The potential of social media to demarcate the catchment of commuters of the Gautrain public transit system

Abstract
The paper systematically analyses the phenomena of being smart in light of the growth of information and communications technology being incorporation into urban development. In a setting where technological advancements are taking hold of every aspect of peoples' lives this paper explores the potential of geographical location data derived from social media platforms can be used to demarcate the catchment area of the Gautrain. The study demonstrates how social media data can be used to analyse the variations in the catchment area of the various nodes with the Gautrain system, with the use of word counts and linguistic measures to interpret posts made on web 2.0. Consequently the users' opinions about the Gautrain were used to interprete the variations in volumes of commuters over a 6 month period. The results of the study will assist in identifying areas of potential expansion and areas in need of intervention, hence highlighting points of interest within Gauteng catchment area.

KEYWORDS: Web 2.0, Gautrain, Commuters, Location data.

1 INTRODUCTION

Contemporally the phenomena of being smart, is taking hold of planning systems, from the business world to city planning. In a setting where technological advancements are taking hold of every aspect of peoples' lives, with more and more people beginning to incorporate these into their daily lives. Phenomones such as smart phones, smart cars, and smart cities, have become a norm which is deeply entangled into the World Wide Web. Accordingly technology seems to have leapfrogged greatly from the olden days were planning could only be reactive. As it is now possible to predict the where and when road congestions will occur, just by analysing the flow of traffic.

But what is it to be smart, does it only involve decision making or the use of advanced technology? Kling & Pozdnoukhov (2013) define a smart city as one were investment in human, social capital and traditional (transport) and modern (ICT) communication infrastructure fuel sustainable economic development. Moreover this would lead to a better quality of life, notwithstanding the value of wise management of natural resources, through participatory governance. From the works of Khan (2014) the smart city concept is seen in another light, by creating a trajectory of the development of a city, from the days when people used horse drawn carts to modern times of smart cars. One key factor seems to stands out which has led to continued development which is the continuous technological advancements. Thus for cities to acme, there is a need for planning that is more holistic which incorporates more efficient and sustainable means of advocating for development as shown in figure 1. With regards to the paper, there is currently a gap in knowledge systems on how to use smart systems in transporational planning. Accordingly the paper will seek to identify a means of demactaing trip generation locations that is were commuters are originating from as a means to understand to movement patterns and characterises of the commuters.

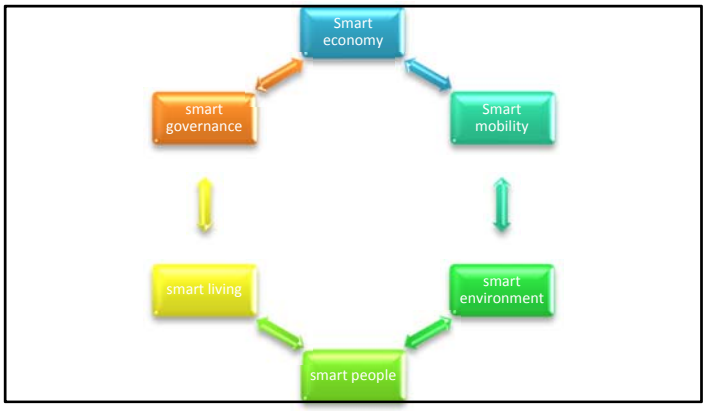

Figure 1: Aspects of a smart city

2 LITERATURE REVIEW

The etymology of the words origin and destination (O-D) relate to a point or place where something begins and a place where a trip ends, therefore the O-D would refer to the point of inception of a trip and where it ends (Oxford University Press, 2015). Bricka et al (2009) further articulates how O-D surveys are a manner of representing spatial interactions of the flows of entities between pairs of geographic locations. This description is further evident in how O-D surveys have historically proven to be a usefully tool which provides a comprehensive depiction of travel patterns within a given location or city region. In transportation planning O-D surveys are carried out to understand the factors that influence the why, how and where of trips generating from an area or towards an area. Gao et al (2012) have further expressed how data collection and analysis has historically been undertaken through travel surveys, traffic counts and positional technologies.

The travel survey based technique involves the collection of raw data based on travel behaviour of the target sample. The data collection methods encompass conducting interviews with key informants, distribution of questionnaires, observations, and video recording with the general nature of this data relating to the demography, movements patterns and socio-economic characteristics. The desired outcome of these O-D surveys is to obtain information relating to displacements of a population, and hence formulating models for planning or restructuring the transport network (Wolf et al. 2003). However this exercise has proven to be time consuming and costly to conduct, with a strain on human resources to conduct interviews and distribute questionnaires. O-D surveys through having their disadvantages still carry the merit of allowing the researcher to obtain valuable information, which is usually obtained only via face to face interaction.

After the new millennium, Global Positioning System (GPS) technologies were used in O-D surveys (Stopher & Greaves, 2009). Accordingly Wolf et al (2001, p1) have also express how GPS data, “provides second-by-second position data with accuracies of three to five metres, as well as highly accurate velocity and time data, introduce a whole new level of comprehensiveness and accuracy to travel surveys”. The utilization of GPS in O-D surveys has brought with it the ability to automatically record trips, whilst recording the geo-location, speed and distance travelled. This which previous could not be accurately record from trip surveys. Also as from the year 2000, the automatic vehicle identification (AVI) systems, were used in O-D surveys (Asakura et al. 2000). The AVI system involves of placing two charge-coupled device cameras positioned separately, with 5-10 kilometres between them. As a motor vehicle passes the first camera, the number plate is recorded and also recorded as it passes the second camera. These recordings are then used to calculate the fraction coefficient, which is the volume of vehicles in a link at a time interval entering from an area at preceding time interval. Combining the fraction coefficient with the traffic volumes observed using roadside traffic counts, the least squares model representing the flows is formulated (Asakura et al. 2000). However the utilization of the AVI system has resulted in an accumulation of big data, and without appropriate analytic

techniques the analysis for the survey becomes time consuming resulting in an accumulation of errors. Over the years this methodology has been improved, through placing numerous cameras at various locations, hence through an analysis of the photographs, the origin and destination of the various motorist are identified over a larger area.

2.1 Emerging Approaches

In the recent past years there has been a rapid growth in the incorporation of social media data in transportation studies. Lorenzi et al (2014) have articulated how a middle class individual's life now revolves around the use of smart-phones. The continued development of smart phones has led to these devices having inbuilt mobile location sensors. Furthermore this has given rise to an increase in development of mobile applications which rely on these location sensors (such as Facebook; Instagram; Strava Metro; Google Maps). The data generated by these applications has the potential of being used to analyse the day to day movement networks of man. However in analysing this data, set backs were identified by Lorenzi et al (2014) in that the information measured was subject to noise and uncertainties, hence leading to imprecise results if these were not excluded in the analysis.

2.1.1 Web 2.0

The notion of Web 2.0 originated in the year 2004, when scholars realised that a second generation of web had emerged, which introduced new possibilities that were previously unattainable. This growth in the web has led to new technological advancements and an incorporation of smart phones and Web 2.0 platforms, such as social media platforms (Facebook, Twitter and Instagram), this could be reflected visually through Mapbox (figure 2.3). The locations of individuals, with smart phones in Gauteng can be seen and also the software allows for a comparison on the type of operating system of these smart phones that is the percentage of android; Iphone; blackberry and other users within the particular area. However while data generated automatically generally has limitations, the sheer granularity has the potential to open new research avenues, which were pervious unattainable from data acquired from field surveys (Bauer et al. 2012). This data could be adopted in the current research in that the geo-tagged data from social media can be mapped to create a continuous spatial density to show the trip generation.

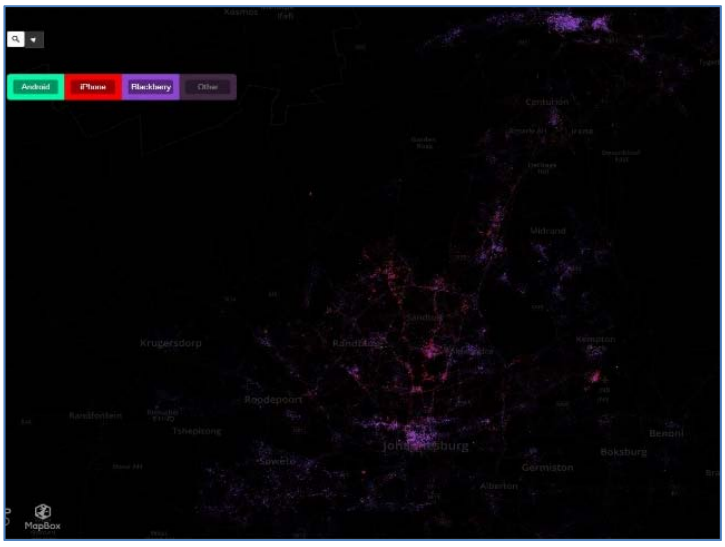

Figure 2: Showing smart-phones in Gauteng; Source Mapbox; 25-06-2015

2.1.2 Big Data

The term big data has become pervasive and yet its definition is still unclear as it has been used to express various notions from social media, large quantities of statistics and real-time data. This advancement has facilitated development, as it has made it possible for people to become more connected, as they can express their views and this information is collected from various sources in a

bid to inform decision making. With regards to urban planning, this has reduced the time taken to respond to service delivery grievances, as the community can easily information council of any grievances via mobile applications, thus bridging the gap between the ordinary citizen and local authorities.

Although big data has a huge potential to assist planners, it has some inherent short-comings in that it can be exploited. Privacy of potentially sensitive information depends on restrictions such as the capability to extract, examine and correlate data (Bryant et al. 2008). In abide to preserve peoples' privacy many platforms have given the user the ability to control what inform can be shared via the internet, and an example is how mobile phones give the user the choose whether to share their location or not. Also big data may include regulated information such as ownership details of property, hence such data should only be used according to set regulations.

2.1.3 Internet of things

The 21st century generation's lives has now become emissary reliant on the internet, as millions of people upload and share information on a daily basis. This internet of things has made it possible for people and devices to stay connected and build relationships over large distances, from meetings being moved from the traditional board rooms to being held via skype or directions being given via tom-tom instead of paper maps. The origins of the notion internet of things (IoT) can be traced back to 1982, when a modified coca cola dispenser (Carnegie Mellon University, 1982), could give information on the temperature of the coca cola inside, this was made possible by the dispenser being connected to the internet (Farooq et al. 2015). Then 10 years later Joy in 1999 hinted on the possibility of device to device connectivity and Ashton (2009) proposed the term internet of things. Shen & Liu (2011) have outlined how IoT is founded on the ideology of allowing the autonomous transfer of useful data through invisibly embedded diverse uniquely distinguishable devices, "which are sensed by sensor devices and further processed for decision making, on the basis of which an automated action is performed" (Farooq et al. 2015, p 1). This expresses how data can now be transfered and analysed almost automatically through machine to machine interaction, with little human interface.

3 OBJECTIVES /RESEARCH QUESTIONS

The Gautrain seeks to expand their rail network system, however before new bus routes and train stations can be chosen, there is a need to first identify the extent of the Gautrain's catchment area. However there seems to be a challenge in identifying the catchment area of the various nodes within the Gautrain transport network system, as their market base is wide and generally dispersed (Musakwa, 2014). The use of geo-location based data has been identified as a possible solution to defining the nodes (catchment areas) of public transit systems (Stopher & Greaves, 2009). As geo-location data can take the form of geo-tagged data from social media and crowd sourced geo-location based data, perhaps spatial analysis of this data can facilitate in deriving a model that represents the nodes that define the hot and cold spots of the Gautrain system. To create a geo-visualized model for identification of trip generation.

4 APPROACH & METHODOLOGY

An explorative research design was adopted for the research, as it allows the research to unpack the potential of web 2.0 data to inform transporation planning. As this study is premised on the utilisation of social media big data to monitor the points of interest of Gautrain users, it becomes evident that privacy concerns arise. The data under analysis carries with it sensitive personal data of the users, that is the user's name and unfiltered tweet or facebook post, the researcher had to ensure that the data was only used for academic and planning purposes.

The social media data, was obtained from Echoecho. Echo-echo is an independent private company that collects and analyses social media data from various web 2.0 platforms. Using sentiment analysis and semantics analysis Echo-echo untangle the big social data to derive meaning from these large quantities of text. The results are captured live and analysed as a means to visualise a brand's 'voice' and where discussions are blossoming or stagnating. Besides the location and social maps, Echo-echo measures segmentation; psychology; language and linguistics; and word clouds and topics.

5 RESEARCH ANALYSIS & FINDINGS

The Gautrain Management Agency has over the years reinforced its presence through the use of social media (Musakwa, 2014). This has led to the Gautrain being able to be at the centre of all social communications surrounding the brand and hence created a direct communication line between existing commuters, potential commuters, stake holders and the GMA as evident in figure 3 and 4. The GMA has integrated social media tactics with the marketing strategy, in a bid to grow the brand as a smart transportation alternative which is safe, reliable and efficient commuters within the Gautrain province (Musakwa, 2014). Furthermore social media has been used to address any adverse or negative perceptions commuters may have about the GMA, this has in turn significantly improved the public's perceptions about GMA as a brand and also lead to more commuters perceiving the Gautrain as a reliable source of transportation. As a means to control public opinions and also facilitate public participation the GMA, has been engaging with the public from the initial development process of the Gautrain. This social media presence that has been growing over the years has led to the Gautrain having more than 120000 twitter follows and having an average feed rate over 75000 posts for the period January to June 2015 (see figure 5), consequently the GMA now uses social media as a mitigation and evaluation tool of their operations. Moreover, opportunities of analysis the social media data exist, as this could led to a greater insight on the Gautrains commuter points of interest.

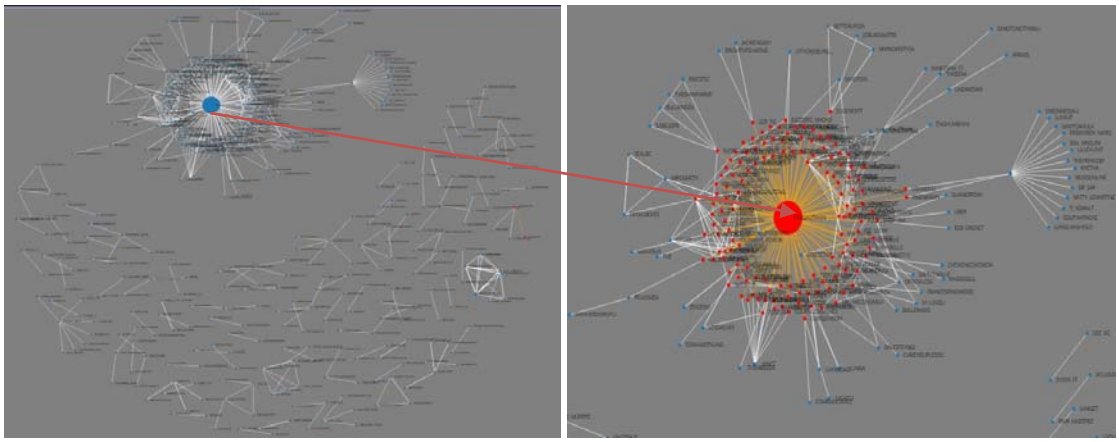

Figure: 3 :Gautain Network Nodes 2015: Source Echoecho, 2015

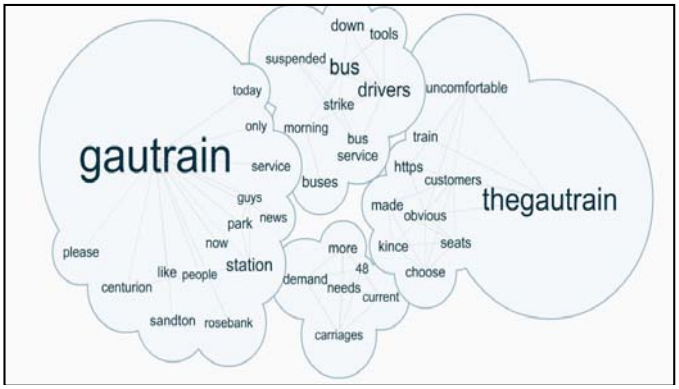

Figure 4 : Gautain Word cloud: Source Echoecho, 2015

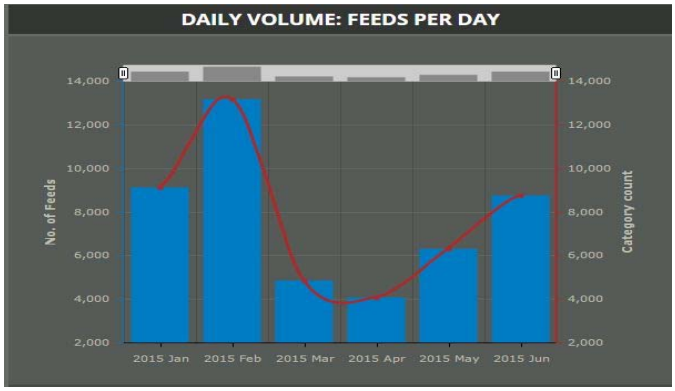

Figure 5: Feeds from January to June 2015: Source Echoecho, 2015

Furthermore with social media, nothing ever seems to die. Once a thread or conversation is initiated, this could either build or harm a brands name. As more and more business have realised the unique value of social media, they are now striving to build their online presence as not only a marketing strategy but also as put of public relations. The public conceptions about the GMA are periodically generated in reports which bring together users' posts on the various social media platforms on the Gautrain (Musakwa, 2014). Through the use of Echocho, these conceptions can be viewed to either show positive or negative emotions, with the posts from January to June 2015 showing 5.6% and 3.9% respectively. These emotions can generally be defined as intense feelings or judgement made with regards to a particular person or brand. Hence the continued measurement of these negative and positive emotions is essential to protect the GMA brand. Consequently the users' opinions about the Gautrain were grouped into the following linguistic measures;

- Space- how much are people talking about the space around them when they refer to the brand.
- Perception- how perceptive is the audience are they open to greater interpretation through seeing and hearing or are they driven by a lower propensity of understanding.
- Motion- words relating to movement how long people are standing in a queue is the queue moving or stationary.
- Time- How much are factors of time being mentioned in updates.

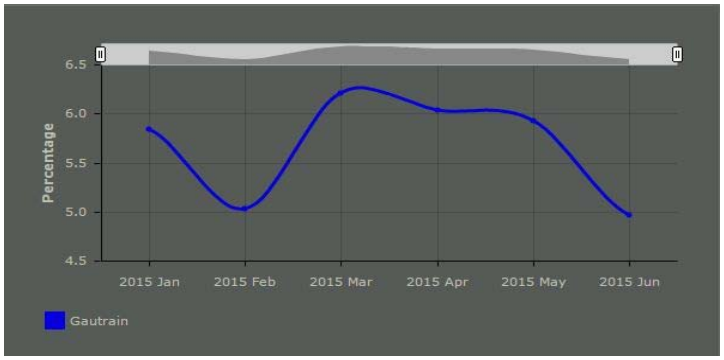

Figure 6: Linguistic measure Space: Source Echoecho, 2015

Pertaining to the linguistic measure of space (see figure 6), the public's options were positive at the beginning of the year, with over 5.5%, however there was a steep fall toward February. This could either be due to an end of a promotion or lack of feed back by the GMA to commuter posts. However as the months progressed, GMA was able to have improved public emotions. This could be due to improved service delivery or improved public relations. Research on the determinants of person perception with regards to service delivery has greatly facilitated the growth of brands. With regards to the Gauteng commuters, their main priority is obtaining a reliable, fast and efficient public transportations means, which would lead to an increase in the number of commuter using the services of a particular brand. Henceforth the occurrence of delays within a system will adversely lead to less favourable impressions being made. There seems to be a need for improvement with regards to the publics' perception of GMA

(see figure 7), as generally the posts have been negative through the months with little improvement around the May period.

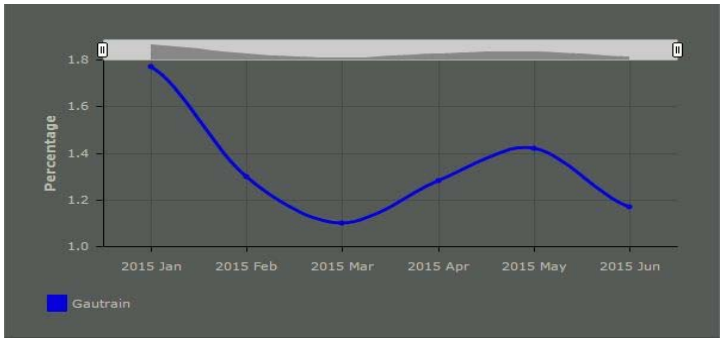

Figure 7: Linguistic measure Perception: Source Echoecho, 2015

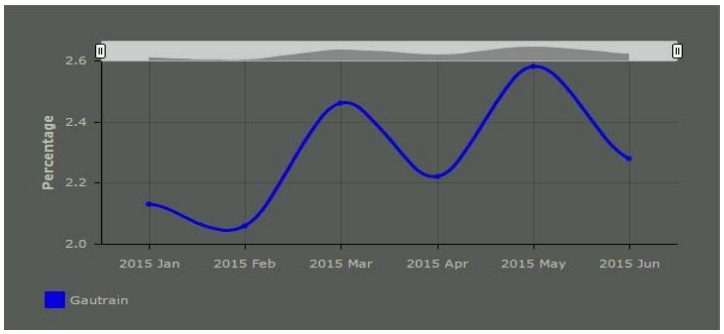

Figure 8: Linguistic measure Motion: Source Echoecho, 2015

Improving the positive magnetism of railway transportation is essential for sustainable development. The manner in which commuters' relate to a brand that is the linguistic measure of motion can either build or hinder the growth of the brand. Studies have shown encumber of waiting seems to be significant in the commuters perceived waiting time and relating to commuter satisfaction. Consequently improving wait times perceptions will lead to better service delivery and commuter satisfaction. With regards to the Gautrain (see figure 8) commuters social media posts at the beginning of the year reflect poor emotions, and with a steep increase towards the March. Hence generally it seems commuters' linguistic measure for motion is generally not constant with sharp variations of time.

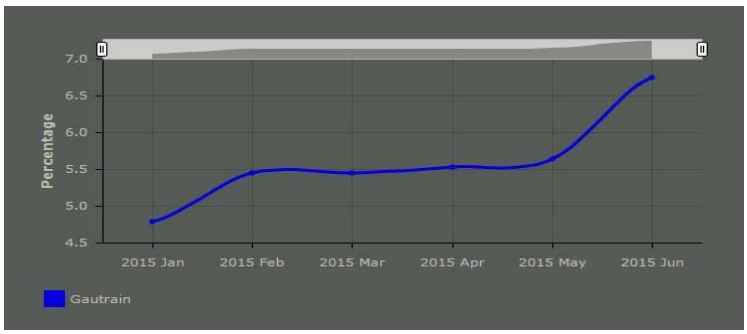

Figure 9: Linguistic measure Time: Source Echoecho, 2015

As historical research on transit users' perception of travelling time has proven to have severe implications on mode choice. Generally commuters' perception is based upon the mode of transportation being able to follow its time table, thus any sudden deviations would lead to adverse emotions. Most commuters of the Gautrain use the railway system as an alternative to driving or using commuter taxis on the highways. Accordingly, the introduction of the Gautrain was ideally meant to ease traffic congestion as the railway system would serve as an alternative to driving or using commuter taxis on the highways. As can be seen in figure 9, commuters linguistic measure on time, has greatly improved as there has been a steady growth in the number of positive comments on social media.

5.1 Word Count: Time series analysis

Posts published within the time period of January 2015 to June 2015 were used to profile of the level of interaction on social media platforms by the participants within the study area. Word counts were henceforward used as the parameterisation to measure the frequency of main topics of interest being discussed during this respective time period. In trying to visualise this data it soon become evident that creating an illustration of all the individual-words in the data set would lead to scribbled visualisation of the big data, hence Tagul was used to present the top words used in the discussions on social media.

The results shown in figure 10 reveal that within the Gauteng city region most of the posts in the data set were comments of users and the Gautrain engaging, as the word 'Gautrain' and 'theGautrain' were mentioned 30586 and 13211 respectively. With sure a high frequency in the mention count this shows the Gautrain's brand has grown over the years, since the Gautrain initial inception in 2010. Meaning the strategies of using social media as a means of building communication linkages and collaborations between the Gautrain and potential commuters are working. "By listening and participating in social media conversations, Gautrain has the opportunity to build authentic, two-way relationships with social media users including mainstream media users and social media opinion leaders who require content to be instantly available, mobile and shared in social networks" Musakwa (2014, p 725). This in conjunction with the mention count frequency measured using Echoecho it becomes lucid that the Gautrain brand has grown over time and that social media interaction is a decisive marketing tool for Gautrain. The first two months of the year saw a steady increase of posts on the social networks, which lead to an increase in commuters using the Gautrain. This growth could be due to social word of mouth that is social media users engaging recommend to other users to use the Gautrain due to satisfaction with the service received.

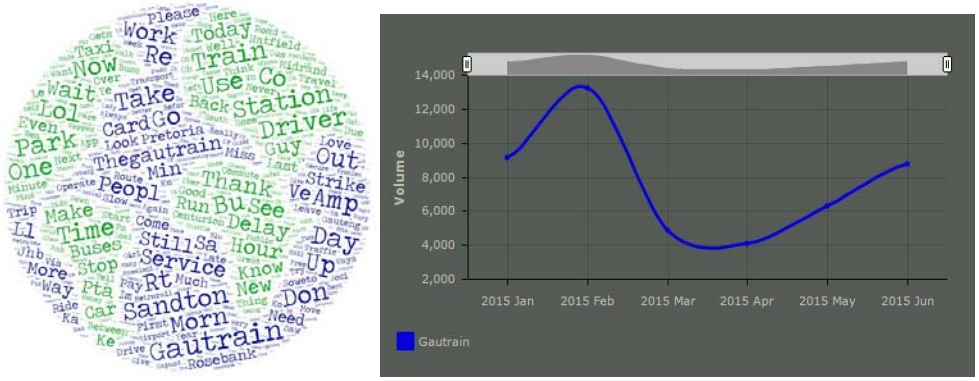

Figure 10: January to June Top words: Gautrain 30586, theGautrain 13211; bus 7105; station 6964; train 4100; sandton 3416; service 2936; no 2847

An analysis on Echoecho's emotion focus reveals attention-grabbing results with positive emotions being at 5.6 % and negative emotions being at 3.9 %. As emotion influences peoples' preference and mode chose, introducing emotion detection allowed for direct unspoken feedback measurement and evaluation. Hence the emotion focus count could be used to explain the sharp fall between February and April. It is safe to say identifying the expressed emotion in the social media posts is challenging, as twitter users are only able to use 140 characters in their posts hence their emotions can end up becoming hidden in posts which do not have explicit reference to key words such as disappointment or happy. Henceforth further analysis using a time series approach for the months needed to be undertaken to analyse for such patterns and determining remedies to improve commuters' perceptions.

The results on the time series analysis for the 6 months reveals that social media interaction between Gautrain and social media users seems to be highest during the middle of January with rapid increases and declines on the days thereafter with the 27th being the last recorded with high volumes as shown in figure 11. The low volumes between 1st and 10th of January 2015 could be due to that most people generally leave the province and go on holiday during the festive season. Consequently less people

would be using the services of the Gautrain, hence the low volumes. The increase in social media posts after the 10th of January also concurs with this hypothesis as generally most people would be returning to work and schools would be opening during this period. The effect of the end of the holidays also seems to spill over on to February as the level of social media post is still high at the beginning of this month as shown in figure 12. However due to the bus driver strike that occurred during this month, the peaks from the 1st to 12th of February could also be a direct result of social media users discussing this. Consequently this lead to negative emotions raising to 4.8% and positive emotions being 6.2%, such an imbalance generally shows dissatisfaction with regards the brand. However with time, the level of interaction between the Gautrain and social media users begins to reduce steadily, due to the half-life of social media posts.

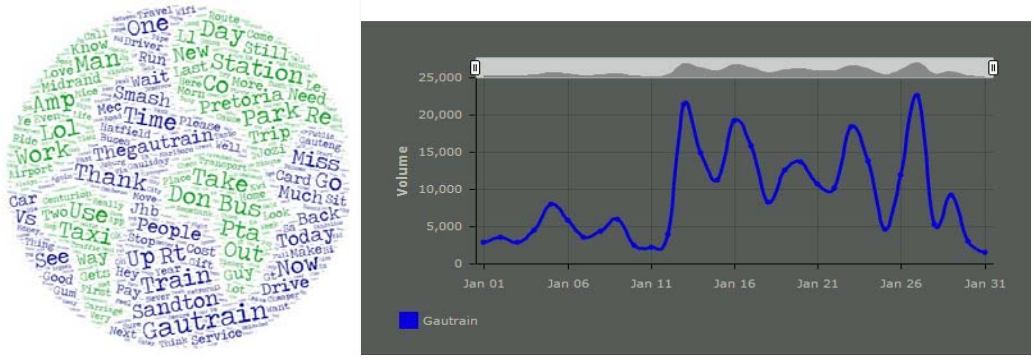

Figure 11: January: Top words: Gautrain 7,343; Station 1905; theGautrain 1714; friend 962; add 961; train 857; bus 791; Sandton; 693; and man 654

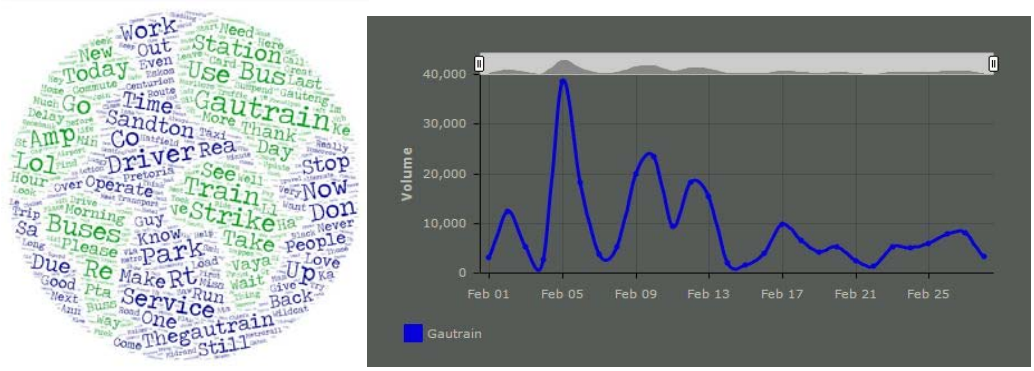

Figure 12: February Top words; Gautrain 8109; theGautrain 4574; Bus 3575; strike 2350; drivers 1938; station 1637; service 1562; train 1087; sandton 1042; no 957

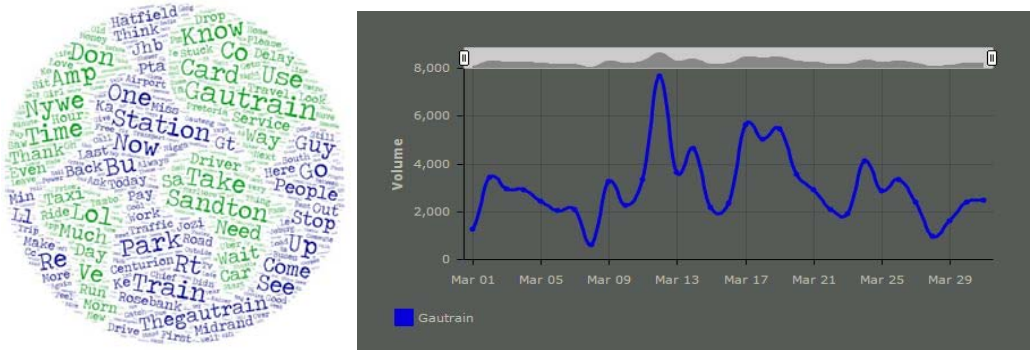

Figure 13: March Top words Gautrain : 3193; theGautrain 1238; station 889; bus 501; sandton 394; get 342; now 342; train 324; like 286

Whilst social media is currently a multifaceted marketing channel, the potential use of this big data represents a gold mine for planners. With regards to the Gautrain, the month of March and April had fluctuating volumes and an increased level of positive emotions being 6.3% and negative emotions

being 4.7% as shown in figures 13 and 14. This could be linked to the quick resolution of the bus driver strike and the Gautrain continuously engaging with the public to ensure all concerns were addressed quickly. However a challenge becomes evident, that is of ensuring a control on the levels of participation between social media users and the Gautrain. Whilst such issues are native to social media it seems a tedious exercise, to expand the life of a social media trend. However history has shown that expanding the life-line of a trend acts as a reinforcing agent which builds a brand and allows it to compete in the market.

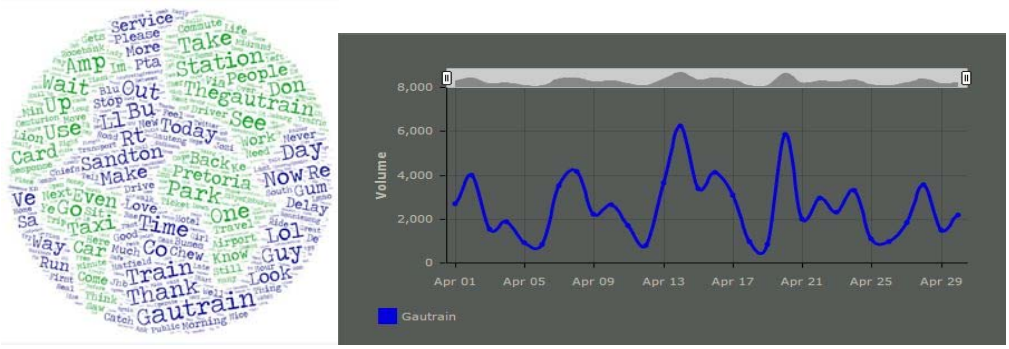

Figure 14: April Top words: Gautrain 2483; theGautrain 1383; station 655; like 429; bus 369; park 316; train 290; gum 282

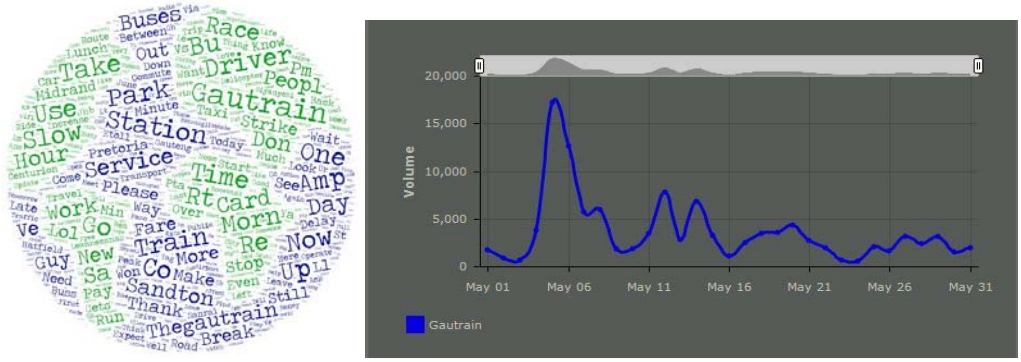

Figure 15: May Top words: gautrain 3951; theGautrain 1815; bus 1428; station 1010; #gautrain 577; drivers 547; sandton 389; up 357; get 354

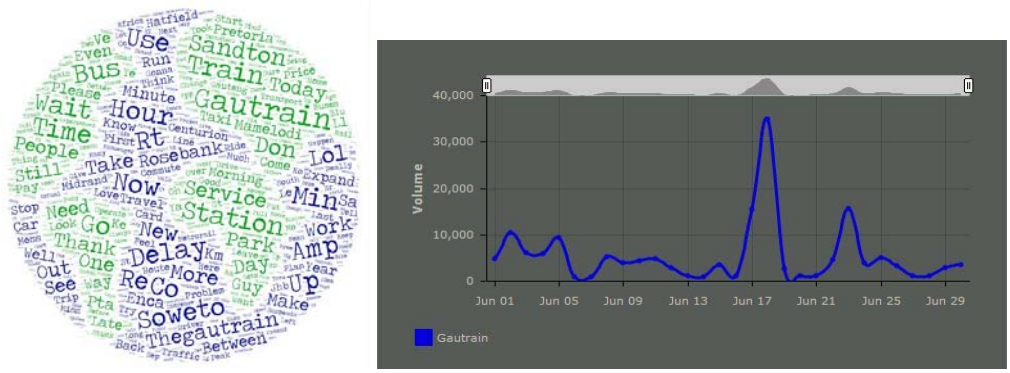

Figure 16: June Top words: gautrain 5507; theGautrain 2487; train 1201; station 868; Soweto 786; #gautrain 765; sandton 691; mamelodi 645; now 559

The involvement of communities in social media has evolved through the years, as public discussion on issues such as strikes and service delivery are now possible through platforms such as Facebook and Twitter. Hence regarding the looming strikes the terms ‘bus and drivers’ had high volumes, during the last months of the study period. Also these lead to an increase in negative emotions rising to 5.0% and positive emotions being 6.3% in May and June seeing a reduction in negative emotions to 4.6% and 5.9% positive emotions as shown in figure 15 and 16. Consequently having such an imbalance in emotions over time could lead to commuters not using the Gautrain as they would lose faith in the

brand, as most users who rely on the Gaubus had to find alternative means to commuter to and from points of interest. In order for the Gautrain not to lose brand loyalty they had to ensure they responded quickly whilst reassuring the public the grievance were being resolved, with minimum disruptions to service delivery.

6 RESEARCH CONTRIBUTION

This examination over the 6 months period demonstrates how social media interaction is a crucial tool in crisis control and brand management. However as shown by the abrupt variations in the emotions of the posts one major drawback of social media then becomes evident, that is posts have a half-life. If companies are unable to continuously adapt, this will affect the company's brand and customer loyalty. This initial analysis of the social media data was used to unpack the pre-existing nature of the crowd sourced data from users living within the Gauteng province and was used to explain the monthly variations in the data set. Furthermore an analysis of the geographical location of social media posts on Echocho shows the Gautrain's geographic presence that is the extent of social media buzz that is generated by the GMA brand. Initially the social media data was viewed at a holistically level for the year 2015. The population count for the social media posts was subdivided as follows for visualisation purposes:

- Blue representing 10 posts or less
- Yellow representing 100 posts or less
- Red representing 1000 posts or less
- Pink representing 10 000 posts or less
- Purple representing 100 000 posts or less

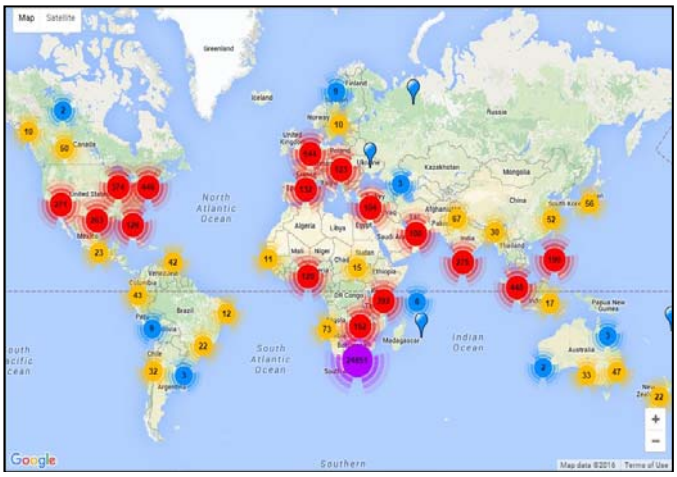

Figure 17: World twitter and facebook posts: Source Echoecho, 2015

At a worldwide scope (as shown in figure 17) the catchment area of the Gautrain seems to cover the Southern parts of North American, with over 1500 users; West and South Europe having over 2000 users and Austria having the least number of users engaging on topics involving the Gautrain. The possible node that serves these users could be the Gautrain station at OR Tambo international air-port as most of these users would ideally use the air-port to enter the Gauteng. Whilst at a regional level in South Africa (as shown in figure 18), the Gauteng province has the most users engaging with the Gautrain, this may mainly be due to the Gautrain directly serving people in this province. The Western Cape was the next province with the highest users engaging with the Gautrain, this could be due to the close historical economic relationship between the two provinces as these two have the highest contribution to the country's gross domestic product. The Limpopo province has also a large number of users who engage with the Gautrain, this may either be due to two reason, namely the province serves as an entry point (road) for the rest of the African in South Africa or due to the growth in economic activities in Limpopo, this has led to growth in the socio-economic relationship between the provinces.

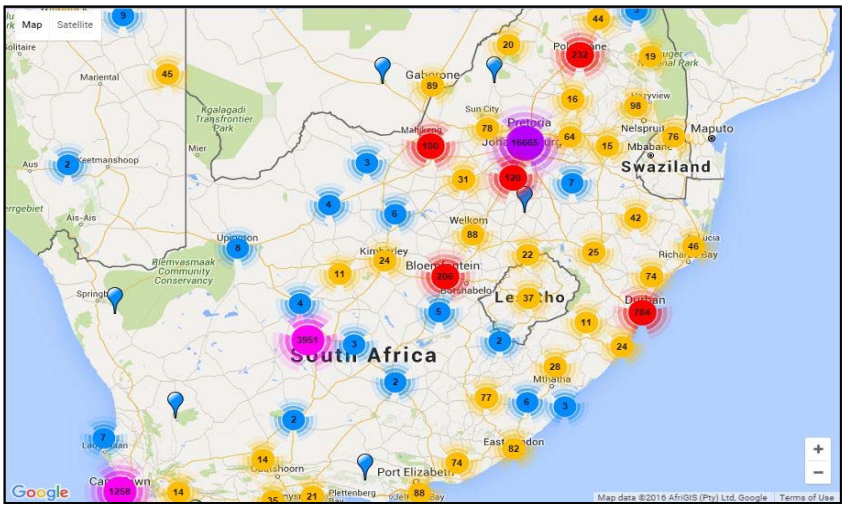

Figure 18: S.A twitters and facebook posts: Source Echoecho, 2015

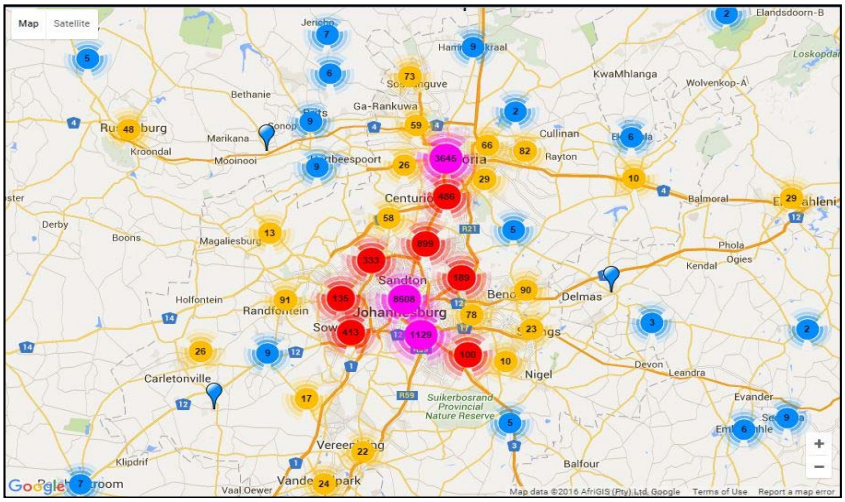

Figure 19: Gauteng twitter and facebook posts: Source Echoecho, 2015

Having an efficiently and effectively run public transportation system is a crucial and indispensable factor for the any developing city region. However as the provision of public transportation is a multifaceted process, with intertwining elements such as culture, politics, finances and shareholder interests. Smart means of monitoring and mitigating the challenges faced in the provision of public transportation need to continuous be developed. The Gauteng city region is likewise faced by this challenge, with the region being the economic hub of South Africa, this has greatly affected the operation of the Gautrain, as more and more people require a fast and reliable transportation means to traverse in and out the metropolitan cities. This is evident by the numerous social media posts made at the existing nodes such as at Park Station; Sandton and Pretoria as shown in figure 19. Also this poses a challenge to clearly demarcate the catchment areas of these nodes whilst visualising the areas of hot and cold spots to be identified for future developmental projects.

7 RESEARCH LIMITATIONS

When using the assumption that people who post about the Gautrain are either existing or potential commuters, then yes the model does show the sphere of influence of the Gautrain. However in the real world situation, this may not always be the case. As there is currently no clear manner to ensure that everyone who posts about the Gautrain is a current or potential commuter. Also as the research follows an explorative approach, a means of validating the results is needed. Hence perhaps in futher studies an experiemental approach could be adopted, which would ensure there is a control and experiement as a

means to check the validity of the results of the model whether it does accurately represent the variations existing in the nodes.

8 DISCUSSION & CONCLUDING REMARKS

In using web 2.0 data as the input for transportation planning many obstacles still need to be tackled. Currently knowledge gaps still exist, with regards to how to exploit this data to inform planning. Hence a bridge is still needed to link what is available (big data) and what could be done (planning). As no model can be used as a one glove to fit all situations, a need to continuously develop and renew planning models exists. Critics have highlighted that the data used in the analysis is only a representation of twitter and facebook users, and that not everyone uses these platforms, as some prefer to use Instagram or Flicker. However as the GMA has shown keen interest in focusing their marketing resources on these two platforms (Musakwa, 2014), the assumption that commuters will generally respond or post were they will receive feedback comes into play. As a result the social media big data from twitter and facebook presents the majority of the Gautrain's sphere of influence. Consequently this reveals how the analysis is deeply imbedded in the utilization of big data for the model to be effective. An understanding of correlation between the observed social media data points is of great importance, as noted in the word count and linguistic measure. However this also is dependent largely on time that is the duration of the study. As over time, more insight on the data can be gathered, hence a seasonal or yearly co-relational analysis can be developed to identify the key factors that influence.

REFERENCES

Asakura, Y., Hato, E. & Kashiwadani, M., 2000. Origin-destination matrices estimation model using automatic vehicle identification data and its application to the Han-Shin expressway network. Transportation, pp. 419-438.

Ashton, K., 2009. That Internet of Things' thing. RFID Journal

Bauer, S. et al., 2012. Talking Places: Modelling and analysing lingustic content in Foursquare. pp. 1-10.

Bricka, S., Zmud, J., Wolf, J. & Freedman, J., 2009. Household Travel Surveys with GPS. Transportation Research Record. Journal of the Transportation Research Board, 2105(5.1), pp. 51-56.

Bryant, R., Katz, R. & Lazowska, E., 2008. Big-Data Computing: Creating revolutionary breakthroughs in commerce, science and society. Washington, DC: Computing Community Consortium.

Carnegie Mellon University, School of Computer Science, 1982. "The Only Coke Machine on the Internet".

Farooq, M., Waseem, M., Khaiji, A. & Kamal, T., 2015. A Review on Internet of Things (IoT). International Journal of Computer Applications (0975 8887), 113(1), pp. 1-7.

Gao, S., Yang, J., Yan, B., Hu, Y., Janowicz, K & McKenzie, G. (2012). Detecting origin-destination mobility flows from geotagged tweets in greater Los Angeles area.

Khan, S., 2014. Historical evolution of Durban's public transport system and challenges for the post-apartheid metropolitan government. New Contree, Volume 70, pp. 173-194.

Kling, F. & Pozdnoukhov, A., 2013. When a City Tells a Story: Urban Topic Analysis. pp. 1-10.

Lorenzi, D., Vaidya, J., Chun, S., Atluri, V. (2014). Enhancing the government services experience through QR codes on mobile platforms. Government Information Quarterly, 31: 6-16.

Musakwa. W, 2014. The use of social media in the Gautrain in Gauteng Province, South Africa: analysis and lessons learnt. REAL CORP 2014 Proceedings. . Page 721-727

Oxford University Press, 2015. Oxford Dictionaries. [Online] Available at: http://www.oxforddictionaries.com/definition/english/origin [Accessed 19 May 2015].

Shen, G. & Liu, B., 2011. "The visions, technologies, applications and security issues of Internet of Things", in E-Business and E -Government (ICEE). pp. 1-4.

Stopher, P. & Greaves, S., 2009. "Missing and inaccurate information from travel surveys – pilot results". 32nd Australasian Transport Research Forum.

Wolf, J., Oliveira, M. & Thompson, M., 2003. Journal of the Transportation Research Board. "Impact of underreporting on mileage and travel time estimates: Results from global positioning system-enhanced household travel survey transportation research record", Volume 1854, pp. 189-198.