Mobile GIS: A tool for informal settlement occupancy audit to improve integrated human settlement implementation in Ekurhuleni, South Africa.

B. T. Mokoena*, W. Musakwa

*Dept. of Town and Regional Planning, University of Johannesburg, Doornfontein, Johannesburg, South Africa – Alumni
baleseng.mokoena211@gmail.com

Dept. of Town and Regional Planning, University of Johannesburg, Doornfontein, Johannesburg, South Africa –wmusakwa@uj.ac.za

KEY WORDS: Mobile GIS, occupancy audit, informal settlements, integration, Ekurhuleni

ABSTRACT:

Upgrading and relocating people in informal settlements requires consistent commitment, good strategies and systems so as to improve the lives of those who live in them. In South Africa, in order to allocate subsidised housing to beneficiaries of an informal settlement, beneficiary administration needs to be completed to determine the number of people who qualify for a subsidised house. Conventional methods of occupancy audits are often unreliable, cumbersome and non-spatial. Accordingly, this study proposes the use of mobile GIS to conduct these audits to provide up-to-date, accurate, comprehensive and real-time data so as to facilitate the development of integrated human settlements. An occupancy audit was subsequently completed for one of the communities in the Ekurhuleni municipality, Gauteng province, using web-based mobile GIS as a solution to providing smart information through evidence based decision making. Fieldworkers accessed the off-line capturing module on a mobile device recording GPS coordinates, socio-economic information and photographs. The results of this audit indicated that only 56.86% of the households residing within the community could potentially benefit from receiving a subsidised house. Integrated residential development, which includes fully and partially subsidised housing, serviced stands and some fully bonded housing opportunities, would then be key to adequately providing access to suitable housing options within a project in a post-colonial South Africa, creating new post-1994 neighbourhoods, in line with policy. The use of mobile GIS therefore needs to be extended to other informal settlement upgrading projects in South Africa.

1. INTRODUCTION

Urbanisation is a phenomenon that has been growing in recent decades in the global south (Turok & Borel-Saladin, 2014), manifesting itself in the form of informal settlements (Freire, et al., 2014). It is said that Africa is urbanising rapidly from a 15% growth in 1960 to 40% in 2010 (Freire, et al., 2014). According to UN-Habitat (2010), 60% of Africans will be living in cities by 2050, and triple in the forthcoming 50 years. This will change the form of regions and mandate policy makers to maximise on urbanisation towards sustainable and inclusive growth (UN-Habitat, 2015). The New Urban Agenda (UN-Habitat, 2015) furthermore calls for policy makers to shift their mind sets from viewing urbanisation as a problem but rather as a tool for development. The current model of urbanisation is not sustainable, nor is it adequate for social and economic prosperity. The strategic and integrated approach taken in the new Strategic Plan for 2014-2019, recommends that we apply a more systemic approach that goes beyond addressing the symptoms of urbanisation, but linking urbanization and human settlements to sustainable development by focusing on the prosperity, livelihoods and employment of the greater population (UN-Habitat, 2015). This will be done by paying attention to the basic needs of the millions of people living in poverty within towns and cities, as well as urban slums (UN-Habitat, 2015).

Unlike Africa, other parts of the world’s urbanisation patterns have been led by industrialisation and economic development; defined by the increase of productivity from proximity and the concentration of activities, confirming’s a relationship between per capita income and urbanisation (Freire, et al., 2014). Africa’s urbanisation patterns however indicate a migration of lower level income groups and less investment of infrastructure (Freire, et al., 2014). This poses a challenge for cities that are unable to accommodate large population influxes that ensure favourable living environments (Freire, et al., 2014).

Migration and urbanisation as described in the Migration and Urbanisation in South Africa Report (2006) is often observed as the result of raptures of economic exploitation, political tension, environmental disasters or violence. South Africa has a particular history, which informed its migration and urbanisation patterns over the past few decades, known as apartheid. The apartheid system historically endeavoured to restrict and control the population movement as well as its settlements patterns in the rural areas or Bantustans (Turok & Borel-Saladin, 2014). Thus tyrannical laws were enforced on people such as the well-documented population control, Group Areas Act and pass law that caused impermanence in the urbanisation process (Turok & Borel-Saladin, 2014; Harrison, et al., 2008; Zuma, 2013). These events resulted in inadequate urban planning in the urban areas which lead urban settlements into sprawling peri-urban areas (Harrison, et al., 2008). This also meant that people were forced to live in ethnically homogenous homelands with limited access to land, resulting in the transition from an agrarian to a cash-based rural economy in order to provide large numbers of labour migrants (Zuma, 2013).

This unique phenomenon of urbanisation has affected the housing sector drastically over the years, with the influx of people coming into major cities in South Africa, most of whom find themselves living in informal settlements (Huchzermeyer, 2004). This posing a
challenge for the state in how it deals with housing people in the low income bracket or with no income at all.

2. THE DISCOURSE OF UPGRAADING INFORMAL SETTLEMENTS PRE AND POST 1994

The chronic of informal settlement upgrading policy in South Africa needs to be narrated for us to understand the context, which we find ourselves today. In Marais & Ntma’s (2013) reflection through the lens of (Huchzermeyer, 2004), discussions of this unfolds in a manner that suggests that until recently, policy had not been developed for the new post-apartheid government for informal settlement upgrading. “Breaking New Ground”: A comprehensive plan for the development of informal settlements was developed only in 2004, after the realisation that the housing subsidy scheme alone was not sufficient on the large inflow of black people in informal settlements (Huchzermeyer, 2004). The upgrading of informal settlements was strongly opposed during the apartheid era, further enforced by the demolishing and forced removals of these settlements (Platzky & Walker, 1985). The control of urbanisation which was abolished in 1985, saw the influx of people in many cities (Huchzermeyer, 2004). Controlled urbanisation was driven by the then government to control land development which meant that land was largely available for middle-income groups or white people and not for low-income households or black people (Huchzermeyer, 2004). The land that was made available for black people was mostly funded by loans provided by the National Housing Commission. The large influx of people by the early 1990 in the form of land invasions on open land across South Africa (Seekings, 1991) put pressure on the apartheid government to embark on large-scale upgrading of informal settlements (Marais & Ntema, 2013). The Independent Development Trust (IDT) was then established around that time providing site-and service and settlement upgrading to approximately 100 000 households across South Africa providing water, sanitation, electricity and ownership, packaged in the capital subsidy of R7 500 per household (Marais & Ntema, 2013). The upgrading process by the IDT, Huchzermeyer (2004) agues and critics, where driven by the market orientated development approach (neo-liberalism) with very minimal community participation as well as a technocratic approach of a one size-fits-all, putting people on the periphery in greenfield projects. Stating that the new post-apartheid government had followed to a large extent in the same footsteps with its new housing policy (Huchzermeyer, 2004). With a once off capital subsidy, at the time, of R15 000 for people earning R0-800 per month focusing on the lowest income earners, working on a sliding scale, and providing R9 000 for people earning R2500-R3500 per month (Huchzermeyer, 2004). Breaking New Ground together with its Informal Settlement Upgrading Programme introduced a concept of promoting social inclusion through community based subsidies and participatory planning of layouts, poverty eradication by securing tenure and assistant livelihoods through well located land and reduction of vulnerabilities through providing economic and social facilities (Huchzermeyer, 2006). These objectives focused on the rehabilitation of land for upgrading, moving away from the overemphasis of Greenfield development as well as focusing on mixed-income groups rather than just on low-income groups (Huchzermeyer, 2006; Marais & Krige, 1999). Emphasising rather on the need for better-located mixed-income and mixed-use housing projects.

3. ENUMERATION OF INFORMAL SETTLEMENTS:

Enumerations process is an important part of informal settlement upgrading, especially when there is no data that exists for a community, but also useful to have regularly updated data. The purpose of enumeration is to provide basic information on all individuals and households (Karanja, 2010) in a particular settlement. This section highlights some of the challenges faced in different enumeration processes.

An informal settlement was enumerated in Nairobi, Kenya, organized by the inhabitants, to examine the detail and scale, depth and nature of poverty within the community (Karanja, 2010). The experience gathered from this enumeration demonstrated that the way in which information is gathered as well as who gathers it, is as important as the information itself (Karanja, 2010). This is because it influences not only the quality and detail of the data collection and its verification but the nature of the occupant’s involvement with planning and implementation of informal settlement upgrading (Karanja, 2010). The challenges and pressures faced in this enumeration process where from people that where not residents of the community who wanted to also get plots for friends or family members, who reside in other communities or for adult children living with them (Karanja, 2010).

The mapping and enumeration of informal Roma settlements, located in Serbia Southeastern Europe were conducted with the attempt to response to some of the housing challenges confronting the communities (Vuksanovic-Macura, 2012). This exercise was conducted several times over a number of years by non-governmental organisations and Roma associates. The settlements, as described by Vuksanovic-Macura’s (2012), are characterised by very poor living conditions because of historic social exclusion and intolerance faced from formal communities (Vuksanovic-Macura, 2012). In her paper she also indicated that the difficult housing challenges that these settlers face had not been adequately addressed by national and local authorities. With the aim of assisting communities, mapping and enumeration initiatives were undertaken over the years for settlement upgrading, socio-economic empowerment purposes, to name a few (Vuksanovic-Macura, 2012).

The method used for enumeration and mapping were questions on household information and about the physical house and plot. The mapping component described the extent of the settlement spatially. Photographs were also taken. The data gathered was entered into a spreadsheet programme and the maps were digitized, forming a database (Vuksanovic-Macura, 2012).

---

As a final reflection the paper highlighted that while the information gathered by the NGO’s was good for the initial phase, given the limited scope and funding, it was insufficient if procedures based on the law was required. This is because official documentation is needed should an upgrading require formal administrative procedures by way of producing an urban plan, legalizing illegally constructed houses and initiating court procedures to transfer ownership (Vukanovic-Macura, 2012). Having said this Geographic Information Systems (GIS) according to (Chitekwe-Biti, et al., 2012) has transformed the extent in which information can be used for planning purpose across the world. Internationally there has been a great development of skills allowing communities to effectively and accurately link information collected during household level enumerations with mapping. Therefore GIS has created the possibility of linking social data with spatial data (Chitekwe-Biti, et al., 2012).

Consequently, the aim of this study is to use smart mobile GIS, to enumerate and conduct an occupancy audit in Ulana informal settlement, Ekurhuleni, South Africa. The remainder of the paper is structured as follows; the next section provides a description of the study area, Ekurhuleni, followed by the methodology, which describes how the occupancy audit was conducted. Lastly the results and discussion are presented

4. STUDY AREA: EKURHULENI AND ULANA SETTLEMENT

The Ekurhuleni Region is a Metropolitan made up of an amalgamation of nine towns together with its townships3, historically known as Eats Rand. Each town consisting of suburbs, industrial areas and the black residential areas attached to them (Bonner, et al., 2012; City of Ekurhuleni, 2013). It has a total surface area of 1 975 km² with a population of 3 178 870 million (City of Ekurhuleni, 2013). It has a strong manufacturing sector and is regarded as the transportation hub of South Africa, having the busiest airport in Africa4 and as such seeking towards branding itself an Aerotropolis. It also has South Africa’s largest railway hub in Germiston, linking it to all major population centres and ports in the Southern Africa region5.

Ulana Settlement is an informal settlement located in Boksburg, Ekurhuleni (Figure 1).

Figure 1: Location of Ulana Settlement

Ulana has a population of 3 092 households according to the recent occupancy audit completed. Ekurhuleni has the second highest number of households living in informal settlements in the country6, with a total of 119 settlements and a population of approximately 160 000 households6. Upgrading and relocating these settlements requires consistent commitment, good strategies and systems in order for the lives of those who live in them to be improved incrementally.

While Enumeration of informal settlements remain an important part of any upgrading project, this paper sets to demonstrate how web based mobile GIS can be used as a tool for enumeration, to foster integrated human settlements. Given the challenges of urbanisation in cities of the Global South, and more specifically within the unique context of South Africa (and Ekurhuleni), the chronicles of upgrading informal settlements pre and post 1994 proved (and still proves) to be challenging.

5. METHODOLOGY

A web-based Mobile GIS tool was used to conduct the occupancy audit in Ulana, as it is often difficult and cumbersome with other techniques. Data collection was conducted from 13 March 2015 to 30 March 2015. A follow up data collection session was done on April 18-19 for occupant’s who were unavailable in March. A GPS was used to collect the location of the dwelling units (DU) whilst an android-based tablet was used to collect attributes of the DU, images as well as the household attributes. This information was stored and synchronised in a server (Figure 2).

During data collection each household and dwelling unit was allocated a barcode to ensure that repetition and redundancy was eliminated. In addition, a drop standard down menus and validation rules were put in place as quality assurance (Figure 3).

---

3 Townships in South Africa often referring to the location where the black people reside
6 EMM: Urban Settlement Grant (2012)
There were minor challenges experienced during data collection such as some members of the households without identification documents. In addition, it was difficult to obtain proof of income document.

These very high densities in Ulana further negatively affects the quality of life as well quality of place for its inhabitants. The living conditions in which informal settlers live are well described by Fanon (1961) who went on to write that ‘It is a world without spaciousness; men live there on top of each other ...built one on top of the other ...a hungry town, starved of bread, of meat, of shoes, of coal, of light ... it is a crouching village, a town on its knees, a town wallowing in mire’. Similarity the dwelling units in Ulana are built close to each other as well as the poor material used exposes them to fire hazards. The use of candles, paraffin or gas in small, crowded spaces, usually results in fires that can quickly spread in crowded areas such as Ulana. Health conditions continue to deteriorate in informal settlements due to the squalid and deplorable conditions. Dewan et al. (2012) have noted that unplanned urban growth can lead to invasive disease infections such as the dengue infection. Furthermore, there have been reports of infectious diseases such as typhoid fever in Gauteng owing to the squalid conditions; meanwhile there has been diarrhoea outbreaks recorded over the past years in informal settlements (SA News 2016; Govender, Barnes & Peiper, 2011). Similarly, the high densities also pose a health risk as they are a fertile breeding ground of pests such as cockroaches and rats in South African cities (ENCA, 2014). The Ekurhuleni Municipality acknowledges that it has a rat problem as a result of informal settlements such as Ulana.

In addition, the high densities combined with the poor material used for building in informal settlements expose residents to weather elements such as rain, cold weather and wind. The poor storm water drainage and waste disposal systems in Ulana settlement also result in high levels of vermin and mosquitoes, causing disease and infection to spread rapidly. Poor access to good sanitation and clean water makes it difficult for people to maintain a good hygiene standard. The lack of good lighting on roads and pathways exposes individuals to crime and violence. There is also a lack of public and private spaces for recreation in Ulana, which often results in substance abuse, and a lack of sufficient food leading to malnutrition (HDA, 2014). Informal settlement upgrading then becomes imperative, in order to improve the lives of those who live there.
6.2 Occupancy Audit, Ulana Settlement

To allocate subsidised housing in a project to beneficiaries, an Occupancy Audit and Beneficiary Administration needs to be completed. The audit revealed that Ulana Settlement has 3,092 households residing within the community, with a total population of 7,031. The total population includes the head of household, spouses as well as dependents. Crucial to the project is knowing how many people qualify for subsidised housing. Table 1 illustrates that according to the findings, 1,758 households potentially qualify for subsidised housing. Which means that only 56.89% would potentially be allocated houses.

<table>
<thead>
<tr>
<th>Category</th>
<th>Quantity</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Potential Beneficiaries</td>
<td>1,758</td>
<td>56.89%</td>
</tr>
<tr>
<td>Single number of Dependence</td>
<td>834</td>
<td>26.97%</td>
</tr>
<tr>
<td>Income exceeds R3500</td>
<td>396</td>
<td>12.8%</td>
</tr>
<tr>
<td>Non SA Citizens</td>
<td>40</td>
<td>1.29%</td>
</tr>
<tr>
<td>Previously approved</td>
<td>64</td>
<td>2.0%</td>
</tr>
<tr>
<td>Total Number of Households</td>
<td>3,092</td>
<td>100%</td>
</tr>
</tbody>
</table>

Table 1: Occupancy Audit

To determine potential qualifiers for a housing subsidy, the identity document number of respondents was filtered through the Housing Subsidy System (HSS), which is a database and information system that is linked to the National Housing Database and provides information on individual beneficiaries who have applied for housing. The National Department of Human Settlements is the overseer of the HSS and the Provincial Department of Human Settlements are responsible for managing the database. Recently some local municipalities were also given responsibilities for the administrating and managing beneficiary applications. Housing officials are able to log on to the system and check the names of applicants on its database with those of the national database to see if applicants qualify for homes based on the qualification criteria of the housing programmes. The identity document, monthly household income, dependence and age are some of the determining factors for a housing subsidy. This information was all gathered to determine potential beneficiaries. Officials will verify these beneficiaries as the final process.

Applying smart technology to collect data for occupancy audit is reliable, up-to-date, flexible and it facilitates informed decision making as opposed to traditional cumbersome techniques such as use of paper based questionnaires. At present the data that currently exists for informal settlements for the Ekurhuleni Local Municipality is outdated and only contains the approximate number of households in each settlement without detailed demographic attributes as collected using the smart mobile GIS. This unfortunately becomes difficult for the Human Settlements Department to plan adequately for informal settlement projects and may lead to waste of financial resources. Having accurate and updated information that also contains potential beneficiaries will enable the department to plan more efficiently for its people as well as optimising resources. For example the current system of upgrading informal settlements often results in inadequate and poor planning as well as wasting of resources. For instance In-Situ upgrading project for Ulana informal settlement, named Balmoral Extension Four (Figure 5) resulted in an adoption of a layout plan with 1,092 opportunities in the form of single stands, zoned Residential 1, each Erf size being 78m².

It is important to mention that the Township Layout for Balmoral Extension Four was (Figure 5) completed prior to the audit and therefore, although well located, the number of stands available will not be sufficient to accommodate the needs of the community. Hence the use of smart mobile GIS tools to collect attributes of informal settlements is an invaluable tool in evidence based upgrading of informal settlements.

![Figure 5: Balmoral Ext 4 layout](image)

7. THE PROFITS OF INTEGRATED HUMAN SETTLEMENTS

The United Nations Housing Policy Guidelines For Developing Countries (1976) in its definition for housing, stated by an Ad Hoc Expert Group in 1962, that “housing is not ‘shelter’ or ‘household facilities’ alone, but comprises a number of facilities, services and utilities which link the individual and his family to the community, and the community to the region in which it grows and progresses”. Another Ad Hoc Group in 1970 concluded that “In fulfilling of social needs, housing plays both a direct and indirect role, and both roles are decisive. In its direct role housing serves as the area where the individual becomes capable of experiencing community and privacy, social well-being, and shelter and protection against hostile physical forces and disturbances. In its indirect role housing serves as the area where an abundant supply of social relationship and services are accessible, such as places for social intercourse, education, sports, social welfare and health protecting services, shopping and transportation”. These definitions illustrate and suggest that housing is very complex and therefore solutions on how best to deliver housing needs to be continually sought.

The conventional way of designing for low income neighbourhoods is unfortunately not ideal, if the needs of the community where to be comprehensively addressed. The reality according to the findings is that informal

---

7 National Housing Code: Vol 3 Individual subsidies
settlements are represented by a diverse group of people and should be planned as such. Integrated development translates this diversity, by providing opportunities for fully subsidised, partially subsidised, serviced stands, rental and bonded housing; creating ‘new post 1994 neighbourhoods’.

The benefits that follow integrated development create openings for ‘entangled spaces’, spoken of and defined by Nuttall (2009) as ‘A condition of being twisted together or entwined, it speaks of an intimacy gained, even if it was resisted, or ignored or uninvited. It is a term which may gesture towards a relationship or a set of social relationships that is complicated, ensnaring, in a tangle, but which also implies a human foldeness’. Integration through strong planning interventions would then eliminate social injustice within the urban space (Oelofse, 2003), encouraging policy makers to view urbanisation as a tool for development and not as a problem (UN Habitat, 2015).

Even though the integrated approach is not a new concept⁸, what this paper introduces is a tool that would improve implementation of this approach, within the informal settlement upgrading program. The solution is therefore the collection of attribute data of informal settlements which helps in the development of reliable, evidence based planning. The use of mobile GIS therefore needs to be extended to other informal settlement upgrading projects in South Africa as well as other cities in the global south. The data collected using smart mobile GIS tools is not only useful for enumeration and beneficiary administration but goes on to assist local government in implementing social, economic and spatial integration in line with policy⁹. Calling for the need to place people in well located areas. This will allow planners to plan efficiently. The use of smart technology also demonstrates that successful planning for housing delivery is not rhetoric but is based on sound information from communities concerned. However training is required to ensure the successful usage of smart mobile GIS tools.

8. CONCLUSION

Although the challenges faced by cities in the global south cannot all be solved in one day, this paper sought to provide a solution to the challenge of urbanisation in Ekurhuleni South Africa, manifesting itself through the form of informal settlements within its jurisdiction. A case study was used to demonstrate how the current model of settlement planning proved to be insufficient to comprehensively accommodate the needs of a diverse community. The paper demonstrated how integrated development can be implemented through the use of web based mobile GIS, by determining the number of potential qualifiers for each informal settlement upgrading project in order to assess the needs of communities.

9. REFERENCES


---

⁸ Cosmo City: An integrated development located in Johannesburg, South Africa can be viewed as an example, see: http://www.joburg.org.za


THE POTENTIAL OF STRAVA DATA TO CONTRIBUTE IN NON-MOTORISED TRANSPORT (NMT) PLANNING IN JOHANNESBURG

K Selala a and W. Musakwa b

a University of Johannesburg, Faculty of Engineering and Built Environment, Doornfontein, Johannesburg, South Africa-kadibetsyos7@gmail.com
b Department of Town and Regional Planning, University of Johannesburg, Doornfontein, Johannesburg, South Africa-wmusakwa@uj.ac.za

KEY WORDS: Non-Motorised Transport, cycling, crowd-sourced data, Strava, Origin, Destinations

ABSTRACT:
Transportation is one of the most fundamental challenges of urban development in the modern world. Cities are currently moving towards sustainable transportation, which includes non-motorised transportation (NMT). The city of Johannesburg has a goal of becoming a smart city, with sustainable development and smart mobility. Encouraging the use of non-motorised transportation and public transport has the potential of reducing the use of private motorised transport and therefore its negative consequences within the city of Johannesburg. There is limited research on cycling and how cyclists interact with cycling infrastructure within the city. The lack of such knowledge hinders proper planning for NMT. This research studies the ability of geolocation-based services to provide such information which can be useful in planning for NMT. Strava Metro application is the example of such geolocation-based services. The potential of this app to influence planning is investigated. According to the information acquired by Strava Metro, there are more cyclists in suburbs at the centre of Johannesburg and towards the north and towards the east, and in Kliptown in the south western part of the city. There are indications that the numbers of cycling activities recorded by Strava Metro are affected by the availability of gated communities, income levels, crime levels and the provision of infrastructure.

10. INTRODUCTION AND LITERATURE REVIEW

Transportation is one of the most fundamental challenges in urban development in the modern world (Jou, 2011). Accommodating motorcars has been an important theme of modern planning in many parts of the world (Todes, 2009; Behrens, 2004). Traffic congestion presents significant environmental, social and economic costs (Navarro et al., 2013; Neutens et al., 2012). The consequences include high resource consumption, externalities such as emitting greenhouse gases (Kenworthy, 2003).

High levels of dependence on the motor car, and the low densities developments make access difficult for those without this form of transport (Newman & Kenworthy, 1996). The emphasis on planning for mobility in cities abandon the significance of non-motorised forms of transport in developing countries (Behrens, 2005). Little attention is paid to the needs of NMT users for road space, crossings, and other amenities, resulting in high levels of accidents (Todes, 2009).

To solve the traffic problem, several solutions have been suggested. One was to reduce the trips people have to take, meaning that people have to travel less (Waterson et al., 2003). This could be possible through mixed land use neighbourhood (Marquet & Miralles-Guash, 2015). The affordability, safety and convenience of public transport (Filippi et al., 2013), and to encourage non-motorised transport (Visser et al., 2003). Non-motorised transport is any form of transport that does not rely on battery and/or fuel combustion driven mechanisms (Yazid et al., 2011). This is a fully human powered mode of transport. Examples include walking, cycling, per-ambulating, using donkey, horse or human-drawn carts/trolleys and rickshaws (Mammon et al, 2008).

Cities should strategically improve the transit and non-motorised mode alternatives to avoid traffic congestion (Frank et al., 2010). Within the area of public policy is where intervention can occur, through the increase in densities and concentration, through mixed use development, housing location, the design of buildings, space and route layouts, PT oriented development and transport development areas, car-free development, and through establishing size thresholds for the availability of services and facilities (Banister & Hickman, 2006). In recent years, promotion of NMT has moved up multiple policy agendas, including in relation to health, transport and climate change (Blanco et al., 2009). Favourable urban form is a critical factor in creating sustainable and energy efficient and greenhouse minimising urban transport systems (Blanco et al., 2009).

Despite the interest and the availability of additional funding for NMT, there are limits to how and where it can be applied (Kenworthy, 2003). Decisions about where and how to invest in NMT infrastructure are hindered by lack of empirical studies on cycling activities (Ali Hassein, 2015). Prioritising investments into improving bicycle facilities is critical for metropolitan areas (Handy & McCann, 2010). It is quite clear that cleaner air and safer streets come at a price, and only with well thought out long-term policies (Mohan, 2002). It is critical for metropolitan planning organizations and cities to effectively and accurately assess demand, connectivity, safety, and community goals (Handy & McCann, 2010).

Most public transport trips involve an element of the use of NMT at the beginning or end of each trip (Mohan, 2002). It is possible to redesign the existing roads to provide a safer and more convenient environment for NMT modes (Mohan, 2002). Most cities promote NMT by demarcating bicycle lanes, cycle tracks sideways roads, or bicycle paths that are not in conjunction with current roads (Broach et al., 2012). City planning experience in Beijing in China to Portland and in the USA suggests designing streets so that walking, cycling and the use of rickshaws become safer and more pleasant (Lamondia & Moore, 2014).
1.1. NMT in South Africa

Political and technical decision-makers in South Africa are often not interested in public transport and NMT because of their attachment to cars (Todes, 2009). This is because cycling in South Africa is a relatively dangerous activity as cyclists are not very visible and are generally not considered equals by motorists in terms of access to road space (Todes, 2009). Cycling in South Africa is seen as a means of recreation or a mode of transport for the poor (Gwala, 2007). On many roads, there are no places for pedestrians and cyclists (Mohan, 2002). Current analysis on transportation patterns show that the focus mainly remains on prioritising the private vehicle with large investment in ‘roadways for movement’, as opposed to promoting PT and NMT (Dewar & Todeschini, 2004). In the city of Cape Town, a public bicycle system has been introduced (Struwig & Anderson, 2013). In theory, this was supposed to encourage cycling within the city since people did not have to buy their own bicycle (Jennings, 2014). Provision of infrastructure for NMT is an interesting and credible solution that needs to be seriously considered in South Africa (Mbara & Celliers, 2013).

According to the Gauteng Household Travel survey (2003) cycling, accounts for 0.2% of trips in Johannesburg. Average travel time spent cycling to work is 42 minutes and to education is 16 minutes. The majority (approximately 90%) of the walkers and cyclists are striders; however, about 10% of people are regarded as stranded, some walk or cycle for longer than 30 minutes because they cannot afford motorised transport (Navarro, 2013). Most cycling destinations are/were within a radius of 10km. As many trips in Johannesburg are longer, NMT routes also need to help people reach the PT networks (City of Johannesburg, 2013).

The City of Johannesburg aims to improve its NMT. Hence, one of the objectives of the Johannesburg Road Agency is to ‘build, maintain, and manage PT and NMT infrastructure to support walking, cycling and the use of PT’ (Johannesburg Roads Agency, 2000) and the adoption of the street alive programme as part of its safety and liveable city delivery program (Visser et al., 2003). The City’s roads have traditionally been designed to cater more for motorists than they do for other road users (Visser et al., 2003). This program seeks to reverse this by prioritising the reallocation of road space to accommodate all users and by promoting road safety (Gwala, 2007).

There have been few surveys that have looked beyond motorised, peak-period and commuter travel, and these provide inadequate insight into the importance NMT in South African cities (Behrens, 2005). As a result, there is little information on the patterns, needs and trends of NMT users, particularly cyclists. This could be because data collection using conventional methods such as surveys is often costly, cumbersome, time consuming, out-dated and unreliable.

Perhaps using geolocation based services is a solution as it has been demonstrated elsewhere. As an example, the city of Salem used Strava to analyse the cycling infrastructure and cycling activities within the city (Strava, 2014). The results acquired were reliable (MacMichael, 2014), this motivate the study to follow such a technique to analyse the non-motorised activities within the city of Johannesburg. There are many other geolocation based services that people use to track their cycling activities and those that are known are: the ‘Runtastic Road Bike GPS App’, ‘Map my Ride GPS Cycling Riding’, ‘The Sport Tracker’, ‘Eco Bicycle’, most of this apps are available as mobile apps.

Strava Metro is a data service providing “ground truth” on where people ride and run. Millions of GPS tracked activities are uploaded to Strava Metro every week from around the globe (Strava, 2014). In denser metro areas, nearly half of these are commutes (Strava, 2014). These activities create billions of data points that, when aggregated, enable deep analysis and understanding of real world cycling and pedestrian route preferences (Strava, 2014). Strava Metro’s mission is to produce state of the art spatial data products and services to make cycling, running and walking in cities better (MacMichael, 2014).

Using Strava Metro enables informed and effective decisions when planning, maintaining, and upgrading cycling and pedestrian corridors (MacMichael, 2014). The previous years have experienced a remarkable success in the use of mobile devices (Muhammed et al., 2011). This has placed governments in most countries under pressure to improve telecommunication technologies (Freire & Paincho, 2014).

There is limited research on cycling and how cyclists interact with cycling infrastructure within the city of Johannesburg (Behrens, 2005). The lack of information hinders proper planning for how and where to invest in NMT infrastructure in the city. For example observations in the recently constructed cycling lanes in the Central Business District (CBD) are hardly used for cycling activity. Accordingly, this study proposes the use of geolocation based serves “Strava” as a way of acquiring information that can help understand the needs and the trends of NMT user and ultimately plan for NMT infrastructure. Consequently, this study aims to determine the usefulness of geo-location-based services in providing cycling trends and patterns in Johannesburg. Lastly how information obtained from Strava can be used to plan for NMT in Johannesburg is also investigated.

1.2. Study area

The City of Johannesburg metropolitan municipality area in South Africa was chosen as a study area. It was chosen because the city has an NMT policy and is developing NMT infrastructure. Yet little is known about the cycling patterns and trend in the city. Johannesburg is located within the Gauteng and together with City of Tshwane and Ekurhuleni is referred to the Gauteng City Region (Figure 1).

Figure 1: The City of Johannesburg map
Johannesburg consists of approximately 4.4 million people, accounting for about 36% of the Gauteng population and 8% of the national population (Todes, 2012). Johannesburg is the economic and financial hub of South Africa. Its economy and contribution to the national economy, has grown substantially and performed well on all major indicators when compared to other Metros. It contributes about 17% of the national GDP and approximately 47% of Gauteng’s economy. Johannesburg ranked 47 out of 50 top cities in the world as a worldwide centre of commerce (the only city in Africa) (Angel et al., 2005). However, massive challenges in terms of urban poverty, inequality, unemployment, food insecurity, social exclusion and underdevelopment still remain.

The city consists of seven regions which are Regions A (Diepsloot, Kya Sand); B (Randburg, Rosebank, Melville, Northcliff, Parktown); C (Roodepoort, Constantia Kloof, Northgate); D (Soweto, Dobsonville); E (Alexandra, Sandton); F (Inner City); and G (Orange Farm, Lenasia). The municipal city's land area is approximately 1,645 km$^2$ with a moderate density of approximately 2,364 km$^{-2}$ (municipal demarcation board, 2013). The population density is unevenly distributed throughout the municipal area. At a regional level, Region D is the most densely populated with 24.4% followed by Regions G (16.7%), F (13.4%), A (12.6%), E (11.8%), C (11.6%) and B (9.4%) respectively.

Johannesburg is a sprawling city with a twofold transport system. It consists of a car-based system in the most developed areas, where automobile use is almost compulsory. In most neighbourhoods, particularly poorer and deprived neighbourhoods transport is either by way of commuter trains, minibus taxis and buses. Commuters and shoppers walk long distances to access facilities (Beavon, 2002). Johannesburg is a highly structured, fabricated and interlocked poly and multi-centric urban mosaic. On the one hand, it stands for “the luxury city and city of control” as reflected by Sandton, for example.

The walled and gated communities in the northern suburbs, for example, Sandton and Midrand, reflect the “gentrified city and the city of advanced services”. The middle class South African areas such as Triumph (Sophia town) lend themselves to the “suburban city and city of direct production”. The largely previously segregated townships for Africans and Coloureds which are leftovers of apartheid highlight the “tenement city and the city of unskilled work”. Finally the “abandoned city and the city of the informal city” are depicted by the informal settlements which are generally located on the edges of townships or in any vacant spaces such as reflected by the inner areas, for example, Hillbrow. While the foregoing description does not do justice to a complex urban environment, it certainly provides a basis to frame the analysis (Beavon, 2002).

11. METHODOLOGY

Research design

The research is exploratory, and therefore utilises the Exploratory Spatial Data Analysis (ESDA). ESDA is a collection of techniques to describe and visualise spatial distributions, identify atypical locations or spatial outliers, discover patterns of spatial association, clusters or hot spots, and suggest spatial regimes or other forms of spatial heterogeneity (Anselin, 2012). Accordingly data form Strava Metro for the City of Johannesburg was collected and ESDA applied to reveal the cycling patterns, trends and distribution in Johannesburg for the year 2014. Strava data was chosen because it has been proven to be a useful source of data for cities such as Portland Oregon and Brisbane (Strava, 2015). Moreover to the best knowledge of the authors there are no complete datasets in Johannesburg that describe the spatial pattern and cycling patterns in the City.

Data description and preparation

Data from Strava Metro was obtained from Strava from Johannesburg. Strava Metro utilises data from Strava, which is a GPS enabled smartphone application that tracks bicycle rides and uploads the data to an online community of other users. Strava let’s athletes all over the world experience social fitness sharing, comparing and competing with each other's personal fitness data via mobile and online apps (Smith, 2014). The GPS in mobile devices store points, which are then stored as Big Data on Strava Metro. Strava Metro therefore packages this data in GIS format to enable cities to better understand cycling patterns. Consequently this data was acquired and analysed in a GIS to reveal the cycling patterns, trends and distribution in Johannesburg for 2014.

Strava Metro has three licences, namely, (1) Streets, (2) Nodes and (3) Origins and destination licences. At present we were unable to acquire the streets or nodes licence which give better insights into cycling patterns. Nevertheless currently the city of Johannesburg does not possess any information on cycling patterns, hence the Strava Metro would be a good start into providing such information. Accordingly, we acquired the Origination and Destination licence that records the start and end of cycling polygon activities. Data received from Strava was in dbf and shapefile format. The dbf contained all the cycling attributes whilst the shapefile contained the location (suburbs) of where the cycling activities took place in Johannesburg. Accordingly the two where joined using the join function in ArcGIS so as to spatially analyse the data. The data was then projected to Transverse Mercator projection.

The data was analysed and visualised in ArcGIS 10.3 software. Cycling patterns where analysed on the basis of the type (whether recreational or commuting), time, frequency as well as the origin and destination. The analysis was at city and neighbourhood level. At neighbourhood level the Spatial statistics, spatial analyst and map algebra functions of ArcGIS were utilised to calculate the cycling trips, the originating and destination polygons as well as the intersected polygons.

12. RESULTS AND DISCUSSION

The number of cycling trips recorded by Strava Metro for Johannesburg was 84297 for year 2014. Only 20% of the cycling trips are for commuting whereas recreational trips accounts for approximately 80% of the cycling trips in the city of Johannesburg. The number of trips recorded in a month is presented the graph in figure 2 below.
According to the graph on figure 2 above, November has the most cycling trips (with 10997 trips) the second highest is January with 9880 trips. The lowest is June with 4660, followed by July with 5151. There are higher numbers of cycling trips were recorded at the beginning of the year and towards the end of the year. During the middle of the year in, June and July, the number of cycling trips decreases significantly.

This could be because it is summer during the beginning of the year and also towards the end of the year and it is winter during the middle of the year. This confirms that cycling is affected by physical aspects such as weather, and climate, which provides reasons why it could be a seasonal activity. Johannesburg can get very cold during the winter season with temperature plummeting to -5 degrees and the months just before and after the winter season can also experience lower temperatures. The cold climate might make it difficult to cycle during the middle of the year.

According to the data provided by Strava, there are preferable hours for cycling in a day. Figure 3 below shows cycling trips per hour of the each day for the year in 2014.

There is a higher number of cycling trips in the morning, the number decreases as it approaches midday, and starts increasing after midday, during the afternoon and declines again in the evening after 16:00 until midnight. The highest number of cycling trips was recorded at 06:00 and at 04:00 and the lowest numbers of trips were recorded from 20:00 to 03:00. Most of the people who leave for work and school leave their origin (which is usually their place of residence) early in the morning. This explains the higher number of cycling trips in the morning between 04:00 and 07:00 as compared to the rest of the day.

Similarly, most recreational cyclists prefer cycling in the morning before they start with their daily activities and in the afternoon. Most of the recreational cyclists use cycling as a means for daily exercising, therefore it is preferably done in the morning and afternoon at their spare time. During the day, people are at work or at school and therefore the number decreases as most people will be busy with other activities.

The number of trip origin in a polygon and the number of trip destinations in a polygon are illustrated in figure below. Figure 4 (a) below shows cycling trip origins whereas Figure 4 (b) denotes the number of trip destinations.

Cycling is a short distance means of transport and therefore considering the sizes of the Strava data polygons, some cycling trips started and ended within the same polygon. The maps above also indicate that most polygons have higher number of trips origins than the number of trip destinations. It is also indicated that except Kibler Park at the south eastern part of Johannesburg, most of the suburbs with higher cycling O-D’s are from the centre of the city towards the north of the city and some on the east.

Most of the suburbs are along the N1 and the M3 routes. The suburbs are from Houghton Estate, Parkview, Waterval Estate at the centre and as one move north there are Hyde Park, Linden, Johannesburg, Morningside, and Kleve Hill Park and at the north of the city there is Carlsbad. Suburbs at the centre are Bergbron, Randparkrif, Windsor and Willowbrook. Most of these suburbs are concentrated by high income households. Some are situated closer to Sandton which recently has a higher economic value even as compare to the CBD.

According to an income distribution map provided in the Integrated Development Plan (IDP), it indicated that low-income earners mostly concentrate the south of the city of Johannesburg. These places with high population densities and low income rates are places such as Soweto and Orange farm, the CBD and Ivory Park. From these results one can assume that the level of income is directly proportional to the number of cycling activities. It might not directly influence the cycling activities but, it is related to property values and resource availability.
The collection of this data requires the use of a smartphone and for a person to have a certain level of technological or telecommunication knowledge. Low incomes are known to have a higher level of illiteracy. Most of the low income households might not afford the smartphone and therefore cannot contribute any means of crowdsourcing. The use of Strava Metro is not yet well known especially to such communities in Johannesburg which are mostly the southern and the south western part of the city. This is where most of the townships and the informal settlements are located. This is where the use of the non-motorised transportation is mostly not safe. The high levels of non-motorised risks are predicted at these areas. Cycling at the poor townships in Johannesburg raises a lot of safety issues and this might also explain the limited number of cycling activities in the low income townships in the south.

Although some of the people who use cycling as a means of commuting transport use this means of transport by choice there are some of the people who use this mode of transport because they are stranded and cannot afford any other means of transport besides walking and cycling. This would also explain the less percentages of commuting trips recorded by this mobile application. Most of the recreational cyclists use cycling as a means of exercising to keep their bodies fit, and it is preferable for the cyclist to exercise in the morning and afternoon at their spare time. This explains the high numbers in the morning hours and the afternoon. During the day, people are at work or at school and therefore the number decreases.

The number of cycling activities in a polygon could be affected by the number of gated communities that are in a specific polygon. The results of the data collected indicated that most of the cycling trips are spread towards the north of Johannesburg. These are the areas where there are the most gated communities in the city of Johannesburg. Gated communities have an impact on cycling in such a manner that, they are safer and cycling activities are provided for in most of them. Most of these gated communities are closer to people’s work places, shopping centres and schools, therefore, it is easier to use NMT in these areas. Most of the households in these gated communities and even in most parts of the region are high income households in Johannesburg.

13. IMPLICATION TO POLICY

The study has several implications for policy. The world is moving towards the use of intercommunication technology (ICT) and governments have to keep up with the development of technology. These implications include, providing an easier way for data collection and therefore, the decision-making process and providing an up to date insights of the cycling trips in the city of Johannesburg.

Traditional ways of data collection are popular in South African researches. However this study demonstrates the utility of Strava to collect data at a consistent spatial extent and for a continuous period thereby monitoring cycling activities. Crowdsourced data is also frequently self-updating, whereas it is not easy to frequently update data using the traditional way of data collection such as traffic counts which are often cumbersome, expensive and unreliable.

The frameworks (non-motorised framework and non-motorised transport policy) for non-motorised transport in Johannesburg does not use the collection of the actual number of people using non-motorised transportation as a strategy for the provision of infrastructure, but use criteria such as proximity to prioritise areas. It is assumed that the people use the NMT for only short distance. These policies then prioritise the areas closer and within the city centre because residential areas are in closer proximity to services. To a certain extent, using the criteria does work, but sometimes the infrastructure is provided without the knowledge of the number of people who needs such infrastructure. Some of the areas that need cycling infrastructure more are often the last on the priority list.

Most people who live in such areas do not use cycling. This neglects the fact that the poor mostly stays in areas far from services and therefore the stranded have to cycle for long distances because they cannot afford public transport. These areas should be the first on the priority list since according to Spatial Planning and Land Use Management Act (16 of 2013) the poor are to be prioritised when it comes to such infrastructure provisions. Such knowledge to the stakeholders will help in more informed frameworks and policies and the priority list will be restructured.

Information from Strava can therefore assist municipalities in planning for infrastructure as it shows the cycling trends, patterns and behaviour. For example the current cycling lanes built in the Johannesburg where built without evidence based research as hardly any cyclists utilise these lanes. Perhaps the City of Johannesburg can follow cities such as in Portland Oregon who have used data from Strava to plan for NMT infrastructure. From a community perspective Strava can also assist planners and the residents to know where most of the cyclists are and safe routes to take when cycling, particularly in South Africa with a high crime rate. Moreover data from Strava can perhaps make it possible Johannesburg to achieve smart city status.

There are still things that need to be done for the city of Johannesburg to achieve its goals. This involves more promotion of the use of NMT. This includes hosting more cycling activities, encouraging cyclists to contribute to crowdsourced data by using the Strava app to record any NMT trip. This will help provide adequate infrastructure, where needed and will make cycling safe. That is when more people might decide to use NMT after all. It will also make it easier for the decision makers, policy makers, and most of the leaders to lead by example and stop being too attached to the cars.

14. CONCLUSION

To conclude, it is clear that crowd-sourced data (Strava Metro) can be a useful tool when it comes to decision-making and the formulation of policies. It can be an essential tool to plan for non-motorised transportation in Johannesburg and throughout the country. There is still a lot of work to be done before this application can be used to its maximum potential and before people move to NMT. With proper education, the right investment, maximum enforcements of traffic laws, crowdsourced data a lot can be accomplished with regards to the efficient use of motorised transportation as well as provision for information that can be used in transportation planning in Johannesburg.
REFERENCES


TWEETS AND FACEBOOK POSTS, THE NOVELTY TECHNIQUES IN THE CREATION OF ORIGIN-DESTINATION MODELS

H. K. Malema*, W. Musakwa*

*University of Johannesburg, Faculty of Engineering and Built Environment, Doornfontein, Johannesburg, South Africa-hopekoketsomalema@gmail.com

Department of Town and Regional Planning, University of Johannesburg, Doornfontein, Johannesburg, South Africa-wmusakwa@uj.ac.za

KEY WORDS: Geolocation based services, Big data, Social media, Pattern analysis, Network movements, Origin-Destination models, Kriging, Transportation planning.

ABSTRACT:

Social media and big data have emerged to be a useful source of information that can be used for planning purposes, particularly transportation planning and trip-distribution studies. Cities in developing countries such as South Africa often struggle with out-dated, unreliable and cumbersome techniques such as traffic counts and household surveys to conduct origin and destination studies. The emergence of ubiquitous crowd sourced data, big data, social media and geolocation based services has shown huge potential in providing useful information for origin and destination studies. Perhaps such information can be utilised to determine the origin and destination of commuters using the Gautrain, a high-speed railway in Gauteng province South Africa. To date little is known about the origins and destinations of Gautrain commuters. Accordingly, this study assesses the viability of using geolocation-based services namely Facebook and Twitter in mapping out the network movements of Gautrain commuters. Explorative Spatial Data Analysis (ESDA), Echo-social and ArcGis software were used to extract social media data, i.e. tweets and Facebook posts as well as to visualize the concentration of Gautrain commuters. The results demonstrate that big data and geolocation based services have the significant potential to predict movement network patterns of commuters and this information can thus, be used to inform and improve transportation planning. Nevertheless use of crowd sourced data and big data has privacy concerns that still need to be addressed.

1. INTRODUCTION

Over the past 30 years Origin-Destination models (O-D) have evolved from static to real-time dynamic traffic models (Zhou & Mahmassani, 2007). These models have been crucial in establishing Intelligent Transportation Systems (ITS) for a city, since they provide predictions of traffic flows and network movements of commuters amongst other things (Hu & Liou, 2014). These predictions allows officials to identify road networks which are the most busiest, the times in which traffic is most congested as well as the modes of transport used by commuters. This kind of information is essential in transportation planning since it presents an opportunity for transportation policies to be improved and for ITS to be used in transport management (Hu & Liou, 2014).

An effective transport management strategy is one of the most important elements that contribute towards the sustainability of a city. Not only does it assist with the mobility of goods and people in a city, but an effective transport management system also impacts positively on the economic and environmental aspects of transportation (Gao et al., 2012). Most cities around the world, however do not have these effective management strategies in place, hence there is a rising percentage of road accidents, the use of private automobiles and traffic congestions experienced (Fernandes et al., 2012). This predicament thus calls for the efficient, reliable and integrated planning of transportation systems, especially in the developing nations.

Various scholars have thus, identified Intelligent Communication Technologies (ICT) as efficient tools which have the potential to assist with the effective management of transportation systems (European Commission, 2010). This is seen evident in cities such as Barcelona and Dubai, given that both cities are well renowned for being prominent smart cities established to date (Bonnel et al., 2015 & Dassani et al., 2015). These cities have transportation systems that incorporate ICT which is in the form of smartphones, apps, sensors and WiFi in the mobility of their citizens. South Africa has recently caught on to this trend, as it has established its first rapid train system known as the Gautrain, which is a first of its kind to be established in Africa.

1.1 The Gautrain Project

The Gautrain is a mega-engineering project which is the very first rapid-transit train to be launched in the African continent in June of 2010 (GMA, 2010). The train project aims to reduce traffic congestion and encourage the use of public transportation systems in Gauteng, whilst being a means to realise smart mobility through the incorporation of ICT. The Gautrain is built and operated using some of the most advanced technologies in world and this makes for the success of the train project (GMA, 2010). In addition, the Gautrain has a phone application which displays the time schedule and alerts of delayed trains, thus allowing commuters to manage their travelling trips accordingly. Moreover, the app has a feature which permits commuters to calculate the amount of train fares from one station to the other, based on the time of the day. Such an app provides commuters an insight on the trip that is about to be pursued (Liu & Teng, 2015).
The railway network of the Gautrain runs through the Gauteng province (Figure 3), linking three metropolitan cities: the City of Johannesburg (COJ), the City of Tshwane (COT) and Ekurhuleni Metropolitan Municipality (EMM). To date there is little known about the travelling behaviours and patterns of the Gautrain commuters, hence the study assesses the feasibility of using modern technologies as tools to create origin-destination models for Gautrain commuters. This approach is being tried and tested as there is evidence of insufficient spatiotemporal data obtained for this project.

In recent years, positioning technologies such as sensors and social media networks have been incorporated into transportation planning and particularly in the creation of O-D models. This novel approach can be credited for its ability to create sustainable and efficient transportation network systems. Detecting the geographical location of consumers from advanced technologies such as the Web 2.0 and big data together with social media network sites, has provided a platform for instant communication on real-time traffic updates to take place (Chen & Liu, 2014). The use of these technologies has made the exchange of data far much easier and as a result, this subject matter has become of great interest in the research domain particularly in transportation planning (Hongyan & Fasheng, 2013; Peters et al., 2013).

Likewise, this study aims to assess the feasibility of using geolocation-based services such as Twitter and Facebook as data mining tools to map the movement network patterns of the Gautrain commuters. The remainder of the paper is structured as follows; the next section presents the literature study on Origin-Destination models, which is briefly followed by a summary of the study area. A segment discussing the methods and materials used to conduct the study follows this. Lastly the proceeding section highlights the key results and implications of big data in planning.

1.2 Origin-Destination studies

O-D studies and models are a fundamental element in transportation management and have since been identified as being key tools in informing transit planning (Bohte & Maat, 2009; Kling & Pozdnoukhov, 2012; Gao et al., 2012). Previously, O-D studies used conventional methods such as household surveys and traffic counts as data mining tools (GAO, et al., 2012; Jin et al., 2014). With time, these methods proved to be rather spatially limited, outdated, cumbersome, unreliable, expensive and tedious (Gao et al., 2012; Lu et al., 2013). Hence, this novel approach of using positioning technologies to capture and trace commuter’s demands and network movements in the attempt to establish real-time, reliable and spatially expansive transportation datasets used to optimize the use of street networks (Lu et al., 2013).

Positioning technologies such as smartphones, social media networks and Geographic Positioning Systems (GPS) which make up big data and Web 2.0, serve as data collection tools in the creation of O-D studies (Farah, 2014). This application of these technological devices and sensors contributes towards the establishment of smart cities since the technology allows for the easy exchange of traffic information to take place instantly. In addition, these technological devices assist with the monitoring of traffic congestion and road accidents, amongst other things (Caragliu et al., 2009).

Figure 1 depicts the connectivity and transfer of information that is typical in a smart city. Smart cities are described as cities which invest in both human and social capital, traditional as well as modern Information Communication Technologies (ICT) which fuel sustainable economic growth and high quality of life, through having prudent management of natural resources encouraged by participatory governance (Caragliu et al., 2009).

![Figure 1: The Gautrain and its routes](image1)

![Figure 2: Illustration of smart city Source: turktelekomint, 2015.](image2)

The smart city concept is centred on the idea of mankind incorporating technologies in the planning and operations of cities with the intent to improve the quality of lives of citizens whilst creating sustainable cities. To date, there are numerous smart cities established around the world with the majority of these being in the First World countries. Some of these cities include Barcelona, Amsterdam, London and Dubai to name a few (Bakici et al., 2013).

Dubai is one of the most renowned intelligent cities established in the Third World countries. This city is highly commended for effectively employing ICT in the governance of the city, which subsequently promotes transparency, easy access to information, sustainable transportation systems and so forth (Dassani et al., 2015). Cellphones, have identified as a very important gadget in upholding the standards of Dubai, since this ubiquitous

25
device provides instant access to datasets and information that is crucial for citizens.

1.3 Geolocation-based services

The past 30 years of technological developments has led to an era of social media revolution which has seen almost half of Chinese students using mobile social media sites such as Facebook and Twitter to acquire information on various issues (Xu et al., 2015). The establishments of these communication systems allows for large amounts of data to be captured, stored and exchanged on smartphones, tablets and other mobile technologies, instantly (Peters et al., 2013). This era has brought with it innovative approaches of retrieving data for various uses such as the improvement of service delivery, customer services as well as transportation planning. As a result, social media networks like Twitter, apps such as Waze and sensors like GPS have thus, become popular for their ability to provide real-time traffic updates (Hongyan & Fasheng, 2013).

The acquisition of data through the above mentioned technologies is made effortless and convenient, due to the ubiquitous nature of smartphones. Not only do these technological devices offer access to the Internet, they also provide geographical location of consumers through the use of GPS sensors (Kaya et al., 2014). This feature makes it easy to identify and navigate to the location of a consumer or a loved one. Moreover, information on the activities undertaken by consumers at a specific location is also provided through social media applications (Hasan & Ukkusuri, 2014). These types of mobile applications are commonly referred to as geolocation-based services. The most popular amongst these are Facebook, Twitter and Foursquare as they contain a check-in feature that provides the geographical location of consumers as they navigate in and around cities (de Abreu Freire & Painho, 2014).

These services have become very popular in research because of their ability to provide rich data that has the potential to improve basic service provision such as road infrastructure for a particular area (Hasan & Ukkusuri, 2014; Peters et al., 2013). Despite their use in various fields, the geolocation based services are at the forefront of providing data in the transportation domain. Their ability to provide information on time, date, routes and activities rendered at a specific time, is a distinct feature that makes these services essential for providing information for Origin-Destination models (Filippi et al., 2013). In addition, the information shared on geolocation based services provides commuters the opportunity to plan their trips accordingly should there be an alert made of an accident, road constructions or traffic congestion in a particular road network.

A study revealed that nearly one in five smartphone consumers in the United States of America uses geolocation based services whilst commuting (Comscore, 2011). This statistic demonstrates the potential use of smartphones as data collection tools since they are utilised everywhere and at any given time. The data collected through this device if exchanged and shared through crowdsourcing, can be used for various reasons, including the creation of Origin-Destination models (Filippi et al., 2013). Accordingly, the O-D models will be used to provide an insight as well as predictions on individual travel behaviours (Wechsl, 2014). This predicted travel information is of great significance as it allows transportation planning policies to be drafted based on up-to-date data (Kaya et al., 2014).

1.4 Privacy concerns

The use of crowdsourced data in various fields is becoming increasingly popular. This is due to the ubiquitous traits of cell phones, which allows for the capturing, recording and sharing of spatio-temporal data (Farah, 2014). Crowd sourced data is commended for empowering citizens by allowing them to provide an input on issues and matters concerning development in their communities, amongst other things (Blatt, 2015). Conversely, crowd sourced data has been widely criticized for the reliability or lack of the data obtained. Scholars frequently raise this concern and they argue that data and opinions provided by ordinary citizens with no level of expertise on certain issues cannot possibly provide substantial information that could inform policies (Blatt, 2015; Callister, 2000).

Smartphones and crowdsourced data provide large amounts of significant information particularly relating to traffic (Filippi et al., 2013). However, cell phone use raises privacy concerns when consumers share data on the Internet. This is identified as one of the weaknesses of crowdsourced data since information such as location, identification number, contact details, etc., can be made available for anyone to access especially through social media sites (Blatt, 2014). Once this kind of information is obtained by people malicious intent, it can consequently be abused.

1.5 The use of Smartphones in South Africa

Smartphone’s are renowned for being efficient data collection tools (Xu et al., 2015), however in South Africa these technologies are yet to be fully exploited for uses other than mere communication devices. It is in the past 10 years, however, that these technological devices, together with crowdsourcing software have emerged and are being explored. Some of these software include echo-social and Mapbox which have the potential to inform transportation planning based on the feeds from commuters. The former has the ability to capture data on the content and location of commuters who use social media sites whilst commuting (Loebal, 2012). The latter, however, is used to capture the location and number of social media users based on the different smartphones used at a specific time period. Both these software have the potential to provide essential information to planning authorities on the movement patterns and behaviours of South African commuters, if used correctly. With South Africa’s poor public transportation management and planning policies, these novelties could be used to improve on the transportation system of the country, whilst providing an attempt at creating smart mobility.

2. STUDY AREA

South Africa is considered to be one of the most developed countries found on the African continent. This
country has been ranked amongst the top ten countries to have the highest GDP in the African continent (IMF World Economic Outlook, 2015). With that said, South Africa like many other African countries still lacks essential infrastructure, relevant skills and resources that would aid the development of the country (Douangphachanh & Oneyama, 2014). Consequently, the country experiences a multitude of issues related to transportation planning, urbanisation, poverty and so forth.

Planning in South Africa was previously conducted using the apartheid statutory approaches hence, the settlement planning in the country resembles segregation which current planners are attempting to rectify through modern planning theories (Todes, 2011). Similarly, the transportation planning policies are outdated, resulting in poor transportation management systems that give effect to poor public transportation systems, the preferred use of private automobiles, traffic congestion and heavy emissions of carbon monoxide (RSA, 2008). These policies were often informed by household surveys and traffic counts (RSA, 2008), which were previously considered the main methods to acquiring information on the travelling behaviors of commuters.

The Gauteng province is the ‘Economic hub’ of South Africa and contributes 3.3% of GDP towards the country’s economy (RSA, 2012). This province further boasts with diversity, which ranges from a variety of ethnic groups from across Africa, to the various economic activities that take place in the major cities (GMA, 2010). The Gauteng province is reported to have 58% of its populace being economically viable. It is comprised of three metropolitan cities: the City of Johannesburg (COJ), City of Tshwane (COT) and Ekurhuleni Metropolitan Municipality (EMM) as illustrated in Figure 2 above. These metropolitan cities combined create an economic region, which is mostly comprised of the Finance, Real Estate and Business sector as well as the manufacturing sector. These metropolitan cities, in addition also create a corridor, which is clustered with numerous industrial developments along the N1 highway. Furthermore, the cities share the Gautrain railway, which is Africa’s, first high-speed railway.

3. METHODS AND MATERIALS

3.1 Methodology

Spatial and quantitative techniques were used to determine the trip distribution of Gautrain commuters. Tweets and Facebook posts containing information on the location of Gautrain commuters were used to derive the concentration levels of the commuters in various neighbourhoods of Johannesburg, Ekurhuleni and Tshwane. The Exploratory Spatial Data Analysis (ESDA) was employed as the research design since it assists in revealing patterns, trends and distributions spatially.

3.1.1 Data collection and preparation

Cadastral data of the municipal boundaries for the three metropolitan cities and social media feeds (tweets and Facebook posts) were collected from municipalities and the Echo-social software, respectively. This software allows for the active monitoring and engagement between customers and company owners to take place in the social media space. The data obtained from Echo-social contains information on the thoughts and opinions of Gautrain commuters.

The data used for this study was captured over a 6 month period starting from 1st January 2015 to 1st June 2015. The Echo-social software contained relevant data which was only captured for the above mentioned dates; hence the study used information recorded within the 6 month period. This data was captured in an excel spreadsheet, which contained crucial information on the coordinates and location of consumers who posted and tweeted about the Gautrain as seen in Figure 3 below.
The coordinates contained in Table 1, were used to visualise the geographical locations of the tweets and Facebook posts using the ArcGIS software. Identifying the location of the tweets and Facebook posts highlights the areas which have the most social media activities.

The excel spreadsheet data (Table 1) with the enumerator values initially contained certain columns and rows with invalid information and duplicates of entries needed to be deleted, prior to running the GIS analysis. There were 18634 entries left from the original 64043 entries after deleting the invalid columns and rows. Subsequently, the spreadsheet was loaded onto the ArcGIS software and this allowed for the location where the tweets and posts were made to be visualised and analysed spatially. One of the first analysis to be completed on the data using the ArcGIS software, was the kriging analysis, which requires one to create Z-values using point data and digital elevation model (DEM) data.

3.1.1 Kriging and fishnet analysis

Kriging is an interpolation analysis which uses point data to determine the spatial correlation of points in relation to each other (ESRI, 2015). This particular analysis is significant for this study since the study uses social media feeds (in the form of point data) to determine the concentration levels of tweets and Facebook posts made by Gautrain commuters. These concentration levels were used to determine the areas of origin and destination for the Gautrain commuters. Below (Figure 3) is the Kriging mathematical formula which is used to calculate the point data prior to executing the results from the analysis.

\[ \hat{Z}(s_0) = \sum_{i=1}^{N} \lambda_i Z(s_i) \]

where:
- \( Z(s_i) \) = the measured value at the \( i \)th location
- \( \lambda_i \) = an unknown weight for the measured value at the \( i \)th location
- \( s_0 \) = the prediction location
- \( N \) = the number of measured values

Source: ESRI 2015

The Kriging analysis was executed on both the Gauteng boundary and the three municipal boundaries, using a maximum distance of 1000m/1km. This radius distance is large enough to cover the social media feeds from commuters coming in and going out of the Gautrain stations, hence it was used.

There were more analysis' conducted on the social media point data after the Kriging analysis were completed on both the Gauteng province and the three municipal boundaries. These new analysis included the fishnet and counts in polygon which uses the Geospatial Modelling Environment (GME) software. The former analysis was used to create grid cells of a 5000m² area for both the Gauteng and municipal boundaries. The rationale behind creating these grid cells is to ensure that the concentration levels of the point data are calculated on equal grids so as to enhance the accuracy of the results. The latter analysis however, was used to create the cold and hotspots of social media activities through converting point data to concentration levels within the grid cells created using the fishnet analysis. The results obtained from both the fishnet and counts in polygon analysis are useful in identifying the neighbourhoods which have more social media activities as indicated in Figure 6.

4. RESULTS AND DISCUSSION

This section of the paper presents the results obtained from the Kriging, Fishnet and Count in polygon analysis. The results have been presented in the form of maps. Following this is a section that provides an interpretation of the results obtained. The results obtained have been used to infer a rationale and to provide logic for the analysis obtained pertaining to the concentration levels of tweets and Facebook posts.

Figure 4: Maps showing the concentration levels of the tweets made in the Gauteng province

Figure 5: Maps showing the concentration levels of the tweets made in the Gauteng province.
The output from the Kriging analysis is maps indicating commuter density levels. These density levels are measured using high and low concentration values, which represent hot and coldspots, respectively. The hotspots are represented by the red and orange colour and the coldspots by the colour green and yellow. The former are symbolic of the locations with high commuter density levels, whereas the latter are typical of locations with low commuter density levels. All the locations with high commuter density levels have been used to refer to both the origin and destination since the data used for this study did not stipulate the areas of origin and destination of commuters. Hence, it was difficult to identify which areas were the origin and destination of the commuters using social media networks.

Figure 4 demonstrates commuter density levels of social media activities within the Gauteng province. The concentration levels of commuters using social media in the province ranges from the value 1097 to 1782 with the former representing coldspots and the latter hotspots, respectively. The areas shaded in orange are moderately concentrated and have commuter density levels close to the maximum concentration value of 1782. Likewise, the locations shaded in yellow represents commuter density levels which have a minimum concentration value that is close to the commuter density level of 1097.

Running the Kriging analysis on such a large scale as the province produced useful results since the outcome indicated the locations within the province which had high and low commuter density levels. Nonetheless, it was also necessary to run the analysis on a smaller scale so as to identify the neighbourhoods with the most social media activities. Thus, there is Figure 5 & 6, which is the outcome of both the Kriging and Counts in polygon analysis run on the three metropolitan municipalities.

Figure 5 illustrates that the stations with most commuter density level and social media activities are Park, Rosebank and Sandton. The commuter density levels in these stations are high because of the locations which they have been established in. For instance, the Gautrain Park station is located near the largest train station in Africa. This station is commonly known as Park station and it is renowned for offering a variety of modes of transport to commuters from all over the county and Africa. These modes of transport range from buses to taxis and trains. It is for this reason that Park station has been recognised as a good exemplary of Integrated Transportation System (ITS) as there is an integration of the BRT station, Gautrain and Gaubuses, Taxi’s and the Metrorail trains all located within the same proximity

Both the Rosebank and Sandton stations are located in close proximity to the Rosebank mall and Mandela Square mall, respectively. These are the two most popular nodes in the Gauteng province since they are known to provide a variety of entertainment. Apart from the entertainment aspect, most people travel to and from Rosebank and Sandton on a daily basis because of work. As a result of their popularity, these nodes attract the greater population and since the Gautrain stations are located in close proximity to them, most people choose to travel by the Gautrain to get to their desired destinations.

The City of Johannesburg according to Figure 5 has the most concentrated levels of social media activities as a result of having 5 of the 10 Gautrain stations located within its jurisdiction. The Johannesburg CBD together with neighbourhoods such as Cresta, Parktown, Braamfontein, Ormonde and Randburg to name just a few are some of the commuter density hotspot areas in the municipality. The high concentrated levels of social media activities in these neighbourhoods can be attributed to the fact that most of the neighbourhoods are located close to the Gautrain stations mentioned above. For example, Braamfontein is in the same vicinity as the Gautrain Park station, thus there is such a high concentration level of social media activities.

There are numerous neighbourhoods south of the Johannesburg city which have high commuter density levels. These neighbourhoods include Soweto and Meadowlands which do not have Gautrain services in operation. Therefore, one can assume that the high concentration levels of social media activities about the Gautrain are a result of having the residents of these areas using the Gautrain services to travel up north to places such Centurion, Midrand and the likes. For such people, the Park Gautrain station is the first point of departure and this station can thus be assumed to be their place of origin once they reach the Johannesburg CBD.
Figure 6 is the results obtained from running the fishnet and counts in polygon analysis. As mentioned above, both the fishnet and counts in polygon allow for an analysis to be made on a smaller scale by creating the grid cells which enhance the level of accuracy of the results, as shown in Figure 6. The counts in polygon analysis basically calculate the number of points found within a polygon. In terms of this paper, the point data refers to the twitter and Facebook posts which have been calculated within every 5000m² grid cells. The results obtained thereof, range from 0-4399 and these represents the commuter density levels.

The results obtained in Figure 6 demonstrates that there are more social media activities taking place in neighbourhoods located within the vicinity of the Gautrain stations. With that said, Braamfontein and Brooklyn are some of the neighbourhoods which are illustrated to have the most highest commuter density levels. These neighbourhoods are located around the Park and Hatfield station, respectively. The number of tweets and posts made in and around these stations range between 660 and 4399. The location of the Hatfield station encourages the use of the Gautrain for people travelling down south to places like Centurion, Sandton and even to the OR Tambo international airport. The Hatfield station is located north of the province and close to the University of Pretoria as well as affluent neighbourhoods such as Hillcrest and Brooklyn and this makes travelling to the south using the Gautrain more time effective as one is able to avoid the traffic congestion found along the N1, especially during peak hours.

The Rosebank, Sandton, Midrand and Centurion stations have commuter density levels ranging from 150 – 659 according to Figure 6. These moderate hotspots indicate that these stations are either the origin or destination stations for the Gautrain commuters. Contrary to these results, are the commuter density levels found in Marlboro, Rhodesfield and OR Tambo International stations. These stations have low concentration levels of social media activities due to the following possible reasons: The Marlboro station is located on the outskirts of a township called Alexander, which is notorious for having high crime rates. The crime aspect could be the determining factor for whether a commuter chooses to use the Marlboro station or not, hence it is a coldspot. Both the Rhodesfield and OR Tambo stations are situated in and around the OR Tambo International Airport and as a result they attract social media activity from tourists coming in and out of the country. These stations serve as areas of origin and destination for those tourists and citizens using the OR Tambo Gautrain station. With that said, it can be assumed that there is a small percentage of population which use the Gautrain in these stations to access places in the east of the province.

The city of Tshwane has low concentration levels of social media activities, despite having one of the largest Gautrain station in the city, which is the Centurion station. This could be a result of having most of the commuters in this city having limited Gautrain services, hence so many coldspots. Areas such as Bronkhorstspruit, Ga-rankuwa and Temba are marginally concentrated thus, representing the coldspots. Such areas are furthest away from the Gautrain services; hence they are less users of the Gautrain located there. It can therefore, be assumed that people residing in those areas use other modes of public transport and do not commute with the Gautrain on a regular basis.

The northern parts of the Gauteng province have low commuter density levels. This may be alluded to the fact that people residing in the northern parts of the COT work in the Tshwane CBD, hence they do not utilise the Gautrain services as much as the people travelling from the south of the Gauteng province. This observation implies that areas such as Ga-rankuwa and Bronkhorstspruit are the places of origin for the commuters and the CBD which is located north of the Pretoria station is their final destination. Hence, most of the residents do not have much contact with the Gautrain.

Figure 7 indicates the number of tweets and Facebook posts made by commuters during the six-month study period. It is evident from the graph that most messages about the Gautrain were posted during the weekdays than they were during the weekends. This is most likely because most people use the Gautrain during the week for business purposes such as travelling between home and work as compared to using it for leisure purposes during the weekend. Such information is important as it has the potential to give an indication to company owners of the busiest day of the week. From this information, the operators will then know whether to increase staff or more trains on that particular day or not.

The contents of the messages posted by commuters or consumers are crucial for any business, irrespective of whether it is in the private or public sector. Therefore, their ability to provide real-time data makes social media sites the relevant tools to assist with improving transportation planning by analysing the crowdsourced data. Based on the analysis of the contents of messages, company owners and planning authorities can provide the necessary transportation services and infrastructure in areas where most complaints are coming from. The contents of the tweets and Facebook posts could also be used to inform transportation planning authorities in the public sectors of the need for providing services such as the Gautrain app and timetable for other modes of transport used in the Gauteng province. Such features are necessary as they synchronise the times of the public transportation systems thus allowing consumers to plan their travelling trips accordingly.

5. LIMITATIONS OF THE STUDY

The Echo-social software used to provide data for this paper has highlighted some weaknesses, which may be assumed to be generic with other big data software. The data obtained from the Echo-social software contained various rows and columns of data which were invalid. Consequently, the data had to be cleaned enhance accuracy in the analysis. Similarly, there where cases where the location (coordinates) could not be verified, because users deactivated the location features on their cell phone. Deactivating the location feature on ones smartphone could pose a challenge for companies or authorities who would like to obtain data on the locations of their commuters or citizens.

Another weakness identified was that the location points representing the tweets and Facebook posts did not
specify which areas are the origin and which are the destinations. Therefore, it became difficult to separate the two, once the Kriging analysis was conducted. This Kriging analysis highlighted the commuter density levels of social media activities for the Gautrain commuters. And as a result, the study assigned the concentrated areas to represent both locations of origin and destination.

Lastly, ward data for Gauteng was used to identify which wards have more concentrated levels of social media activities. Wards are not similar in size hence the results may be distorted. Consequently, a fishnet analysis was therefore conducted in order to assign equal grid cells to the municipality boundary so as to obtain more accurate results.

6. IMPLICATION FOR PLANNING AND CONCLUSION

The content of the social media data obtained from EchoSocial proved to be proficient tools to provide an insight on the travelling behaviours of commuters. This data was used in a density analysis called Kriging, which enabled one to identify areas which areas have a concentrated level of social media activities and commuter density. This information could thus, be used in transportation planning to alert and inform planning authorities about roads which are commonly used and congested. Furthermore, the times of day for which the congestion takes place can also be brought to attention. In light of this information, authorities may choose to expand lanes, establish new road network links or work on improving the already existing public transportation systems as an attempt to reduce traffic congestion.

The areas which were found to be highly or moderately concentrated with social media activities need to be assessed for possible Gautrain extension sites. The concentration levels could be an indication that there is a threshold of commuters who could benefit from having the Gaubus services, extended to those areas. According to the results obtained, these potential areas include Soweto, Randburg and Springs, to name a few.

Social media could also be used as a platform for civil society to voice out their level of satisfaction (or lack of) with regards to the services provided by the government and business entities. This is a cheap and convenient way of collecting data from commuters without having to leave the comfort of their homes. Such mediums could also be used to provide recommendations in terms of how to improve transportation planning in South Africa based on the views and opinions of commuters from the Gautrain, as it is renowned for being the first high speed train in the country. This kind of data may probe authorities to improve on transportation facilities and infrastructures for other modes of transportation. From a business aspect, big data could be used to identify customer profile and their respective places of origin.

In conclusion, social media data which contains information on the location of origin and destination of commuters, accurate times and all the other information included in Figure 3, has the ability to be analysed and thus be used to draw travelling pattern and behaviours of commuters.

REFERENCES


Using crowdsourced data (Twitter & Facebook) to delineate the Origin and Destination of commuters of the Gautrain public transit system in South Africa

T Moyo¹, W Musakwa¹²

¹ University of Johannesburg, Faculty of Engineering and Built Environment, Doornfontein, Johannesburg, South Africa - thembanijoel@gmail.com
² Department of Town and Regional Planning, University of Johannesburg, Doornfontein, Johannesburg, South Africa - wmusakwa@uj.ac.za

Key Words: Origin and destination; geo-location data; commuters; kriging; density

Abstract:

The study of commuters’ origins and destinations (O_D) promises to assist transportation planners with prediction models to inform decision making. Conventionally O_D surveys are undertaken through travel surveys and traffic counts, however data collection for these surveys has historically proven to be time consuming and having a strain on human resources, thus a need for an alternative data source arises. This study combines the use social media data and geographic information systems in the creation of a model for origin and destination surveys. The model tests the potential of using big data from Echo echo software which contains Twitter and Facebook data obtained from social media users in Gauteng. This data contains geo-location and it is used to determine origin and destination as well as concentration levels of Gautrain commuters. A kriging analysis was performed on the data to determine the O-D and concentration levels of Gautrain users. The results reveal the concentration of Gautrain commuters at various points of interest that is where they work, live or socialise. The results from the study highlight which nodes attract the most commuters and also possible locations for the expansion for Gautrain. Lastly, the study also highlights some weakness of crowdsourced data for informing transportation planning.

1. INTRODUCTION

The study of commuters’ origins and destinations (O_D) promises to assist transportation planners with prediction models which inform decision making. Conventionally O_D surveys are undertaken through travel surveys and traffic counts, however the data collection exercise for these surveys has historically proven to be time consuming and causing a strain on human resources, thus a need to aggrandize the data sources (Wolf, et al., 2003). As we are now living in the age of the internet of things (IoT) a need for smart analytic techniques has arisen. Bolstered by the current advancements in web 2.0, humanity has gradually departed from the culture of using the internet to send emails to incorporating it into every aspect of their lives (Gao & Liu, 2013). Chandler (2015, p183) articulated how data is now capable of altering the ways in which knowledge of the world is produced and consequently altering the ways in which it can be governed. This information age has subsequently created new prospects for transportation planning by revolutionising how information is managed, collected and analysed to improve transportation systems.

Accordingly with the dynamics of public transportation being multifaceted, the move towards intelligent transportation systems seems to be the logical solution for addressing situations which may unfold, instead of using the traditional reactive approaches. The general consensus amongst scholars is that it is now possible to model the spatial dependence of commuters using geographical location data to predict areas of clusters and outliers ((Wolf, et al., 2003; Stopher & Greaves, 2009; Gao & Liu, 2013; Hasan & Ukkusuri, 2014).

The evolution of analytics

The analysis of social media data can best be expressed through an insight of developments in data analysis which has evolved over the years as highlighted in figure 1. Historically data analysis between 1995-2009 scholars used data as a means to an end, and those between the years 2009-2013 analysing data as both a means to an end and also as the end. In the recent years there has been a paradigm shift with scholars from 2013-2016 analysing data as the end and those post 2014 analysing data as a service (Deloitte, 2014). This move from data as an end to empowering cities as a service opens up new possibilities, as data is no-longer only viewed as either a means to an end or just the end, but as an enabler for decision making and collecting feedback for development. Subsequently this offloads the risks and burdens of data management to a third-party cloud-based provider (Deloitte, 2014). The evolution of analytics has greatly changed the manner in which data is managed, as it has led to improvements in data quality, agility and reduction of cost. Furthermore these analytical tools seem to be a viable resource that will improve operational efficiency, while boosting the quality of urban planning.

Figure 1: Evolution of analytics
Big data
In past 5 years there has been a rapid incorporation of social media data in transportation studies. Lorenzi, et al., (2014) have articulate 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 in-built 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; Twitter and Google Maps). The data generated by these applications has the potential of being used to analyse the day to day movement networks of human beings. 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.

The growth of big data analytics has spawned remarkable captivation globally (Riggins & Wamba, 2015), this can be seen with many city officials engaging with the private sector, in a bid to make use of this big data, such as using Echoecho platform to analyse social media data. This has made it possible for city authorities and private companies to understand the multifaceted aspects of social media data. This will led to inside on how people interact with their immediate environment, through insight of social media big data which is amassed by posts made daily by the social media users. Big data has over the years been described as data sets whose size is beyond the ability of commonly used software tools to capture, manage and process the data within a tolerable elapsed time (Riggins & Wamba, 2015).

Consequently social behaviour identification through the quantification of various aspects of human behaviour is now possible through this big data. With regards to urban planning, the unpacking of big data has led to a reduction in time spent to respond to service delivery grievances, as the community can easily inform council of any grievances via mobile applications, thus bridging the gap between the ordinary citizen and local authorities (Hasan & Ukkusuri, 2014). As mobile devices have become more advanced within built sensors, it is now possible to trace and create a digital foot print showing the movement of people, through the collection of big data from mobile network towers, social media platforms and wifi feeds (Lancey, 2001; Yang, et al., 2012; Chatzimilioudis & Zeinalipour-Yazti, 2013). This has the potential to inform planning in that authorities can identify areas which have the potential for investment by analysing the sphere of influence of various land uses. Furthermore big data can be used to improve service delivery. For example Waze through its connected citizens program has assisted authorities such as in Rio de Janeiro to re-direct traffic to other routes to avoid traffic jams during the rush hour as Waze relies mainly on crowdsourced data from users (Waze W10, 2014).

Crowd sourced data
Crowd sourcing is the activity commonly referred to as a phenomenon in which a large group of people engage in a given task in order to harvest usable information (Estellés-Arolas & González-Ladrón-de-Guevara, 2012). There has been significant growth in utilisation of crowd sourcing as a modus operandi to disentangle the multifaceted issues that exist in the real world. The abundance of data on the World Wide Web coupled with the ability to acquire feedback from crowds has the prospective of altering the manner in which data is synthesized and decisions are made. Brambilla, et al., (2013, p1) outlined that “crowd sourcing can be used to answer questions that are inherently hard for machines but can be handled relatively easily with human input”. This exploratory technique is generally an information-seeking activity where people gradually acquire knowledge about one or more issues of interest.

Meanwhile Chatzimilioudis & Zeinalipour-Yazti (2013) have tested the prospect of using the user’s location as a form of crowd sourcing. Their research creates a trajectory for crowd sourcing activities and data management trends, as the identification of the geographic location of the crowd, will lead to an identification of hot and cold spots in the city. Moreover mobile crowd sourcing platforms such as the Waze connected citizen program have led to an improvement of service delivery and assisted in disaster management (Waze W10, 2014). Crowd sourced data can be analysed from big data collected from social media to identify the various trends circulating on the internet and this data can inform decision making through the storage, processing and analysis of real-time data streams. Map D, Strava Metro, Echo echo and Waze are examples of companies which analyse social media, to analyse people’s views, to identify futurist trends and to advise decision makers in planning.

Musakwa (2014) used social media data to determine commuters’ perception of the high speed railway network, the Gautrain in South Africa. This was made possible with increased access to the internet by commuters. However, currently most of the research which has been carried out only highlights the numerous ways to collect data through crowd sourcing techniques, and little has been done to incorporate this information with geographic information systems to inform decision making and facilitate sustainable developmental practises.

Internet of things
The internet allows the contemporary researcher to access information at any time or place, simply because it allows them to literally access various databases. Riggins & Wamba (2015, p1) have outlined how this “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.” This opens up new possibilities in O_D analysis as people move around with various mobile devices which are constantly sending information to the internet, such as cell phones, tablets and smart watches. This will allow for more accurate location of trip generation and also tracing of the various movement networks.

With time and place having been the two major constraints in origin and destination surveys, the advancements of in crowd sourcing, big data and internet of things present an untapped gold mine of geo-location data. Consequently social behaviour identification through the quantification of various aspects of human behaviour is now a possibility in real time. With
technological advancements machines are now able to handle big data, plus through their ability to process algorithms at real time, they reveal insights about the social media users. These algorithms generate classes which can be used for sieving data according to predefined orders hence leading to a means to analysis existing patterns in the dataset (Big data privacy report, 2014). An example is twitter which uses learning algorithms for analysing big data to inform various twitter users based on their interests the latest trends and news stories.

Data interpretation technique
Kriging models have over the years been used in fields of mining, remote sensing and environmental disciplines to predicate spatial patterns (Cressie, 1991; Auston, 2002; Chahouki, et al., 2010). Kriging can be defined as a geo-statistical local interpolation procedure that utilizes the known locations of data points and distance between them to predict density patterns (Bonaventura & Castruccio, 2005).

Scholars such as Mohammad & Adnan (2011) have utilized kriging to predict bird species occurrences using observed records. Using density maps in GIS which are model-based estimations of data distributions, they used kriging to create ideal and impartial approximations models to predict the location of hot and cold spots for bird species. Their work, consequently forms the basis of this study, as a pre-analysis of the data needs to be done before selecting the appropriate parameters for kriging as a means of ensuring optimal estimates and minimum error.

Public Transport in Gauteng, South Africa
The history of public transport provision (PTP) in Gauteng is driven by various forms of social, economic and political forces that have moulded and shaped it to form its current nature. From the horse drawn cart of the early colonial era; to motorized systems; to the present multi-faced modes encompassing motorized and rail transportation. Khan (2014, p 173-174) have articulated how “the transport landscape in South Africa was largely shaped by colonial and apartheid social and spatial engineering to serve primarily the economic wants and social well-being of the minority white ruling class”. Hence over the years since gaining independence most transportation policies (such as the National Land Transport Transition Act) have tried to advocate for more sustainable means of ensuring the provision of public transportation.

One of the solutions used to regulate transportation is travel demand management (TDM). This with regards to PTP seeks to reduce the amount of motorised travel (Del Mistro & Behrens, 2008), and this has been done in Gauteng through the implementation of the Rea vaya; Metro rail and bus; Ari yang; Putco; Gautrain and Gautbus. However TDM has not been fully implemented as people still prefer to use the mini-bus taxis, as they argue they cater more to their needs as they have more flexible operating hours and that they have successfully penetrated into various their POI. Thus there is still a need to make the formal forms of public transportation more attractive to the commuter.

To address such issues, the National Government identified the use of Intelligent Transportation Systems (I.T.S). ITS refers to “application of data processing, data communications, and systems engineering methodologies with the purpose of improved management, safety and efficiency of the surface transportation network.” (Gauteng 25-year Integrated Transport Master Plan, 2013, p. 6). However ITS are heavy reliant on the collection and analysis of data, to make improvements in travel demand prediction, traffic modelling and O_D surveys. Hence to meet the goals of I.T.S, the following objectives were identified namely being safety; mobility; efficiency; productivity; energy and environment; customer satisfaction (Gauteng 25-year Integrated Transport Master Plan, 2013). Consequently the incorporation of I.T.S in PTP presents a new untapped source of data, and also introducing new aspects to origin and destination surveys such as big data, crowd sourcing, and internet of things. Accordingly there seems to be large market of social media users in South Africa, with Twitter and Facebook having 6.6 million and 11.8 million users respectively (World Wide Worx & Fuseware, 2015). Hence, the aim of the paper was to explore how geographic information systems and geo-location based analytic techniques can be used to define trip generation for the Gautrain nodes.

15. STUDY AREA
The paper mainly was focused on Gautrain Rapid Railway Link (GRRL) and their commuters (figure 1.1). As the Gautrain has been identified as the backbone for public transit provision in the province (Du Plessis, 2010; Gautrain, 2009), and the current gap in knowledge systems exists in how this can become a reality, as the Gauteng province has many inherent problems with regards to public transportation provision. Also given how the Gauteng City Region (GCR) is a cohesive cluster of cities, towns and urban nodes that collectively make up the economic hub of South Africa, generating more than 36% of the country's Gross Domestic Product (GDP), whilst covering less than 2% of the country's total surface area (Gautrain, 2009), an improvement in public transportation system planning becomes necessity to ensure continued sustainable development.

This economical hive is constantly drawing as an influx of commuters traverse through the province on a daily basis, leading to congestion becoming a norm on highways during the peak hours of the day. Bohlwewi Environmental (2002) has outlined how the Gauteng Provincial Government identified the Pretoria CBD, Johannesburg CBD and the airport in East Rand as the most important nodes to be linked by the Gautrain this over time has led to the growth of activities on the other nodes such as Rosebank, Sandton and Hatfield as evident in figure 1.1 (Ruwanpathirana & Perera, 2015). In addition the Gautrain project is still at its inception as the project was only implemented in 2010, and how less than 6 years later, the railway line is still not near completion, with only 10 fully functional train stations, a need for its expansion and integration to other parts of the province still exists.
In the previous year in 2014 the terms Gautrain and station were mentioned 83195 and 19561 times respectively (figure 1.2), on social media platforms namely Twitter and Facebook. With such a high interaction of users, interfacing on social networks, it can be assumed that these are either existing or potential commuters of the Gautrain. Drawing from the background of the statics of how South Africans and the global world have embraced the use of Web 2.0, in their daily lives with Twitter and Facebook having 5 500 00 and 9 600 000 users respectively (Meier, 2013) and the a lack of integration within the public transport provision in Gauteng, which namely encompasses the rail, bus and mini-bus taxis services. The results of the paper shall be used as a means of identifying the extent of trip generation of the various nodes, whilst highlighting areas of crowd clusters, which through collaborative planning could be used as a basis of integrating the existing public transit systems.

Accordingly as this study is premised on the utilisation of social media big data to monitor the points of interest of Gautrain users, that is the demarcation of the sphere of influence of the Gautrain, it becomes evident that privacy concerns arise. As 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 research had to ensure that the data was only used for academic and planning purposes. Also another ethical issue becomes evident that is confidentiality, although information shared on social media platforms is public knowledge, the researcher still could be held liable for any misuse of the data, especially the geographic locations of the posts. Hence the researcher utilised the university’s ethics and code of conduct policy to guide and inform how the data would be safe guarded to protect the interest of the Gautrain and also the social media users.

16. METHODOLOGY

To achieve the goal of the study, the research design adopted an experimental approach which used spatial and quantitative data, in a bid to explore Geographic Information Systems (GIS) techniques which can be used to define trip generation for the Gautrain nodes. This research design hence formed the blueprint of the study from inception to its epilogue. As spatial phenomenon in the real-world is made up three spatial dimensions namely the ‘x’; ‘y’ and ‘z’ with ‘x’; ‘y’ representing geographical co-ordinates and ‘z’ representing elevation, a means to incorporate these in the research was indispensable. Consequently the proposed criteria used for delineating trip generation was established using the model (figure 1.3) which was developed through strenuous trials of various analytic and visualisation modus operandi, until one which showed a real life interpretation of the data was found.
the editing process was repeated until the researcher was satisfied that all the records used in the analysis reflected a true representation of the real world feeds. The raw data was edited as follows, with the following raising red flags, namely being fields missing content and these were removed from the data to be analysed:

- No X co-ordinates
- No Y co-ordinates
- Missing X and Y co-ordinates
- Missing twitter comment
- Missing facebook comment

**Data preparation**

Figure 1.4 shows a summary of the data which was used in the analysis. After data cleaning18633 geo-location social media big data records for the period January to June were uploaded onto Arc Map 10.3, and converted to vector points. The data was then projected to a common projection system (TM 29 Hartebeesthoek 1994). After projecting the data was clipped to the extent of the various sites that is Gauteng; Johannesburg; Pretoria; East Rand. The ‘z’ value for all the sites was determined, by extracting the ‘z’ using the DEM raster layer for the various site extents. Consequently the model was developed using the rule set:-

1. Identify the spatial analytic technique
2. Running the spatial analytic technique
3. Rate the results from the analyses.
4. Comparison of Spatial patterns
5. Identification of clusters and outliers

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Type</th>
<th>Spatial Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Social data</td>
<td>Excel</td>
<td>Undefined</td>
</tr>
<tr>
<td>ZA Dem 90m</td>
<td>Raster</td>
<td>WGS1984 UTM zone 35S</td>
</tr>
<tr>
<td>Wards</td>
<td>Vector</td>
<td>GCS Hartebeesthoek 1994</td>
</tr>
<tr>
<td>Gauteng</td>
<td>Vector</td>
<td>GCS Hartebeesthoek 1994</td>
</tr>
<tr>
<td>Johannesburg</td>
<td>Vector</td>
<td>GCS Hartebeesthoek 1994</td>
</tr>
<tr>
<td>Pretoria</td>
<td>Vector</td>
<td>GCS Hartebeesthoek 1994</td>
</tr>
<tr>
<td>East Rand</td>
<td>Vector</td>
<td>GCS Hartebeesthoek 1994</td>
</tr>
<tr>
<td>Income</td>
<td>Vector</td>
<td>WGS 1984 UTM zone 35S</td>
</tr>
<tr>
<td>Gautrain Stations</td>
<td>Vector</td>
<td>HH94 Lo29</td>
</tr>
<tr>
<td>Gautrain rail tracks</td>
<td>Vector</td>
<td>HH94 Lo29</td>
</tr>
</tbody>
</table>

![Figure 1.4: Dataset](image)

Using the geostatistics wizard, a histogram with 10 bars was created using the z-axis values of the social media big data. An expeditious analysis of the data set was then undertaken this which would show whether the data represented either a normal or abnormal distribution. The histogram consequently shows a very distinct unimodal (one hump) and skewed right, hence relating to the existance both clusters and outliers existed in the data set. In geo-statistical analysis, outliers are considered to be sampling errors however in this analysis these outliers will be analyzed using the criteria weighting.

The ideology of the variogram is premised on the hypothesis that the spatial relation of two sample points does not only depend on their absolute geographical location, but rather on their relative location (Wackernagel 2003). Also Webster & Oliver (2007, p 65) have outlined how “the variogram as a geo-statistical method is a convenient tool for the analysis of spatial data and builds the basis for kriging”. The cloud produced in the variogram represents the lag distances in the data set (Wackernagel, 2003). Subsequently if the data set had produced a discontinuity at the origin, then the height of the discontinuity being the lag distances would be included in the krig, however there seems to be no discontinuity in the data set; and for the sill the value for the data set was $\gamma_{0.10}^5$; whilst the range was 0.80.

![Figure 1.5: Histogram](image)

Kriging was then carried out to visualise these trends identified in the histogram and semi-variogram. The results of kriging from the simulated social media data show that for the various spatial resolution distance bands namely 1km; 1.5km and 2km, there is little variation in the density sizes of the various Gautrain nodes, hence only the 1km spatial resolution distance band was used in the analysis. To conduct the neighbourhood analysis, two methods were done, one using a focal statistics for a fishnet and the other using a count in polygon for the wards data. Initially a fishnet of 5kmx5km was initially created for the Gauteng region and through the use of Geospatial Modelling Environment (Beyer, H.L., 2015), a count in polygon of the social media data was run for the fishnet and the wards data respectively. This created a new field called
count in poly’ in the fishnet and ward data which was then converted to raster format for the focal statistics to be run for the two layers. The analysis on the fishnet was carried out with the neighbourhood being rectangular with a height and width of 5 x 5 km; and secondly with a circular neighbourhood with a radius of 2,850 metres for the fishnet. Lastly for the wards using the count in polygon for the social media data. The data was then reclassified using an interval of 5 classes to visualise the results.

In view of the results an evaluation criterion for the model was then developed through brainstorming and a review of literature. Tools such as list reduction and multi-voting were used to assess which criterion could be used to establish the extent of the clusters and outliers. A respective ranking was consequently produced based on the criterion was then developed with 5 representing areas of high cluster with potential for development and 1 representing areas with little to no potential for development.

17. RESULTS AND DISCUSSIONS

Neighbourhood Analysis
As C. Fiorina has articulated concerning big data, “the goal is to turn data into information, and information into insight.” Hence from the analysis of the focal statistics and comparisons of the results, the wards data shows a clear distribution of hot and cold spots within the study area, with areas near or around the stations having hot spots as shown in figure 1.7. However this representation does not visually represent the real world, as these hot spots take the form or shape of the ward, thus making it difficult to compare density per km². The focal statistics for the fishnet however enables for an unbiased analysis over the surface area per 25 km² as shown in figure 1.8. However using the circular neighbourhood analysis produced a perspicuous hot spots. These hot spot sites can be used as means to justify developing more in these sites. More research is however indispensable to identify whether these as cluster sites of commuters, have any economic; social or historical influence which is attracting them there. Once it has been determined the frequency of the influx of commuters to these sites, the Gautrain railway line or Gaubus may be extent to these sites.
Interpolation
Interpolation was simulated over 100 times using the empirical bayesian krigging technique to reduce the occurrence of errors. By means of the ‘z’ values extracted from the DEM data for the Gauteng province, prediction maps were consequently produced to visualise the trends identified in the preliminary analysis of the social media big data set as shown in figure 1.9. Also it seems that kriging seems quite sensitive to the presence of outliers or misfit values, as some clear indications of cold spots can be seen in northern parts of Pretoria, with the lowest values being recorded falling in the rank 1 in the criteria weighting. Also a hot spot belt seems to emerge in Johannesburg moving towards the East as highlighted in figure. Accordingly from the kriging results the majority of the users seem to be located near the train station locations as shown in figures, this could be due to that the current Gautrain stations are located in melting points of commuters. Examples include the Park node, which is located in the centre of the CBD and acts as an entry point for most regional and local commuters, also given the close proximity to Bree taxi rank and Mtn taxi rank being located only 10 minutes away, this node has a high connectivity level.

As a result the high levels of social media posts around these locations do not necessary mean that the users reside around the train station, but that the train station is located in one of the commuters’ major points of interest. Furthermore the kriging surface and neigbourhood analysis presents hot spots of areas which are currently not being directly serviced by the Gautrain or Gaubus. These are easily identified in areas such as the western and southern parts of Johannesburg; central and eastern parts of East Rand. Accordingly these locations could be areas worth investing into by either expanding the railway lines or bus routes to these as there is clearly a ready market these locations represent points of interest of the potential commuters. The combined maps from the neighbourhood analysis and krig support the study’s hypothesis, and should be used as they visualise the points of interest of the Gautrain commuters using a prediction model with respect to a three dimensional analysis.

18. CONCLUSION

Using the model the study compared and contrast their merits and demerits, depending on the input datasets (Anselin, L. 1996). The GIS techniques offered the researcher various control elements to assist in determining spatial relationships which exist in the datasets. Accordingly the study revealed that the focal statistics presented the most visually accurate means of identifying clusters in the geo-location social media data per square metre. Hot spots were identified in areas near some stations such as Park Station and Sandton, this could mean these have the highest concentration of commuters. Also new hot spots were identified that is areas which are currently not serviced by the Gautrain and these are Soweto and Randburg in Johannesburg; Germiston and Alberton in East Rand; Montana Park in Pretoria. Subsequently these could be possible locations the Gautrain could further investigate as viable locations to expand the railway tracks to. Also through the results from kriging, hot and cold spots are easily identifiable, hence locations with hot spots should be further invested in, and as these are clearly points of interests of the commuters. However further research is still needed, such as running the model whilst incorporating other control factors to determine variations using a time-series analysis, to identify any variations in hot and cold spots over time, thus areas which would present a constant hot spot would clearly be worth investing into.

Also as the purpose of study practices vary, also the applicability of the model should differ. Different scholars conduct research for numerous objectives. Accordingly the current study, was based on demarcating the locations were Gautrain commuters, were coming
REFERENCES


