Supply chain micro-communities in urban areas

Nadia M. Viljoen[a,1], Johan W. Joubert[a,2]

[a]*Centre for Transport Development, Department of Industrial and Systems Engineering, University of Pretoria, Private Bag X20, Hatfield, 0023*

Abstract

An increase in urban freight transport is inevitable as growing urban populations require more goods, more conveniently. A deeper understanding of the geography and trends of urban freight transport must recognise that it is the aggregate result of a complex web of supply chain interactions. To understand the trends, the behaviour of the underlying supply chains must be understood. Using Global Positioning System (GPS) traces of commercial vehicles and network theory concepts, this paper examines the characteristics of supply chain micro-communities in three urban areas in South Africa. The similarity in the structure of these micro-communities across the three, very diverse, areas suggests that the dynamics that drive supply chain interaction are not dependent on local geography. Four prominent archetypes were identified that account for more than half of the micro-communities in each area. Directionality, geographic dispersion and the balance of importance in the micro-communities are studied in the context of these archetypes. This paper presents a first puzzle piece in deducing urban freight transport patterns from supply chain interaction. Furthermore the results are an empirical benchmark that can validate theoretic models of urban supply chain interaction.

Keywords: urban freight, network analysis, supply chain, communities, isomorphism

1. Introduction

UN Habitat estimates that by 2030 approximately 60% of the world's population will live in cities and that 80% of urban growth over the next two decades will take place in African and Asian cities (United Nations Industrial Development Organization, 2018). Urban logistics is one of the necessary evils that accompany urbanisation. Apart from the obvious fact that a growing urban population needs more goods, raised expectations in terms of convenience and choice further increase the demand for logistics services, particularly freight transport. Unfortunately, freight transport also significantly detracts from the liveability of urban spaces. Not only do additional congestion and parking obstructions increase travel times and traffic unpredictability, but even the liveability of urban spaces are hindered by air and noise pollution (Dablanc et al., 2011). The

*Corresponding author
[1]Tel: +27 12 420 2205 (N.M. Viljoen)
[2]Tel: +27 12 420 2843; Fax: +27 12 362 5103 (J.W. Joubert)

impact of urban logistics on the quality of urban spaces and the rate at which urban spaces are developing worldwide has elevated its prominence as a topic of research since 2010 (Lagorio et al., 2016).

When a new supermarket is established in an urban area, it requires that various goods be transported from its many suppliers to the new retail location. These suppliers are often situated on the periphery of urban areas where land is cheaper and zoning accommodates industrial use. A demand for freight transport arises. The logistics sector responds by supplying the necessary services to bring the goods to the store. The demand for freight transport in an area thus emanates from the need to move freight to and from commercial facilities. These facilities are not stand-alone entities, but form part of a greater web of supply chains. The supermarket, for example, may be part of a larger chain of supermarkets that share suppliers, warehouses and transport services. Therefore, the demand for freight transport in an urban area emanates from a complex web of supply chain interactions that require freight to be moved as efficiently as possible.

The textbook representation of a supply chain shows a linear sequence of processes and partnerships that transform raw materials into a final product. In reality, supply chains are not really linear sequences. They are complex networks with autonomous actors. Cause-and-effect is non-linear, hierarchies and structure are ill-defined and just when you think you have it mapped, everything changes. It's more appropriate to think of a supply chain as a Complex Adaptive System (CAS) (Bellamy and Basole, 2013; Choi et al., 2001; Hearnshaw and Wilson, 2013; Kim et al., 2011; Pathak et al., 2007; Tukamuhabwa et al., 2015). The aggregate behaviour of a CAS emerges from the interactions of its autonomous actors. Any policy decisions or interventions that seek to affect the urban logistics behaviour that emerges from complex supply chain interactions must first understand the interactions between supply chain actors on a micro level (Anand et al., 2012; Démare et al., 2017; Marcucci et al., 2017).

Empirical studies that investigate interactions between supply chain actors predominantly use data gathered through surveys. Collecting a sample of such data representative of the overall supply chain system in an urban area would be a punishing effort. It would require expert navigation of confidentiality hurdles and a data collection team of note. Joubert and Axhausen (2011) and Joubert and Axhausen (2013) developed an alternative approach that regards commercial vehicle activity as an indicator of supply chain interactions. Their premise is that if a commercial vehicle travels between two logistics facilities, it represents some sort of supply chain interaction. This approach is arguably less accurate and lacks the capability of defining the nature of the interactions. However, it is far more representative and generalisable as it uses the pervasive Global Positioning System (GPS) traces of tens of thousands of commercial vehicles from all sectors of the economy. Furthermore, the nature of the GPS data enables longitudinal studies.

This paper uses the methodology of Joubert and Axhausen (2011) to identify and investigate the supply chain interactions on a micro level within three urban areas in South Africa. This is done by identifying and investigating the micro-communities of these supply chains. Supply chain literature has before referred to *dyads* or *triads* as building blocks of a supply chain (Choi and Wu, 2009). A dyad consists of a buyer and supplier connected through a transactional relationship. A triad merely adds a third, often intervening, entity such as another supplier or buyer. This distinction seems somewhat arbitrary and theoretical. There is a growing body of knowledge proposing more appropriate models to emulate CAS (Bellamy and Basole, 2013) and Complex Network Theory (CNT) is fast becoming a favourite. From a network theory perspective, we propose that micro-communities are the building blocks of supply chain systems.

The research question addressed is:

> *What are the characteristics of the collection of micro-communities that make up the supply chain network in the Gauteng Province (GT), the City of Cape Town (CoCT) and the eThekwini Metropolitan Municipality (ET) and are these similar across the three urban areas?*

The study used the GPS traces of commercial vehicles that were operational in these three areas during February 2014. A complex network of supply chain activities was extracted from this data for each of the areas. Then concepts from complex network theory and graph theory were applied to extract and investigate the micro-communities.

The paper is structured as follows: In Section 2 we review studies that have applied network theory to modelling supply chains before exploring different data sources that could be used to define empirical supply chain networks. Section 3 briefly introduces the three urban areas. Section 4 elaborates on the dataset and how it was processed. This is followed by the methodology used to extract the initial supply chain networks in Section 5. An overview of the micro-communities extracted from these networks is given in Section 6 before their characteristics are scrutinised in Section 7. The insights are summarised and the paper concluded with mention of future work in Section 8.

2. Related Works

2.1. Modelling supply chain networks

As a CAS, supply chains exhibit both *structural complexity* (the interconnectedness of firms) and *adaptivity* (dynamic learning) (Pathak et al., 2007). A selection of studies that illustrate the ability of CNT to model these two attributes is discussed here.

CNT is a well-suited approach to making sense of the *structural complexity* of supply chains. Hearnshaw and Wilson (2013) set a standard in applying CNT to supply chains by comprehensively mapping different facets of a supply chain to complex network constructs. They investigated which topological characteristics make for an efficient supply chain. Yan et al. (2015) add to this marriage between supply chain management and social network theory by proposing three theoretical typologies based on the concept that within any supply chain there is usually a firm that acts as the "nexus" supplier.

There are also studies that follow a more empirical vein, like the one of Kim et al. (2011) who modelled material flow and relationships in the automotive industry using a complex network framework. This improved on the insights from a previous investigation of the same case study. CNT has come into vogue as the availability of datasets increased. Beckers et al. (2017) challenge the trend of interpreting network analysis results in this domain without qualitative, local knowledge. Their study focusses on the hierarchy of the logistics network in Belgium. In doing so they illustrate the value of combining local knowledge with the results of an iterative community detection algorithm.

These and similar studies provided good descriptions of the structure of a supply chain, but CNT has also been successfully employed to understand the *adaptivity* of a supply chain. Vulnerability is one element of adaptivity and a number of authors have investigated it. The first approach to studying vulnerability uses simulations of random errors and targeted attacks to determine where and how networks break apart, while the second uses concepts from epidemiology to study the spread of risk or damage through a network.

Using the first approach, Thadakamalla et al. (2004) evaluated the resilience of the random, small-world and scale-free topologies and suggested what the implications of their findings could be for real-world supply chains. But the empirical data with which to model such real-world supply chains were not readily available. The closest empirical parallels in the next decade or so was the intense study of global shipping networks made possible through Automatic Identification System (AIS) technology. The initial studies in this domain were more focussed on the topologies of global freight shipping networks and how these related to global economic patterns (Ducruet and Notteboom, 2012; Ducruet and Itoh, 2015). Other studies investigated vulnerability in these networks using the first approach and works like that of Viljoen and Joubert (2016) go as far as to interpret these vulnerability observations in terms of potential supply chain impacts. Although related, these studies modelled freight movement on a highly aggregated level, not on supply chain level.

Nair and Vidal (2011) extended the purely topological perspective of Thadakamalla et al. (2004) by adding inventory flow to the random and scale-free formulations. Meanwhile, in a study customised for military supply chains, Zhao et al. (2011) proposed alternate vulnerability metrics that took into account the fact that different facilities performed distinct functions. Both the addition of inventory flow and consideration of facility function yielded richer insights regarding the vulnerability of the supply chains.

Using the second approach, Basole and Bellamy (2014) studied risk diffusion in networks where nodes represented individual firms and links represented business relationships. Their results showed that small-world networks are more robust than scale-free networks when it comes to risk diffusion. A following study evaluated financial risk diffusion amongst supply networks from the electronics industry spread across North America, Asia and Europe (Basole et al., 2016). Their results showed that networks that were not as dependent on a few central hubs (scale-free) but instead had relationships that connected distant neighbours (small-world) reduced the impact of risk propagation and increased overall network health.

All of these studies used CNT to better understand the structural complexity and adaptivity of supply chains. However, these characteristics are investigated on a high level of granularity. The insights of these and similar studies in the supply chain domain would be complemented by and understanding of what happens on the micro-community level. But such a true understanding cannot be based only on simulated theoretical models.

Among the supply chain focussed studies discussed, only Basole et al. (2016), Beckers et al. (2017) and Kim et al. (2011) used empirical datasets. These datasets are not readily available in the public domain and in all three cases either required relational, financial or sweat capital. It can be very difficult to obtain data with which to model a supply chain network. Very difficult, but not impossible. Apart from these studies other friends from the urban planning domain offer some insights.

2.2. Identifying the locations and functions of logistics facilities

Within urban planning circles the concept of "logistics sprawl" has become a hot topic (Aljohani and Thompson, 2016). Often interchanged by terms like freight sprawl, logistics polarisation, decentralisation and de-concentration, logistics sprawl is the "spatial deconcentration of logistics facilities and distribution centres in metropolitan areas" (Dablanc and Rakotonarivo, 2010). Crucial to studies on this topic is the ability to identify logistics facilities and determine their geographic location. Some studies go further by recording additional attributes such as facility size, function, age or even industry served.

Giuliano and Kang (2018) accessed records from the ZIP Code Business Patterns register that report, for each ZIP code land parcel, the number of business establishments, employment, and payroll. They approximated the location of these business establishments within the land parcel using a density centroid technique. However, these establishments are only reported on a 6-digit industry code level which lumps warehouses and distribution centres with many other types of business establishments that do not send or receive freight per se. Therefore, in her study of freight decentralisation in US metropolitan areas, Cidell (2010) used the County Business Patterns register which provides more detailed information, albeit at a coarser level of geographic granularity. This study did not approximate facility locations, but rather reported facility density on a metropolitan level. A reliable centralised business register provides a standardised reference point, but one's study is limited to the granularity and detail of that register.

Other authors rely on a combination of data sources instead of a single reference set. Dablanc and Rakotonarivo (2010) triangulated a combination of business registers, the archives of the French yellow pages, and contextual information gathered from literature studies and company interviews to locate parcel transport terminals in Paris. The SIRENE database registers all commercial and administrative establishments while the SITADEL database holds the record of all building permits. Combining these with the database from the Paris chamber of commerce and industry and the yellow pages from the French postal company, La Poste, gave the authors an amalgamated dataset from which to extract locations. Contextual data undoubtedly filled in the gaps. In the case of Coetzee and Swanepoel (2017) there were no reliable central business databases. They combined a systematic search of a popular South African business directory with a random survey of 596 properties within five industrial regions nearby the OR Tambo International Airport to identify surrounding air-cargo related businesses. The industrial regions were identified using land-use data. These combined datasets are rich and layered, but not standardised. The heavy reliance on contextual information also poses a risk to repeatability.

Instead of creating their own context-specific triangulation of data sources, Sakai et al. (2015) and Sakai et al. (2017) leverage the comprehensive Tokyo Metropolitan Freight Survey (TMFS) to study the impact of logistics sprawl on truck shipment efficiency. The Transport and Planning Commission of the Tokyo Metropolitan Region in Japan is the custodian of the TMFS, which is carried out roughly every ten years. The survey covers upward of 30 000 establishments in the Tokyo Metropolitan Area. It collects data regarding both standard freight activity measures and facility & business information — making it one of the most comprehensive surveys of its kind (Sakai et al., 2015). Being a public survey, the methodology is required to be standardised and repeatable. However, the cost of such surveys could be prohibitive and few government agencies would be up for the task. This again limits comparisons with other cities and reduces the observation frequency for longitudinal studies.

These studies from urban planning found ingenious ways of identifying where facilities are, but not how they are connected. For this we refer back to Basole et al. (2016), Beckers et al. (2017), and Kim et al. (2011). Beckers et al. (2017) identified the buyer-supplier linkages of nearly 170 000 unique firms in Belgium from a dataset provided by the National Bank of Belgium. These linkages traced the flow of money as the proxy for supply chain relationships. The study conducted by Basole et al. (2016) combined "multiple public and proprietary data sources" to identify a set of focal companies. The relations between these companies were extracted from the Thomson Reuters SDC Platinum and Connexiti, two databases that define interfirm relationships based on financial transactions.

Our study is a preliminary exploration of urban supply chain micro-communities. It is not limited to any one economic sector nor is the methodology intended to be context-specific. We are searching for common trends and observations that can be generalised to all urban areas. Therefore, the greatest limitations of the data sources used in the cited studies are the lack of standardisation, generality and representativity. To overcome these, we adopted the more unconventional approach of using the GPS traces of freight vehicles to identify logistics facilities and the relational links between them (Joubert and Axhausen, 2013; Joubert and Meintjes, 2015a,b).

Vehicle telematics in South Africa is on the rise. In 2014 an estimated 600 000 commercial vehicles (more than 20% of the total fleet) were fitted with GPS tracking devices (Automotive Fleet, 2015). Furthermore, BergInsight (2015) estimated that the penetration of 21.5% in 2015 would rise to 32.5% by 2020. The GPS devices report their traces to owners or third-party vehicle tracking companies. From these traces it is easy to identify when a freight vehicle was moving or standing still. By applying algorithmic rules one can further differentiate when the vehicle was standing still to perform a *minor* activity such as collecting or delivering freight and when it was performing a *major* activity, such as parking for extended periods at a depot or overnight spot (Joubert and Axhausen, 2011). Minor activities occur at logistics facilities, therefore a geographic cluster of minor activities executed by many different vehicles eludes to the presence of a facility. Despite reservations, Joubert and Meintjes (2015a) show that subjective expert analyses can yield fairly accurate, repeatable and reproducible facility identification. They go on to show that automating this process with density-based clustering algorithms holds promise both in terms of accuracy and consistency (Joubert and Meintjes, 2015b).

The limitation of a GPS-based approach is well recognised. GPS traces do not report on the type of vehicle, its owner, the commodities it carries or the purpose of its trip and thus lacks the "depth" offered by other data sources. A number of researchers are developing methods to address this. In studies such as the one by Yang et al. (2014) this limitation is somewhat mitigated as the GPS data were sourced from a single grocery store chain operating in New York. Combining this knowledge with contextual knowledge of the store's operations offered richer insights regarding freight trip behaviour. When data is sourced from a third party telematics provider such contextual knowledge is either not available or is protected by non-disclosure agreements. Ma et al. (2016) used such an anonymous dataset. They applied a density-based spatial clustering algorithm to identify activity chains and non-hierarchical clustering to establish four clusters of similar freight trip behaviour namely local delivery vehicles, vehicles with long loading or offloading times, small package (express) delivery vehicles and owner-operator vehicles. Although their classifications cannot be confirmed beyond a shadow of a doubt, their work makes a significant methodological contribution in activity-based freight trip generation. Contextual information and advanced data-mining techniques are the two most prominent approaches to overcoming this GPS data limitation.

3. Three urban areas

This study investigated the supply chain micro-communities in three urban areas in South Africa namely the City of Cape Town (CoCT) metropolitan municipality in the Western Cape province, the eThekwini Metropolitan Municipality (ET) in the KwaZulu-Natal province and the entire Gauteng Province (GT). The locations of the three areas, in the context of South Africa, are shown in Figure 1.

GT is the nexus of Southern Africa's industrial and economic activity. Geographically it is the smallest province ($18\,178\,\mathrm{km}^2$), yet it is also the most densely populated in the country with a population exceeding 13 million (Statistics South Africa, 2016). Therefore, although GT consists of five separate municipalities, it is considered a *megacity* according to international classifications (United Nations, Department of Economic and Social Affairs, Population Division, 2015). Three of the province's five municipalities are classified as metropolitan. Historically, the discovery of the Witwatersrand gold reef resulted in the establishment of many towns on the east-west axis. There was also a north-south 'pull' with one of the country's capitals (Pretoria, now Tshwane) in the north and the Vaal River in the south (McKay et al., 2017). Through urban development these towns became the cities that make GT a sprawling polycentric megacity. Each city brought with it its own urban development legacy, resulting in a patchwork-like landscape of land use and road infrastructure.

CoCT and ET are two metropolitan municipalities anchored by two of Southern Africa's most prominent seaports. Both municipalities cover a similar area of $2\,445\,\mathrm{km}^2$ and $2\,556\,\mathrm{km}^2$, respectively. Despite these similarities, their urban form and transport geography are dissimilar due to the different ways in which government has tried to repair the urban development legacy of the Apartheid regime. Natural geography also contributes to the differences.

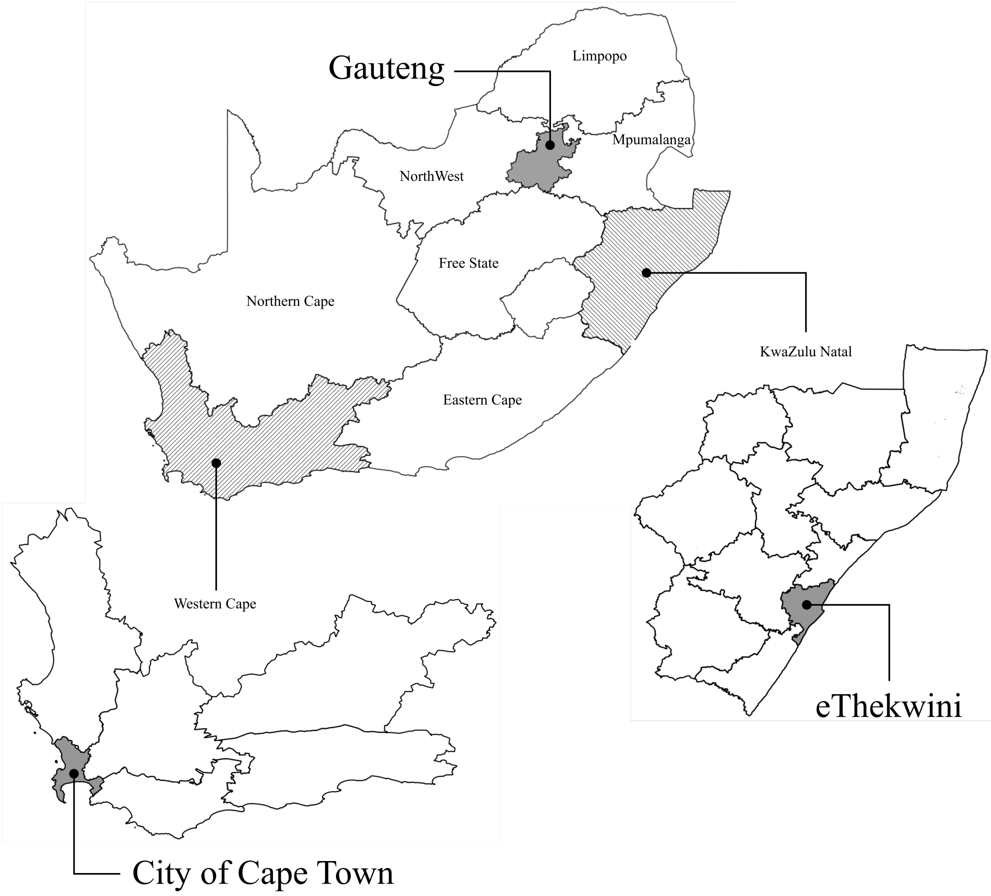


Figure 1: The CoCT and ET metropolitan municipalities and GT province in context of the rest of South Africa.

The urban philosophy that has had the most bearing on supply chains in ET is that of "promoting spatial concentration and efficiency, a compact city model underpinned by the two concepts of an urban core and an urban edge" (Musvoto et al., 2016). This resulted in the bolstering and expansion of the already industrious port city, Durban. Apart from the development in the city core, commercial projects have also spread out from the city along the primary road corridors — particularly along the N2 highway towards the North Coast (Musvoto et al., 2016). Economic activity in ET thus centers around a single nexus and spreads out along the multiple corridor routes that criss-cross the municipality.

Although economic activity in CoCT also originated with its port, industry has become far more diversified, spatially decoupling supply chains from the port itself. CitySpace Planning Cape Town (2012) provides an overview of the economic landscape in the Cape Town Municipal Spatial Development Framework. Although the Central Business District (CBD) is still a hub of activity for the services sector, commercial centres arose inland and along the coast to the east to cater for the medium- to higher-income residential developments located there. The wholesale and retail sectors have taken off in these areas. Manufacturing has also moved from more central industrial areas to the periphery, partly due to reduced land costs and partly due to severe congestion. CoCT is hemmed in by inland mountain ranges and is situated on a peninsula. Thus all traffic into the CBD is channeled onto two primary (mostly choked) road arteries. What remains in the CBD is a concentration of the services sector, especially finance, insurance and business services.

The three urban areas under study are greatly dissimilar in terms of their economic geography. Consequently, their road transport networks and freight traffic trends are also dissimilar. At the outset of this study, the expectation was that this dissimilarity would be reflected in the structure of the supply chain micro-communities.

4. Data collection and processing

Joubert and collaborators maintain a database of Global Positioning System (GPS) traces for tens of thousands of commercial vehicles operational in South Africa from 2010/01 to 2014/05 (53 months). This data were sourced from one of the primary players in the South African commercial vehicle telematics industry. The market drivers underpinning the boom of commercial vehicle telematics in South Africa are protection against hijacking, improved fuel and time efficiencies and monitoring of driver behaviour (Automotive Fleet, 2015). In addition, many insurance companies require vehicle telematics as a prerequisite for cover. These market drivers are relevant to vehicle owners regardless of the size of the fleet, industry sector or area of operation. Therefore, in lieu of more detailed statistics, the assumption is made that the sample of commercial vehicles reporting telemetry data is representative of the total commercial vehicle fleet.

Using the methodology of Joubert and Axhausen (2011), the GPS traces in the dataset were converted into chains of minor and major activities for each unique vehicle. Any commercial vehicle activity chains that had *one or more* of its activities within the urban area during February 2014 were included. As a result, the activity chains of 15 000+ unique commercial vehicles were extracted from the database. The exact size of the commercial vehicle population in South Africa is hard to determine due to the classification scheme of the electronic national administration traffic information system (eNaTIS) that combines both private and commercial vehicles that are larger than station wagons but lighter than 3 500 kg into one category. Based on conservative assumptions, the dataset for February represents between 0.8%–1.9% of the total commercial vehicle fleet and 2.5% of the fleet fitted with telemetry devices (Electronic national administration traffic information system (eNaTIS), 2014).

The sample is biased as it only includes commercial vehicles that chose the specific data partner as their third party provider. However, due to confidentiality, the authors have no further insight into what that bias could be. In 2014, the data provider was one of the primary competitors in the commercial vehicle telematics market. From the GPS traces it is clear that the sample includes vehicles that perform only local activities, vehicles that travel between provinces or all across South Africa and vehicles that cross borders into other Southern African countries.

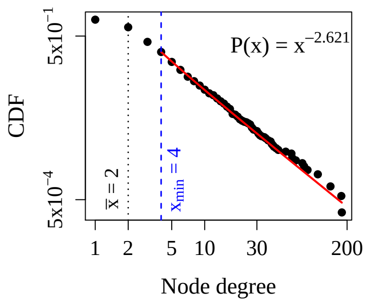
5. Initial supply chain networks

All activity chains that executed one or more activities in the urban areas during February 2014 were used to define the *nodes* and *links* in each of the three urban areas. The nodes represent the logistics facilities that were identified by locating dense clusters of minor activities (Joubert and Meintjes, 2015a). A link existed between two nodes if a commercial vehicle travelled directly between them four or more times during February 2014. The threshold of four was chosen as it relates to approximately weekly trips. Thus the web of supply chains active in each of the three urban areas were summarised in three supply chain area networks.

Every node in these networks has a number of incoming and outgoing links that connects it to other nodes in the network. The degree of the node is the sum of the number of incoming and outgoing links for each node. Figure 2 displays the distributions of this node degree for the three supply chain networks.

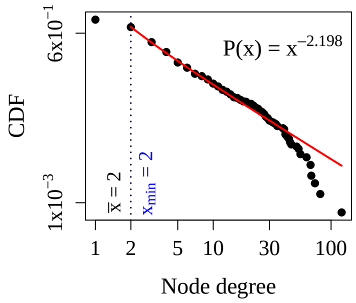
In all three cases the degree distributions obey a theoretical power law. This means that the majority of nodes in the network have relatively few connections to other nodes while a small number of nodes are highly connected hubs within the network. This is clearly illustrated in the graphs by the range from $\bar{x}$ to the extreme values of the x-axis.

Many empirical networks are scale-free (Barabási and Albert, 1999), having degree distributions that obey the theoretical power law $p(x) \propto x^{-\alpha}$ with scale parameter $2 \leq \alpha \leq 3$ (Pastoras-Satorras and Vespignani, 2001). Clauset et al. (2009) cautioned, however, that most of these networks only obey the power law for values of x that lie beyond a specific x_{min}. Using the implementation of Gillespie (2014) in `R`, the x_{min} for each network was determined and power law functions fitted



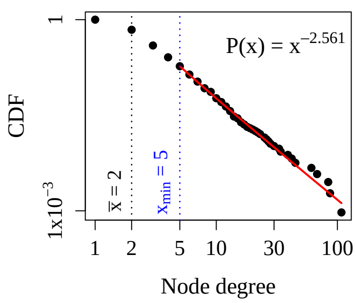
Nodes	3 424
Links	6 222

(a) Gauteng Province



Nodes	1 440
Links	3 308

(b) City of Cape Town



Nodes	1 060
Links	1 965

(c) eThekwini Metropolitan Municipality

Figure 2: Degree distributions of the three supply chain area networks.

accordingly. Using a standard *goodness-of-fit* test we could evaluate our hypothesis. The p-values generated were 0.28, 0.26 and 0.62 for the GT, CoCT and ET networks, respectively. These values were higher than the recommended 0.1 and thus we could accept the hypothesis that the networks are scale-free, for nodes that have a degree greater than x_{min}.

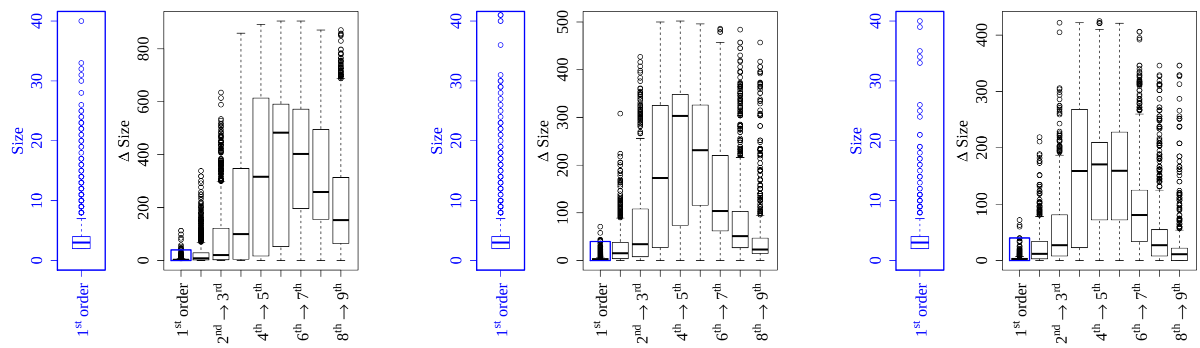
6. Defining micro-communities

Complex Network Theory (CNT) is awash with community detection algorithms and techniques. These algorithms start from the overall network and iteratively segregate communities based on one or more metrics (top-down). A popular algorithm used in scale-free complex networks is the Girvan-Newman (GN) algorithm which separates communities iteratively by cutting the "bridges" (edges with highest betweenness scores) between sub-communities (Fortunato, 2010). Another (bottom-up) approach determines communities by starting with individual nodes and exploring their neighbourhoods. In this section we will explore both approaches to determine which is more suitable to the study at hand.

The First Order Neighbourhood (FON) of a node includes the node itself as well as all of its direct neighbours. The second order neighbourhood of the same node includes the node itself, all its direct neighbours and the direct neighbours' neighbours. Therefore, the order of a neighbourhood indicates the maximum length of the geodesic path between the node in question and any of the other nodes in the neighbourhood. The size of a neighbourhood is the number of nodes it includes. Figure 3 plots the distribution of the sizes of the FONs (blue boxplot on the left of each subplot). Then the increase in neighbourhood size, ΔSize, is measured as the order of the neighbourhoods increase from 1$^{\text{st}}$ to 2$^{\text{nd}}$, 2$^{\text{nd}}$ to 3$^{\text{rd}}$ etc. The right pane in each subplot plots the distributions of ΔSize as the order increases and compares it to the size of the FONs.

In all three urban areas the sizes of the second order neighbourhoods are more than double that of the FONs. As the neighbourhoods increase in their order, ΔSize also inflates drastically. However, the distributions of ΔSize also become much broader, meaning that some neighbourhoods "blow up" in size far more rapidly than others. This increase in ΔSize eventually tapers off. The size and structure of the underlying networks provide a natural limit for how large the neighbourhoods could possibly become.

This paper is aimed at understanding the interrelations of the smallest of supply chain communities — where autonomous actors make decisions. The second order neighbourhoods are already so much larger than the FONs that there was reservation that the complexity of these neighbourhoods would make the waters more murky. Therefore, if using a bottom-up approach to community detection, limiting it to FONs is deemed more appropriate.



(a) Gauteng Province

(b) City of Cape Town

(c) eThekwini Metropolitan Municipality

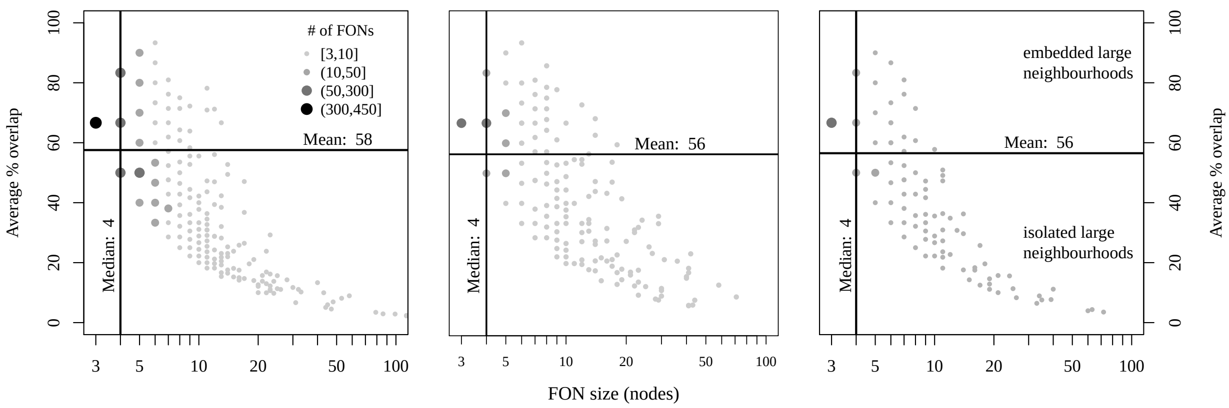
Figure 3: Comparison of the size of the FONs to the increase in size (Δ Size) as neighbourhoods are systematically enlarged.

Initially, the FON of each node was determined, creating a census of FONs for each of the three networks. There were 3 424, 1 440 and 1 060 FONs in the GT, CoCT and ET networks, respectively.

Intuitively many of these FONs overlapped to some degree with some being completely overlapped (swallowed) by larger FONs. The FONs that were swallowed were removed from the census leaving only 1 288, 530 and 416 FONs in the GT, CoCT and ET networks, respectively.

Although none of the remaining FONs are completely overlapped, each FON may have a number of adjacent FONs that partially overlap it to some degree. The average percentage overlap of the FON is thus the mean of the percentage of nodes that are overlapped by each adjacent FON and is plotted against FON size in Figure 4.

Notably the mean of the average percentage overlap and the median of the FON size are close to equal across all three networks. The distributions also have similar shapes. Half of the FONs are small with three or four nodes. In fact, for the GT, CoCT and ET networks respectively, 481 (37%), 145 (27%) and 133 (32%) of the FONs had only three nodes and an average overlap percentage of 66.67%, or two of the three nodes. These results align with the degree distributions (Figure 2). Most facilities are directly connected to only a few neighbours while a minority of facilities have very large FONs. The larger a FON, the lower its overlap with adjacent FONs to the point where the largest FONs are nearly isolated from the rest.



(a) GT (1 288 FONs)

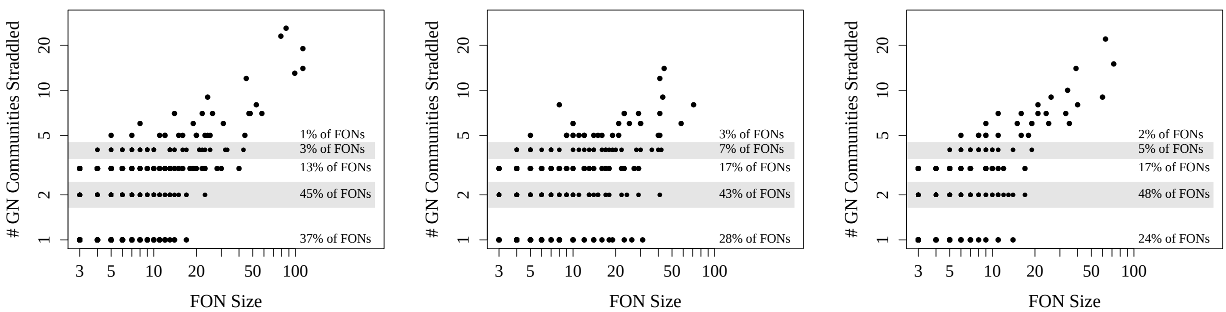
(b) CoCT (530 FONs)

(c) ET (416 FONs)

Figure 4: Percentage FON overlap versus FON size.

Next, we used the GN community detection algorithm from the `igraph` package in `R` (Csardi and Nepusz, 2006) to identify micro-communities using the top-down approach. The GN communities were determined using both weighted and unweighted versions of the networks. Using the weighted networks always yielded more, smaller communities. This is intuitive as more prominent bridges would become apparent when link weights are incorporated in the betweenness calculations. However, the link weight did not play a role in the determination of the FONs, thus it is more appropriate to compare these to unweighted GN communities.

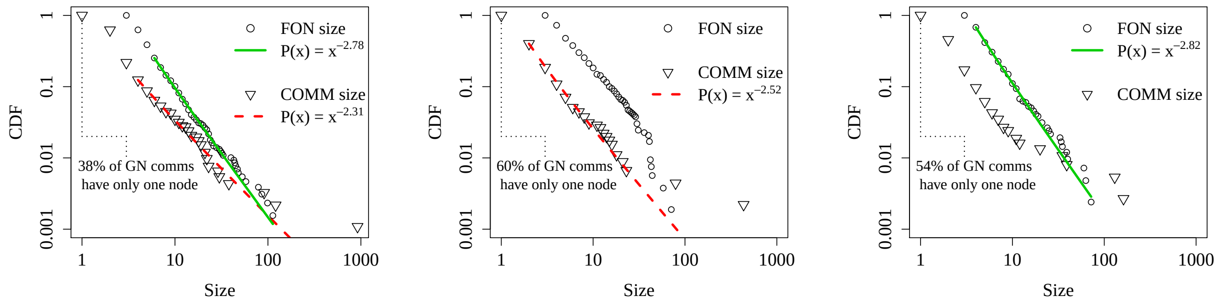
Matching FONs to GN communities was not very successful. Every node in the network belongs to exactly one GN community. In contrast, every node belongs to one *or more* FONs as illustrated by the level of overlap in Figure 4. There were no GN communities that matched a FON exactly (i.e. contained an identical subset of nodes). In two thirds of the cases, FONs included nodes that belonged to two or more GN communities. In Figure 5 each point represents a FON that is plotted according to its size and the number of unique GN communities that FON could be mapped to. The choice of community detection clearly plays a definitive role in the structure of the resulting communities. The choice between the FON and GN approaches came down to an analysis of the sizes (number of nodes) of the micro-communities.



(a) GT network (b) CoCT network (c) ET network

Figure 5: Number of GN communities straddled by each FON according to the FON size.

A comparison of the sizes of the FONs and GN communities shows that the GN communities are smaller than the FONs. The distribution of sizes (Figure 6) approximate power law functions in all three urban areas, meaning that the majority of communities were small, containing less than a handful of nodes. Where the minimum size of the FONs was three, many of the GN communities contained only one node. From a supply chain relationship point-of-view, communities with only one node are not very insightful. Therefore, it was decided to use the bottom-up FON approach instead of the top-down GN approach to analyse supply chain interactions on a micro level. Thus we could conclude from the observations in Figure 4 that the supply chain interactions in each of our three areas have a number of large micro-communities that are closely connected to a central player and then a myriad of very small micro-communities that are enmeshed.



(a) GT network (b) CoCT network (c) ET network

Figure 6: Comparison of community size distributions emanating from the FON and GN approaches. For the GT network the hypothesis that both distributions followed a theoretical power law function could be accepted with p-values greater than 0.1. In the case of CoCT a similar hypothesis could only be accepted for the distribution of the GN community sizes. Meanwhile, for ET the hypothesis could only be accepted for the distribution of FON sizes.

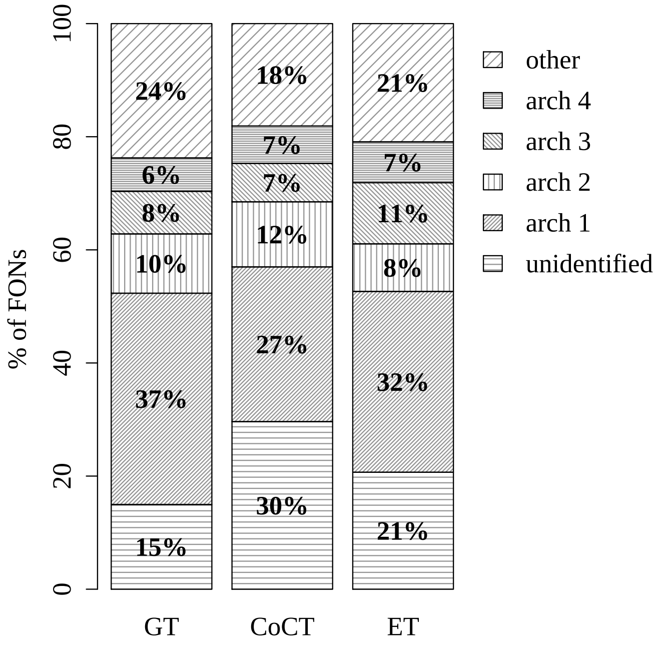
7. Characteristics of the First Order Neighbourhoods

Four characteristics of the FONs were explored to better understand supply chain interactions on the local level. Firstly, to understand the structure of relationships — i.e. how facilities are connected within these micro-communities — common archetypes were extracted. Secondly, it was investigated whether there were any significant patterns in terms of the directionality of the relationships within these communities. Thirdly, the geographic dispersion of the communities were assessed. Finally, the impact of link weights on the balance of importance in the FONs was analysed.

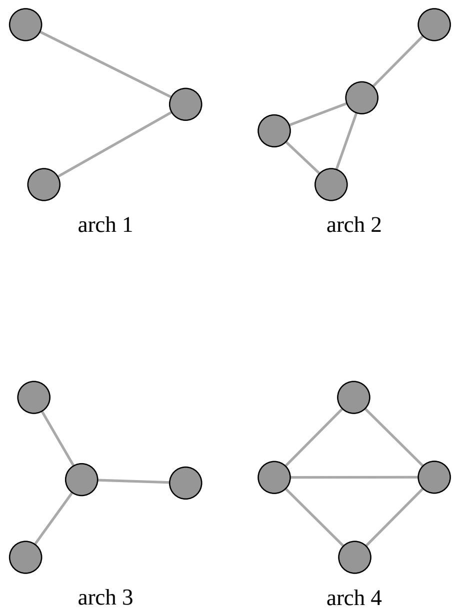
7.1. Common archetypes

To extract common archetypes, the concept of *isomorphism* was used. Two networks are isomorphic if they have the same number of nodes and links and identical degree distributions. In the case of directed networks, it is not only the total degree per node that must be identical, but also the distinction between in- and out-degree so that the two networks match not only in structure but also in directionality. For this initial investigation of FON characteristics, we extracted common archetypes using the undirected versions of the FONs.

A census was made of all archetypes that occurred three or more times in any one of the three urban areas (see Figure 7). Altogether 36 archetypes were identified. This approach left 15% (193) of the FONs in the GT network "unidentified". Similarly, 30% (157) and 21% (86) of the CoCT and ET networks remained unidentified. More than half of the FONs in all three urban areas were characterised as one of the four most prominent archetypes shown in Figure 7b.



(a) Percentage of FONs mapped to identified common archetypes, where "other" is a grouping of arch 5–36.



(b) The four most prominent archetypes recurring in each of the three urban areas.

Figure 7: Common archetypes identified using isomorphism.

The most commonly occurring archetype is the triad (arch 1) where a central node connects two neighbours that are not directly connected to each other. This finding actually reinforces the idea that, on a local level, the direct neighbours of a facility do not interact with each other. This could possibly suggest that interactions or decision-making between supply chain actors is a "private" matter, not involving other neighbours in the micro-community. However, further study is required to confirm this suggestion, especially since the other three prominent archetypes (arch 2, 3 and 4) could challenge this. The second archetype (arch 2) suggests three interconnected neighbours with one actor that is almost an external party to the tightly knit bunch. The third archetype (arch 3) suggests that even on a micro-level, certain facilities play a dominant and coordinating role. Finally, arch 4 suggests a less hierarchical and more connected structure.

Of the 36 archetypes identified, 18 had a hub-and-spoke nature like arch 3. Meanwhile, 10 of the archetypes were less hierarchical and more connected like arch 4. These findings support the rationale behind the three theoretical supply chain archetypes proposed in Viljoen and Joubert

(2017). In that study, the impact of the underlying road network on the vulnerability of each of the three theoretical supply chain archetypes was investigated.

Without further knowledge regarding the functionality of the nodes (i.e. are they retail stores, warehouses etc.), the sectors in which they operate and the companies that own them, one should be careful to make definite statements about the implications of these common archetypes. However, these results do show that there are, in fact, common archetypes to be extracted and that four of these are very prominent. Furthermore, there is similarity among the three, very diverse, urban areas in terms of the most prominent archetypes. A final observation is that Gauteng, the area with the largest supply chain network, has proportionally less uniqueness in terms of the archetypes. Eighty five per cent of the FONs in GT are one of the 36 identified archetypes with more than a third classified as arch 1.

7.2. The influence of directionality

The 36 archetypes discussed in the previous section regard all the FONs as undirected. When the directed nature of the links are factored in, each archetype could be broken down into even smaller groups of FONs that are isomorphic in a directed sense as well. Before delving into such further detail it is necessary to first find the links between what has already been discovered about the structure of supply chain micro-communities and urban logistics behaviour. For this reason, the finer-grained classification of the directed archetypes is postponed to future work. Nonetheless, it is worthwhile to observe any overall trends in directionality of the FONs. Thus it was investigated how many of the FONs are Directed Acyclic Graphs (DAGs). In a DAG the links all point in the same direction so that there are no potential cycles in the graph. Archetypes 1, 2 and 3 all had relatively high proportions of DAG neighbourhoods as shown in Table 1.

Table 1: Percentage of FONs of arch 1, 2 and 3 that are DAGs

Archetype	**GT**	**CoCT**	**ET**
arch 1	55%	39%	53%
arch 2	22%	15%	17%
arch 3	24%	28%	33%
arch 4	3%	3%	7%
other	2%	0%	2%
unidentified	9%	3%	10%

A DAG neighbourhood implies a one-way movement of freight through the neighbourhood.

That is, the commercial vehicles are always travelling in the same direction. This insight could in future be exploited to identify prominent freight-flow highways in the overall supply chain network or to identify facility function in the FONs. Either the origin (source) or final destination (sink) of a certain freight flow is one of the facilities in that FON, or that micro-community acts merely as a conduit for freight movements that were initiated by partners outside of it.

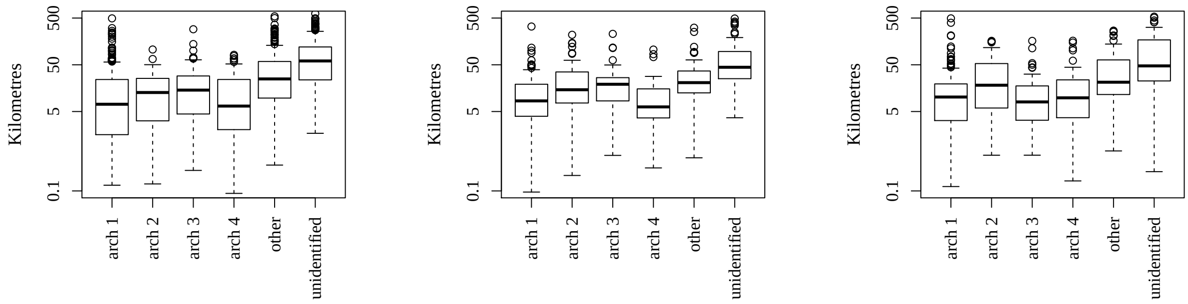
Structure and direction of flow are two important characteristics in supply chains. Another inescapable characteristic is that of geographic distance.

7.3. Geographic dispersion

The question arose whether FONs of different types across the three urban areas would show similar trends when it came to the geographic dispersion of the facilities. Interestingly, the stark contrasts described in Section 3 are not that evident when one considers supply chain micro-communities in these areas. Figure 8 shows the distribution of the diagonal span[3] of FONs in each area according to archetype while Table 2 lists the median. All the distributions display a prominent right tail extending into the hundreds of kilometres. The FONs within the right tail of the distribution would have one or more of their nodes situated *outside* of the urban area in another municipality or even province. This confirms the intuition that long distance freight trips add to the urban logistics activity in each of the three areas.

The sparseness of the right tail in terms of GT deserves some discussion. GT is colloquially referred to as the "gateway into Southern Africa" by politicians. As an inspiring slogan it does no harm but when this viewpoint is erroneously adopted by road freight policy makers it could lead to poor decisions. The fact that there are relatively few connections longer than 100 km in GT echoes the findings of Van Heerden and Joubert (2014) that the majority of urban logistics in the province is *not* the result of inter-provincial commercial vehicles performing logistics activities en-route to some final (international) destination. Direct connections are mostly confined to the province (intra-province). A more viable description is that internationally-bound commercial vehicles probably pass through the province without stopping while freight that is "imported" and "exported" from GT is delivered to or consolidated at facilities on the periphery of the province with the last mile distribution conducted by intra-province commercial vehicles (Van Heerden and Joubert, 2014).

[3]The diagonal span is measured as the Haversine distance of the diagonal of the minimal rectangle that encloses all the nodes in the FON.



y	**Corr.**
FON Size	0.22
Mean link dist.	0.91
Min link dist.	0.63
Max link dist.	0.98
Range link dist.	0.88

(a) GT network

y	**Corr.**
FON Size	0.31
Mean link dist.	0.92
Min link dist.	0.26
Max link dist.	$\to 1$
Range link dist.	0.97

(b) CoCT network

y	**Corr.**
FON Size	0.18*
Mean link dist.	0.94
Min link dist.	0.83
Max link dist.	0.99
Range link dist.	0.91

(c) ET network (*The correlation is not significant with $p = 0.06$)

Figure 8: Distributions of diagonal span for different archetypes across all three urban areas. The tables below the graphs display the Pearson correlation results between diagonal span and the indicated variables (y).

Table 2: Medians of the distributions of diagonal span for different archetypes across all three urban areas.

Archetype	**GT**	**CoCT**	**ET**
	Median diagonal span (km^2)		
arch 1	7	8	10
arch 2	13	15	18
arch 3	14	19	8
arch 4	8	10	10
other	25	21	21
unidentified	60	44	47

Each of the four archetypes are remarkably similar across the three areas. Even compared to each other arch 1–4 are similar in terms of diagonal span. But remember that these archetypes are also nearly identical in terms of the number of nodes they contain.

The remaining 32 archetypes grouped together under “other” have larger diagonal spans overall but again there is similarity across the areas. The sizes of the archetypes range from 5 to 12 nodes. The largest diagonal spans belong to those FONs that are not identified as a specific archetype. The sizes of these FONs range from 5 to 113 nodes. The geographic span of a FON thus seems to be a function of the number of nodes rather than its archetype or the geographic peculiarities of the urban area.

To perform correlation tests between diagonal span and other variables required samples of independent FONs from each urban area. FONs are considered independent if there are no shared nodes (i.e. overlap). Using an algorithmic approach samples of 313 (24%), 114 (22%) and 98 (24%) independent FONs were extracted for GT, CoCT and ET, respectively. Pearson correlation tests were used to determine the correlation between diagonal span (x) and a number of variables (y) as indicated in the tables below Figures 8a–8c. With one exception, all correlations were significant with $p < 0.01$.

It turns out that there is only a weakly positive correlation between diagonal span and FON size. Meanwhile, the linear correlation between diagonal span and the mean and maximum values of the link distances was very strong. This indicates that as the overall area covered by the FON increases, so does the general distance between individual facilities, which is intuitive. However, the correlation between diagonal span and the minimum link distance is not as strong, indicating that even in FONs with extensive geographic dispersion, there are some facilities that are closely situated to one another.

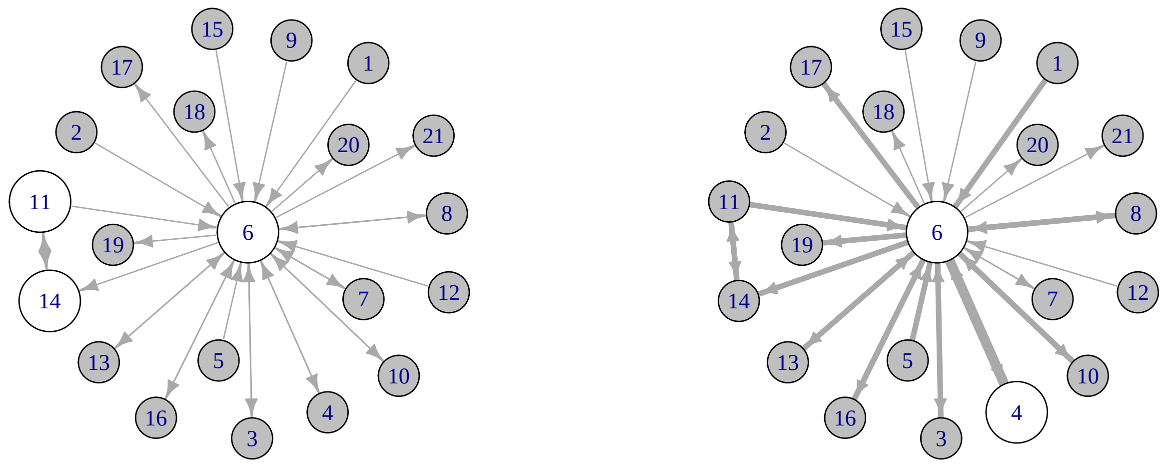
The most striking observation from these results is the similarity of the geographic dispersion across the three urban areas and most prominent archetypes. It should be noted that the Haversine distance used in this analysis is a simplification of the actual road distance. Some may argue that using actual road distances (or even travel time) would yield different results. This is a worthwhile matter for future enquiry.

Finally, no discussion of supply chain relationships would be complete without identifying who the most important players are.

7.4. The balance of importance

The weight of a link represents the number of commercial vehicle trips from one node to another. One perspective could be that this weight is a measure of the strength of the relationship between two nodes. In urban settings where e-commerce, retail distribution and rapid response are growing trends, it is assumed that a strong logistics relationship would result in more frequent truck trips. Thus the relative importance of nodes in the network (often called hierarchy) can be greatly affected by link weights. Our final investigation in this study explores to what degree the "balance of importance" in the FONs is affected when link weights are considered.

Degree centrality is used as the measure of importance[4]. Nodes with higher overall degree (in- and out-degree) are considered to be more important. Figure 9 shows the impact link weights can have when answering the question: "Which are the most important facilities in this supply chain community?" When weights are not considered, the degree centrality of nodes 6, 11 and 14 are in the 95$^{\text{th}}$ percentile. However, when link weights are considered, nodes 6 and 4 are the only nodes in the 95$^{\text{th}}$ percentile. But just how much influence do link weights have in defining the most important nodes in the FONs studied? Table 3 shows that in all three urban areas, in more than 70% of the cases, the set of most important nodes in the FON were exactly the same whether link weights were considered or not. For the remaining FONs, the sets containing the most important nodes were either fully-overlapping, partially-overlapping or distinct sets. Thus for the majority of FONs the link weight does not influence which nodes are identified as most important.



(a) Unweighted. Nodes 6, 11 and 14 are most important.

(b) Weighted. Nodes 6 and 4 are most important.

Figure 9: An illustration of how regarding link weight or not affects the identification of the most important facilities in FON 225 of the ET urban area.

Another key question to ask is: "How much more important are the most important nodes than the rest?" In a typical hub-and-spoke network the hub node would be *far more* important because it has a *much higher* degree. To measure how much more important the most important nodes of a FON are than the rest of the nodes in that FON, the distance (D) between the means of these

[4]The directed acyclic nature of many of the FONs ruled out the use of eigenvector centrality.

Table 3: Percentage of FONs with sets of most important nodes were identical, fully-overlapping, partially-overlapping or distinct when comparing the weighted to the unweighted degree centrality.

Urban area	Identical	Fully-overlapping	Partially-overlapping	Distinct
GT	81%	9%	1%	9%
CoCT	72%	9%	4%	15%
ET	76%	10%	1%	13%

two sets is measured as defined in (1)–(3). $\boldsymbol{I}$ is defined as the set of degree scores, I_n, for the most important nodes. The most important nodes are defined to be those with centrality scores in the 95$^{\text{th}}$ percentile. $\boldsymbol{J}$ is then defined as the set of degree scores, J_n, for all other nodes.

$$\overline{I} = \frac{\sum\limits_{I_n \in \boldsymbol{I}} I_n}{\|\boldsymbol{I}\|} \tag{1}$$

$$\overline{J} = \frac{\sum\limits_{J_n \in \boldsymbol{J}} J_n}{\|\boldsymbol{J}\|} \tag{2}$$

$$D = \overline{I} - \overline{J} \tag{3}$$

If the value of D is relatively large it means that the importance is concentrated among the most important nodes, like in a hub-and-spoke network. To determine whether the link weights tip the scales further towards the most important nodes, the percentage change in D when using the weighted versus unweighted networks is determined as follows:

$$\Delta D = (D_{\text{weighted}} - D_{\text{unweighted}})/D_{\text{unweighted}} \tag{4}$$

Apart from a few outliers, the concentration of importance in the FONs that had identical sets as shown in Table 3 showed very little or no change ($\overline{\Delta D} < 0.001$) when including link weights. The same could be said for the FONs that had fully-overlapping sets and those that had distinct sets. This is interesting in the case of distinct sets as it implies that the concentration of importance remained the same but jumped from one set of nodes to a completely different set of nodes. Only those FONs with partially-overlapping sets saw a marked change in the concentration of importance ($0.04 < \overline{\Delta D} < 0.17$).

Overall it seems that the link weights do not have as significant an impact on the balance of importance in the FONs. This suggests that a node's importance is much more reliant on how many supply chain facilities it connects to rather than the frequency of the interactions between them. However, the caveat to this observation is that the dataset only pertained to one month during which the distribution of link weights were tight around medians of 7 or 8 with a few extreme outliers. Possibly if more than one month were included in the analysis, the link weight distribution would be broader and this could have a more distinct impact on the balance of importance as captured by degree centrality.

8. Conclusion

The motivation to embark on a technical investigation of supply chain micro-communities is the understanding that emergent urban logistics trends, particularly in terms of urban freight transportation, are the result of the interaction of autonomous supply chain actors. Although these actors are part of a dense web of supply chain interactions, freight transport decisions are more likely affected by interactions with close neighbours in micro-communities. This paper used network theory concepts and GPS data to take a first step in exploring how supply chain micro-communities relate.

The value of this study to the field of urban logistics is twofold. Firstly, it presents quantitative network analysis results that elude to potential supply chain behaviour. Heeding the words of Beckers et al. (2017), these results are not yet grounded in contextual or qualitative knowledge and thus should be consumed with caution. The anonymity of the GPS data inhibits contextual analysis and a full understanding of the potential sample bias, but the results serve to guide appropriate qualitative studies to further this field of enquiry. Secondly, this rigourous analysis of an empirical freight transportation network can be used to validate synthetic networks generated for other studies. Agent-based transport simulations in particular could benefit from having these benchmark results with which to validate synthetic populations.

This paper used network theory concepts to define FONs as the building blocks of the supply chains in GT, CoCT and ET. The validity of using this bottom-up approach instead of more common top-down GN community detection algorithm was established.

The validity of defining supply chain micro-communities in terms of FONs was established and four characteristics of these FONs were investigated. It was found that across all three urban areas there emerges four prominent structural archetypes that account for more than half of the FONs. These archetypes have but three or four nodes, supporting the notion that most logistics facilities only have direct freight movement ties with a few other facilities. Most of these smaller FONs overlap to some degree, indicating that although these micro-communities are small, they are also enmeshed and what impacts on one will undoubtedly have spillover effects to the others. Each urban area also has a small number of much larger FONs that are dominated by one node and are not really overlapped by other FONs. A change in the dynamic of these FONs could have a far greater ripple effect in the structure of the network.

Looking at the directionality of the FONs, it was interesting that a high proportion of the four most prominent archetypes are directed acyclic graphs. In these FONs commercial vehicles are always travelling in the same direction. This could indicate that these FONs, are merely conduits on the freight flow highway or that the origin (source) and/or destination (sink) of a specific freight movement is in that FON. This is typically an insight that must be further interrogated by means of contextual data.

Furthermore, it seems that the geographic dispersion of these micro-communities are not influenced by the geographic constraints of the urban area they are in. The four prominent archetypes had similar geographic spans in each urban area and, as discussed, these urban areas have very different urban geographies. The geographic dispersion is also not that closely correlated to the number of facilities in the micro-community.

Finally, the influence of link weight on the balance of importance in each community was investigated. Overall, the importance of facilities within a community is determined more by how many neighbours it has rather than the frequency of vehicle trips to and from those neighbours. Two caveats to this observation is the fact that the data only spans one month and thus the distribution of trip frequency is quite narrow and that all links indicating a trip frequency less than four times per month were filtered out.

Overall, it is striking to see how similar supply chain communities are across three very different urban areas. The observations in this paper open many avenues for future work. Firstly, repeating this analysis over time using the longitudinal data available is a high priority. Secondly, repeating the analysis with GPS data from other urban areas outside of South Africa would also be valuable. Thirdly, the potential sample bias of the dataset and the limitations posed by the data anonymity place constraints on the interpretation of the observations in this study. Understanding the sample bias and resolving the data anonymity through data mining techniques is an ongoing effort amongst the authors.

References

Aljohani, K. and Thompson, R. G. (2016). Impacts of logistics sprawl on the urban environment and logistics: Taxonomy and review of literature. *Journal of Transport Geography*, 57:255–263.

Anand, N., Quak, H., van Duin, R., and Tavasszy, L. (2012). City logistics modeling efforts: Trends and gaps — a review. *Procedia - Social and Behavioral Sciences*, 39:101–115.

Automotive Fleet (2015). Telematics market in South Africa to reach 840k by 2017. Automotive Fleet, Available online: https://www.automotive-fleet.com/128848/telematics-market-in-south-africa-to-reach-840k-by-2017. (Accessed: 9 November 2018).

Barabási, A. and Albert, R. (1999). Emergence of scaling in random networks. *Science*, 286(5439):509–512.

Basole, R. C. and Bellamy, M. A. (2014). Supply network structure, visibility, and risk diffusion: A computational approach. *Decision Sciences*, 45(4):753–789.

Basole, R. C., Bellamy, M. A., Park, H., and Putrevu, J. (2016). Computational analysis and visualization of global supply network risks. *IEEE Transactions on Industrial Informatics*, 12(3):1206–1213.

Beckers, J., Vanhoof, M., and Verhetsel, A. (2017). Returning the particular: Understanding hierarchies in the Belgian logistics system. *Journal of Transport Geography*, In Press.

Bellamy, M. and Basole, C. (2013). Network analysis of supply chain systems: A systematic review and future research. *Systems Engineering*, 16(2):235–249.

BergInsight (2015). Fleet management in South Africa. Technical report, BergInsight.

Choi, T., Dooley, K., and Rungtusanatham, M. (2001). Supply networks and complex adaptive systems: Control versus emergence. *Journal of Operations Management*, 19(3):351–366.

Choi, T. Y. and Wu, Z. (2009). Taking the leap from dyads to triads: Buyer-supplier relationships in supply networks. *Journal of Purchasing & Supply Management*, 15:263–266.

Cidell, J. (2010). Concentration and decentralization: the new geography of freight distribution in US metropolitan areas. *Journal of Transport Geography*, 18:363–371.

CitySpace Planning Cape Town (2012). Cape town spatial development framework, statutory report. Technical report, City of Cape Town.

Clauset, A., Shalisi, C. R., and Newman, M. E. J. (2009). Power-law distributions in empirical data. *SIAM Review*, 51(4):661–703.

Coetzee, P. J. and Swanepoel, P. A. (2017). Spatial relationships and movement patterns of the air cargo industry in airport regions. *Journal of Transport and Supply Chain Management*, 11:297.

Csardi, G. and Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal*, Complex Systems:1695.

Dablanc, L., Diziain, D., and Levifve, H. (2011). Urban freight consultations in the Paris region. *European Transport Research Review*, 3(1):47–57.

Dablanc, L. and Rakotonarivo, D. (2010). The impacts of logistics sprawl: How does the location of parcel transport terminals affect the energy efficiency of goods' movement in Paris and what can we do about it? *Procedia - Social and Behavioral Sciences*, 2:6087–6096.

Démare, T., Bertelle, C., Dutot, A., and Lévêque, L. (2017). Modeling logistics systems with an agent-based model and dynamic graphs. *Journal of Transport Geography*, 62:51–65.

Ducruet, C. and Itoh, H. (2015). Regions and material flows: Investigating the regional branching and industry relatedness of port traffic in a global perspective. *Journal of Economic Geography*, pages 1–26.

Ducruet, C. and Notteboom, T. (2012). The worldwide maritime network of container shipping: spatial structure and regional dynamics. *Global Networks*, 12(3):395–423.

Electronic national administration traffic information system (eNaTIS) (2014). Vehicle population statistics for January/February 2014. eNaTIS, South Africa. Available online from http://www.enatis.com/index.php/statistics/13-live-vehicle-population/383-vehicle-population-statistics-for-januaryfebruary-2014 (Accessed 9 November 2018).

Fortunato, S. (2010). Community detection in graphs. *Physics Reports*, 486(3–5):75–174.

Gillespie, C. S. (2014). *Fitting heavy tailed distributions: the poweRlaw package.* R package version 0.20.5.

Giuliano, G. and Kang, S. (2018). Spatial dynamics of the logistics industry: Evidence from California. *Journal of Transport Geography*, 66:248–258.

Hearnshaw, E. and Wilson, M. (2013). A complex network approach to supply chain network theory. *International Journal of Operations & Production Management*, 33(4):442–469.

Joubert, J. and Axhausen, K. W. (2013). A complex network approach to understand commercial vehicle movement. *Transportation*, 40(3):729–750.

Joubert, J. W. and Axhausen, K. W. (2011). Inferring commercial vehicle activities in Gauteng, South Africa. *Journal of Transport Geography*, 19:115–124.

Joubert, J. W. and Meintjes, S. (2015a). Computational considerations in building inter-firm networks. *Transportation*, 42(5):857–878.

Joubert, J. W. and Meintjes, S. (2015b). Repeatability & reproducibility: Implications of using GPS data for freight activity chains. *Transportation Research Part B*, 76:81–92.

Kim, Y., Choi, T. Y., Yan, T., and Dooley, K. (2011). Structural investigation of supply networks: A social network analysis approach. *Journal of Operations Management*, 29(3):194–211.

Lagorio, A., Pinto, R., and Golini, R. (2016). Research in urban logistics: A systematic literature review. *International Journal of Physical Distribution & Logistics Management*, 46(10):908–931.

Ma, X., Wang, Y., McCormack, E., and Wang, Y. (2016). Understanding freight trip-chaining behavior using a spatial data-mining approach with GPS data. *Transportation Research Record: Journal of the Transportation Research Board*, 2596:44–54.

Marcucci, E., Le Pira, M., Gatta, V., Inturri, G., Ignaccolo, M., and Pluchino, A. (2017). Simulating participatory urban freight transport policy-making: Accounting for heterogenous stakeholders' preferences and interaction effects. *Transportation Research Part E*, 103:69–86.

McKay, T., Simpson, Z., and Patel, N. (2017). Spatial politics and infrastructure development: Analysis of historical transportation data in Gauteng — South Africa (1975–2003). *Miscellanea Geographica*, 21(1):35–43.

Musvoto, G., Lincoln, G., and Hansmann, R. (2016). The role of spatial development framewroks in transformation of the eThekwini municipality, KwaZulu-Natal, South Africa: Reflecting on 20 years of planning. *Urban Form*, 27:187–210.

Nair, A. and Vidal, J. M. (2011). Supply network topology and robustness against disruptions — an investigation using multi-agent model. *International Journal of Production Research*, 49(5):1391–1404.

Pastoras-Satorras, R. and Vespignani, A. (2001). Epidemic spreading in scale-free networks. *Physical Review Letters*, 86(14):3200.

Pathak, S., Day, J., Nair, A., Sawaya, W., and Kristal, M. (2007). Complexity and adaptivity in supply networks: Building supply network theory using a complex adaptive systems perspective. *Decision Sciences*, 38(4):547–580.

Sakai, T., Kawamura, K., and Hyodo, T. (2015). Locational dynamics of logistics facilities: Evidence from Tokyo. *Journal of Transport Geography*, 46:10–19.

Sakai, T., Kawamura, K., and Hyodo, T. (2017). Spatial reorganization of urban logistics system and its impacts: Case of Tokyo. *Journal of Transport Geography*, 60:110–118.

Statistics South Africa (2016). Community Survey 2016 Provinces at a glance. Technical Report ISBN 9780621446661, Statistics South Africa, Pretoria, South Africa.

Thadakamalla, H., Raghavan, U., Kumara, S., and Albert, R. (2004). Survivability of multiagent-based supply networks: A topological perspective. *IEEE Intelligent Systems*, 19(5):24–31.

Tukamuhabwa, B. R., Stevenson, M., Busby, J., and Zorzini, M. (2015). Supply chain resilience: Definition, review and theoretical foundations for further study. *International Journal of Production Research*, 53(18):5592–5623.

United Nations, Department of Economic and Social Affairs, Population Division (2015). World Urbanization Prospects: The 2014 Revision. Technical Report ST/ESA/SER.A/366, United Nations, New York.

United Nations Industrial Development Organization (2018). Sustainable cities: Hubs of innovation, jobs, industrialization and climate action. UNIDO, Vienna, Austria. Available online from https://www.unido.org/sites/default/files/2015-12/13._Sustainable_Cities_Brochure_0.pdf (Accessed 21 March).

Van Heerden, Q. and Joubert, J. W. (2014). Generating intra and inter-provincial commercial vehicle activity chains. *Procedia - Social and Behavioral Sciences*, 125:136–146.

Viljoen, N. M. and Joubert, J. W. (2016). The vulnerability of the global container shipping network to targeted link disruption. *Physica A*, 462:396–409.

Viljoen, N. M. and Joubert, J. W. (2017). The road most travelled: The impact of urban road infrastructure on supply chain network vulnerability. *Networks and Spatial Economics*, 18(1):85–113.

Yan, T., Choi, T. Y., Kim, Y., and Yang, Y. (2015). A theory of the nexus supplier: A critical supplier from a network perspective. *Journal of Supply Chain Management*, 51(1):52–66.

Yang, X., Sun, Z., J., B. X., and Holguín-Veras, J. (2014). Urban freight delivery stop identification with GPS data. *Transportation Research Record: Journal of the Transportation Research Board*, 2411:55–61.

Zhao, K., Kumar, A., Harrison, T. P., and Yen, J. (2011). Analyzing the resilience of complex supply network topologies against random and targeted disruptions. *IEEE Systems Journal*, 5(1):28–39.