks, it be computed as a representation of the number of in-degree in the network. Freeman articulates for the betweenness centrality “a point is viewed as central to the extent that it can avoid the control potential of others.” (Freeman, 1979, p. 224). Contemporary it has been used in research to identify nodes that act as a bridge from one section of the network to another (Frank, 1999).

A notable milestone in centrality measures was the growth in theorems which determine the minimum number of edges required to be added in a network to obtain a  $k$ -edge-connected network (Frank, 1999). These algorithms hence were developed to ensure this property of minimum number of graph-edges, whose addition to a given network would increase the connectivity of the whole network (Narula & Ho, 1980; Graham & Hell, 1985; Pettie & Ramachandran, 2002; Qi et al, 2015).

## 3. METHODOLOGY

### 3.1 Study Area

Sandton is an affluent area situated in the Johannesburg Municipality, Gauteng, South Africa. It is the City of Johannesburg consists of more than 4.4 billion people, which accounts for approximately 36% of the Gauteng population and 8% of the national population (Todes, 2012). Sandton is the business capital of Johannesburg and is characterised by its gleaming towers and the profusion of construction cranes towering above the central suburb. For most South African's Sandton and business are intertwined as most of South Africa's top companies, and the Johannesburg Stock Exchange, are headquartered here. Central Sandton is Johannesburg's premier shopping district, wall to wall with international high street and luxury labels. It's also the location of many of Johannesburg's top business hotels and a high-profile conferencing destination.

Given this background of Sandton, a Gautrain (a high speed train) station was strategically placed at the heart of Sandton City, to take advantage of Sandton economic history and social importance (see Figure 1). After the construction of the train station, The Gaibus was then rolled out as an extension of the Gautrain, to link commuters to and from neighbouring suburbs around Sandton. However currently not all the Gaibus stops are being fully utilised as some have become ‘ghost’ stations as few if non-commuters board or drop off at these stops. Commuters have expressed how the current stops are not located at their points of interest. This has led to a need to augment the current Gaibus system, in order to fully utilise all the existing stops. Moyo, Musakwa & Mokoena (2018) have proposed the use of non-motorised transportation to serve commuters for the first and last mile of their trip. Perhaps the introduction of non-motorised transportation routes to link with the existing Gaibus network will lead to an improvement in the robustness of the network, whilst also meeting the needs of the commuters.

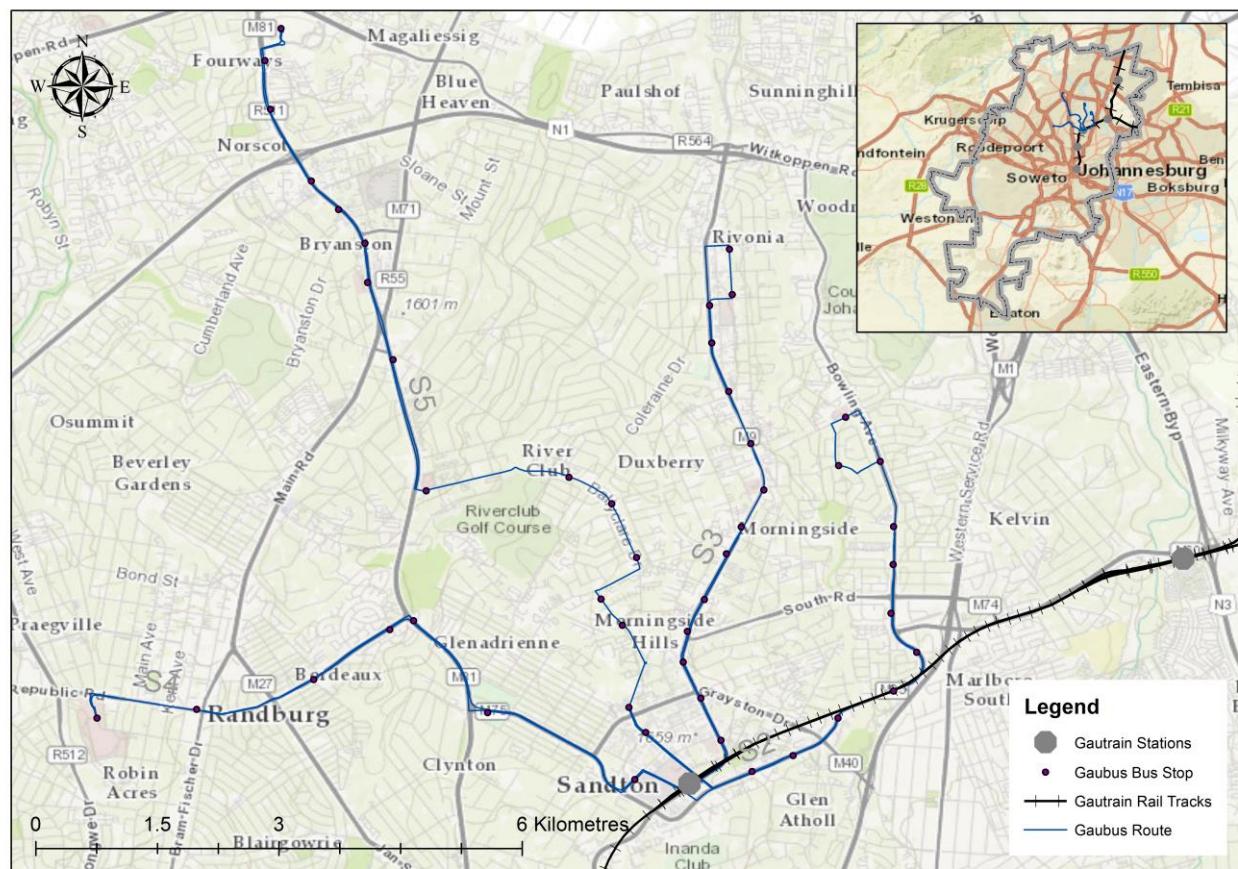


Figure 1: Study Area

### 3.2 Data Collection and Analysis

Using an explorative methodology, the authors collected spatial data in the form of the Gaubus route network, bus stops and cycling data from Strava for the year 2014. Strava is an information service that provides GPS tracked cycling activities, which can be computed in order to make detailed analysis of cycling data. The analysis of the cycling data recorded from by Strava users can be used to understand cycling patterns in different locations for varying spatio-temporal analysis. For the study, the authors used Strava data to determine which locations have the most commuting cycling trips and at which time of day.

To determine the influential nodes of the Gaubus routes, used the following mathematical notation to determine the betweenness centrality:

Let  $G$  be a graph; then the betweenness centrality ( $b_c$ ) of a given vertex  $r$  is defined as a sum over pairs of  $(v; w)$  of vertices other than  $r$ , counting for each pair the fraction  $b_{vw}(r)$  of undirected shortest paths between them that pass through  $r$ . After determining the betweenness centrality we proposed the introduction of new routes or short-cuts to improve the network betweenness centrality. These new routes developed with the following constraints:

- No new edge should be more than 5km
- No edge should be introduced if there is an existing edge between nodes.
- Travel time will be used as the edge weight

### 4. RESULTS AND DISCUSSION

Data collected from Strava from January to December 2014, reveals most cycling activities occur in from the early morning at 4am to 6am, these trips were recorded as being recreational trips (see figure 2). For this study recreational trips were defined as trips made for pleasure and recreation. Typically such trips are not focused on minimising the total travel time but are aimed to optimise recreational enjoyment. Hence most recreational cyclists in Johannesburg prefer cycling in the morning before they go to work or before they start with their day to day trips and in the afternoon when they are done with their day to day trips.

From 6 am to 8am morning peak most the trips were commuting trips. Commuting trips can be defined as trips periodically recurring trips usually between two points of interest namely home and work. In the evening however only a few commuting trips were recorded, this could be due to that most roads are generally congested and most cyclist usually prefer to avoid routes with high vehicular traffic. Going forward with the study we only focused on commuting trips, as most Gaubus commuters utilise the Gaubus for commuting trips.

For the Sandton area and surrounding neighbourhoods we then assessed the number of commuting cycling trips that occur, near or within close proximity of the Gaubus routes, this which was aggregated as shown in figure 3. Most cycling trips occur near the S4 and S5 routes. The two routes are characterised by office parks, shopping centres, residential properties and business

offices. Also given how the slope on these two routes is gentler than on the S2 and S3. This is also true for other parts of the city as cyclist prefer gentle slopes for commuting trips, and steep slopes for recreational trips. With these results in mind,

we proposed the introduction of new routes or edges to link the various Gaibus stops. But before introducing these new edges we computed the betweenness centrality of the existing network (See figure 5)

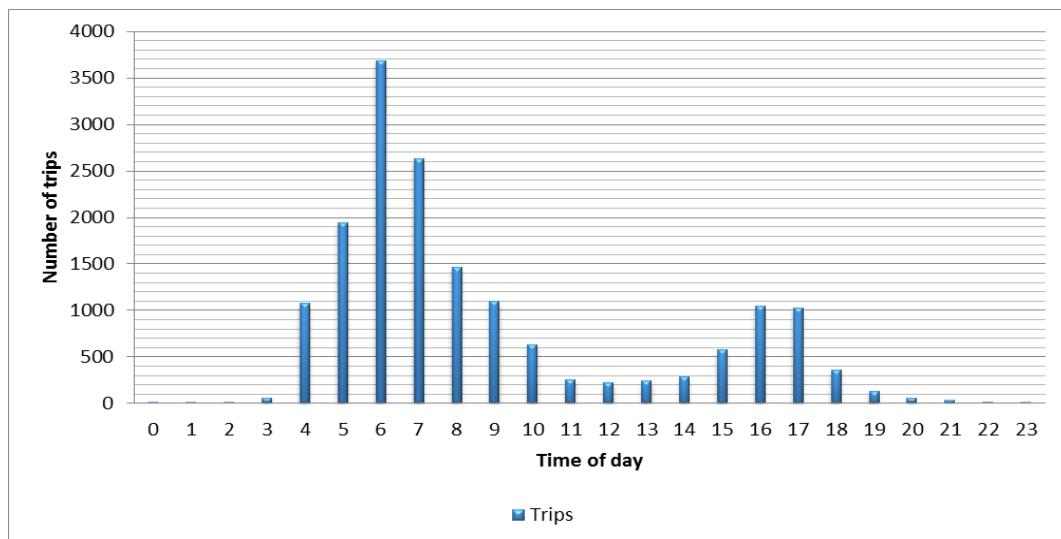


Figure 2: Cycling trips per time of day

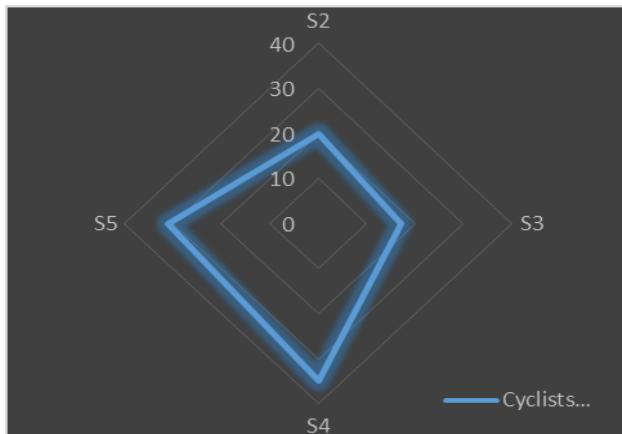


Figure 3: Cycling trips per Gaibus route

The existing Gaibus network influential nodes are currently located near the Gautrain station. These nodes are essential in that commuters can use these to easily switch between the various Gaibus routes. Also bus stops near Ferndale have a high betweenness centrality. This could be due to the close proximity other bus stops, hence commuters have a short distance to walk from their points of interest to the bus stop, hence making these highly accessible nodes in the network.

However besides nodes located in close proximity to the Gautrain station, it is not possible for commuters to switch to the other Gaibus routes. There is hence a need to increase the network robustness and ensure the connectivity in the network is improved. To ensure efficient spread of services and a balance of demand and supply, identifying a minimum set of routes to enhance influential nodes ability improve connectivity is crucial. Using Prim's Algorithm (Moret & Shapiro, 1991), we determine the minimum number of routes to be added to ensure a  $k$ -edge-connected network.

Figure 5 reveals the influential nodes/ bus stops based on the betweenness centrality after introducing new edges. In addition, the higher order influential nodes that appear in red can be used as locations to link the Gaibus to other existing mobility modes besides only non-motorised systems, such as the Rea Vaya (a BRT system in Johannesburg) to service the southern parts of the city, as currently the Gaibus route network only services the northern parts of the city of Johannesburg. Also through the use of an aggregate, the results were normalised to ensure any node that has a direct connection to all other nodes would score 10 and one with no connection would score 0.

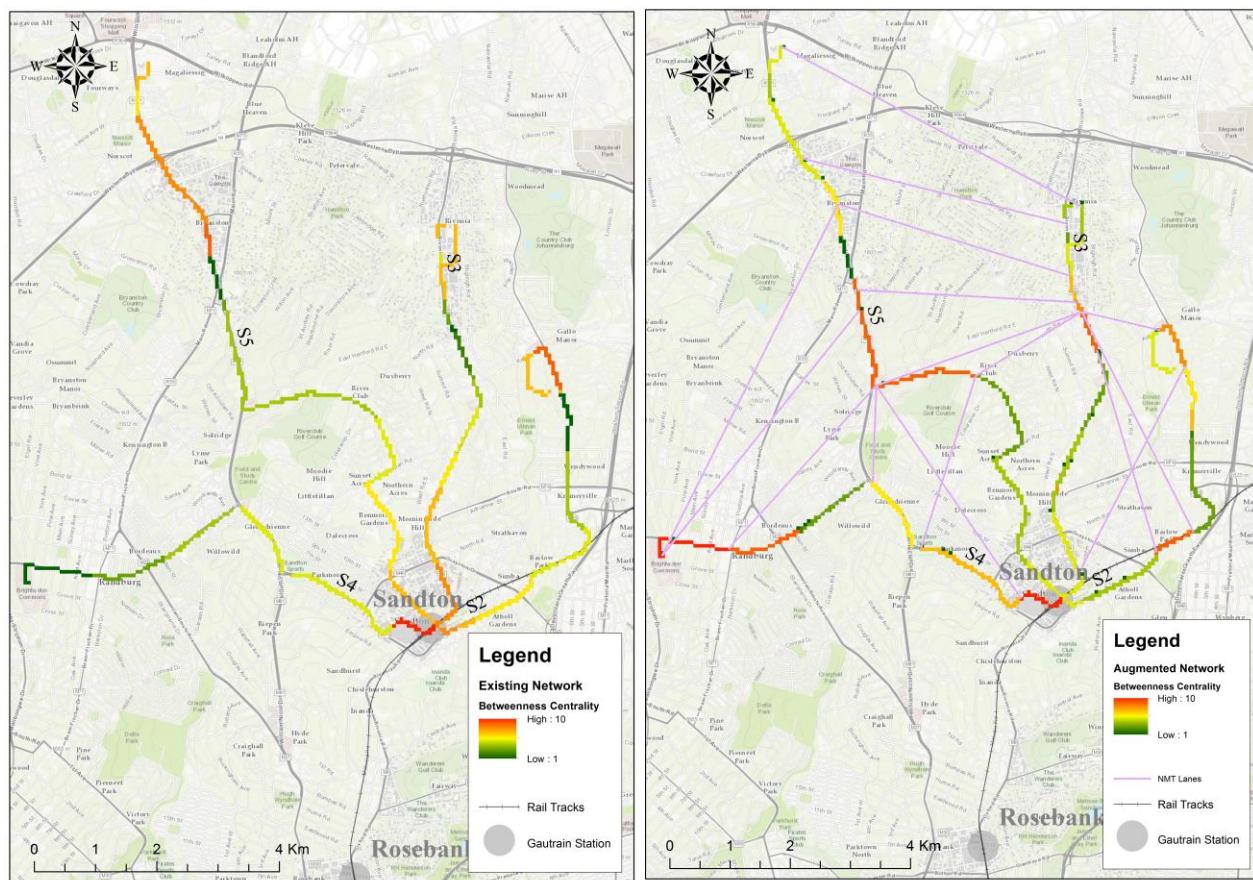


Figure 5: Betweenness centrality of Existing and Augmented Gaubus network

## 5. CONCLUSION

While there have been efforts to promote the use of non-motorised transportation in the city, such as the recently developed cycling infrastructure in the city. There has been little research to guide how to integrate non-motorised transportation with other mobility networks, as currently the two systems are not integrated, hence fail to share synergies.

The paper hence sought to propose how to intergrate these two systems. This intervention, is a viable solution, has it a cost effective means of enhancing the existing mobility network, as the construction of new bus routes could prove expensive. Also given the social and environmental merits for the city to opt for non-motorised transportation, which in turn will lead to a better quality of life for citizens in Johannesburg.

Regarding future research, the authors, propose analysing the centrality of mixed mobility networks, such as for two public transportation providers. This could be used to identify how to integrate the mobility network of the city, as currently the existing public transportation providers are not spatially integrated.

## ACKNOWLEDGEMENT

The research in this paper was funded by the National Research Fund (NRF), South Africa, Grant number 110778.

## REFERENCES

- Ashton, K., 2009. That ‘internet of things’ thing. *RFID journal*, 22(7), pp.97-114.
- Bang-Jensen, J., Gabow, H. N., Jord’an, T., Szigeti, Z. 1999 Edge-connectivity augmentation with partition constraints. *SIAM J. Discrete Math.*, 12(2), 160–207
- Chen, H., Chiang, R.H. and Storey, V.C., 2012. Business intelligence and analytics: From big data to big impact. *MIS quarterly*, 36(4).
- Frank, A. 1992 Augmenting graphs to meet edge-connectivity requirements. *SIAM J. Discrete Math.*, 5(1), 25–53.
- Frank, A., 1999. Connectivity augmentation of networks: structures and algorithms. *Mathematical Programming*, 84(3), pp.439-441.
- Freeman, L.C. 1978. Centrality in social networks conceptual clarification, *Soc. Networks*. 1 215–239.
- Freeman, L.C., 1980. The gatekeeper, pair-dependency and structural centrality. *Qual. Quant.* 14 (4), 585–592.
- Graham, R.L. and Hell, P., 1985. On the history of the minimum spanning tree problem. *Annals of the History of Computing*, 7(1), pp.43-57.

McAfee, A., Brynjolfsson, E., Davenport, T.H., Patil, D.J. and Barton, D., 2012. Big data: the management revolution. *Harvard business review*, 90(10), pp.60-68.

Moret, B.M. and Shapiro, H.D., 1991, August. An empirical analysis of algorithms for constructing a minimum spanning tree. In Workshop on Algorithms and Data Structures (pp. 400-411). Springer, Berlin, Heidelberg.

Moyo, T & Musakwa, W (2016). USING CROWDSOURCED DATA (TWITTER & FACEBOOK) TO DELINEATE THE ORIGIN AND DESTINATION OF COMMUTERS OF THE GAUTRAIN PUBLIC TRANSIT SYSTEM IN SOUTH AFRICA. ISPRS Annals Of The Photogrammetry, Remote Sensing And Spatial Information Sciences, Volume III-2, 2016 XXIII ISPRS Congress, 12–19 July 2016, Prague, Czech Republic. doi:10.5194/isprsannals-iii-2-143-2016. pp 150

Moyo, T., Musakwa, W. and Mokoena, B.T., 2018. AN ANALYSIS TO INVESTIGATE SPATIAL COGNITIVE FACTORS WHICH INFLUENCE CYCLING PATTERNS IN JOHANNESBURG. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, 42.

Narula, S.C. and Ho, C.A., 1980. Degree-constrained minimum spanning tree. *Computers & Operations Research*, 7(4), pp.239-249.

Pettie, S. and Ramachandran, V., 2002. An optimal minimum spanning tree algorithm. *Journal of the ACM (JACM)*, 49(1), pp.16-34.

Qi, X., Fuller, E., Luo, R. and Zhang, C.Q., 2015. A novel centrality method for weighted networks based on the Kirchhoff polynomial. *Pattern Recognition Letters*, 58, pp.51-60.

Sabidussi, G. 1966. The centrality index of a graph, *Psychometrika* 31 581–603.

Wu, X., Zhu, X., Wu, G.Q. and Ding, W., 2014. Data mining with big data. *IEEE transactions on knowledge and data engineering*, 26(1), pp.97-107.

Zanella, A., Bui, N., Castellani, A., Vangelista, L. and Zorzi, M., 2014. Internet of things for smart cities. *IEEE Internet of Things journal*, 1(1), pp.22-32.