International Journal of Geographical Information Science
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/tgis20

Detecting the dynamics of urban structure through spatial network analysis

Chen Zhong[a], Stefan Müller Arisona[ab], Xianfeng Huang[c], Michael Batty[d] & Gerhard Schmitt[a]

[a] Future Cities Laboratory, Department of Architecture, ETH Zurich, Zurich, Switzerland

[b] School of Computer Engineering, Nanyang Technological University, Singapore, Singapore

[c] State Key Lab of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, China

[d] Centre for Advanced Spatial Analysis, University College London, London, UK

Published online: 08 May 2014.

To cite this article: Chen Zhong, Stefan Müller Arisona, Xianfeng Huang, Michael Batty & Gerhard Schmitt (2014): Detecting the dynamics of urban structure through spatial network analysis, International Journal of Geographical Information Science, DOI: 10.1080/13658816.2014.914521

To link to this article: http://dx.doi.org/10.1080/13658816.2014.914521

International Journal of Geographical Information Science, 2014
http://dx.doi.org/10.1080/13658816.2014.914521

Detecting the dynamics of urban structure through spatial network analysis

Chen Zhong[a]*, Stefan Müller Arisona[a,b], Xianfeng Huang[c], Michael Batty[d] and Gerhard Schmitt[a]

[a]*Future Cities Laboratory, Department of Architecture, ETH Zurich, Zurich, Switzerland;* [b]*School of Computer Engineering, Nanyang Technological University, Singapore, Singapore;* [c]*State Key Lab of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, China;* [d]*Centre for Advanced Spatial Analysis, University College London, London, UK*

(*Received 9 September 2013; accepted 9 April 2014*)

Urban spatial structure in large cities is becoming ever more complex as populations grow in size, engage in more travel, and have increasing amounts of disposable income that enable them to live more diverse lifestyles. These trends have prominent and visible effects on urban activity, and cities are becoming more polycentric in their structure as new clusters and hotspots emerge and coalesce in a wider sea of urban development. Here, we apply recent methods in network science and their generalization to spatial analysis to identify the spatial structure of city hubs, centers, and borders, which are essential elements in understanding urban interactions. We use a 'big' data set for Singapore from the automatic smart card fare collection system, which is available for sample periods in 2010, 2011, and 2012 to show how the changing roles and influences of local areas in the overall spatial structure of urban movement can be efficiently monitored from daily transportation.

In essence, we first construct a weighted directed graph from these travel records. Each node in the graph denotes an urban area, edges denote the possibility of travel between any two areas, and the weight of edges denotes the volume of travel, which is the number of trips made. We then make use of (a) the graph properties to obtain an overall view of travel demand, (b) graph centralities for detecting urban centers and hubs, and (c) graph community structures for uncovering socioeconomic clusters defined as neighborhoods and their borders. Finally, results of this network analysis are projected back onto geographical space to reveal the spatial structure of urban movements. The revealed community structure shows a clear subdivision into different areas that separate the population's activity space into smaller neighborhoods. The generated borders are different from existing administrative ones. By comparing the results from 3 years of data, we find that Singapore, even from such a short time series, is developing rapidly towards a polycentric urban form, where new subcenters and communities are emerging largely in line with the city's master plan.

To summarize, our approach yields important insights into urban phenomena generated by human movements. It represents a quantitative approach to urban analysis, which explicitly identifies ongoing urban transformations.

Keywords: polycentric spatial structure; urban movements; complex networks; spatial analysis; smart card data

*Corresponding author. Email: zhong@arch.ethz.ch
Current address of Stefan Müller Arisona: Institute of 4D Technologies, University of Applied Sciences and Arts Northwestern Switzerland FHNW, Bahnhofstrasse 6, 5210 Windisch, Switzerland.

1. Introduction

Urban spatial structure refers to the arrangement of urban space with respect to the set of relationships arising out of urban form and its underlying interactions which are composed of people, freight and materials, and information (Rodrigue *et al*. 2009). These flows have strong effects on transportation, economic growth, social equity, and sustainable urban development, and for a long time, it has been assumed that the actual configuration of urban form that emerges from such interactions is influential with respect to the efficiency, equity, and quality of life of a city's inhabitants. In the past, urban areas have gradually decentralized from their core market centers, and transformations have led to heterogeneous kinds of urban sprawls. Their form has become ever more polycentric consisting of a complex hierarchy of different kinds of centers, neighborhoods, and communities tied together by a multiplicity of transport and information systems. These transformations raise important questions: can one identify this new regime of centers and borders that are emerging from the way people use space for their daily activities through their interactions with one another and with different spaces? In the case of planned developments, do these emerge in the way such development is intended? And can one predict these different kinds of urban forms as a new hierarchy of clusters in terms of the way centers, neighborhoods, and their borders are entangled with one another? Such questions motivate our study and there is now a growing amount of recent research, relevant for planners to validate their designs and to make better use of urban space (Thiemann *et al*. 2010, Rinzivillo *et al*. 2012).

The spatial structure of modern cities has been shaped, in large measure, by advances in transport and communications (Anas *et al*. 1998). The complexity of human movements has redefined the usage of urban space and the arrangement of resources. People, as physical carriers, motivate the transfer of materials, money, people, and information between areas in urban space. Therefore, taking travel as a proxy for spatial interaction, we aim to identify the following essential elements of urban spatial structure:

- **Hubs** refer to the most significant areas that connect spaces between which urban stocks are transferred. These act within the urban structure as spatial bridges between different neighborhoods.
- **Centers** refer to the most relevant areas that accumulate urban stocks, which can differ from hubs but are very often the same.
- **Borders** refer to socioeconomic boundaries that are generated by aggregated travel location choices that subdivide a city into small neighborhoods which we call communities here.

To detect hubs, centers, and borders, we make use of new data sources that record all movements on the public transport systems, which in our case study of Singapore is smart card data from the automatic fare collection system that records the place (stop) and time where a traveler taps in and/or out of the system. Such data sources provide very good resolution of urban mobility covering almost all geographic areas in the city and more than half of the country's 5 million population who use public transport on a daily basis. Using such data sets, we can identify the spatial structure of urban hubs and the borders between them using a variety of network analysis based on the physical properties of the topological and the weighted travel networks and methods for partitioning the network into hubs and centers. Our findings reveal that Singapore is developing towards a polycentric urban form, where new communities are emerging and socioeconomic borders changing. In fact, the time series that we use is very short but given the rapidity with which Singapore is developing – 25 km of metro has been constructed in the last 5 years – we consider that these data provide some sense of how communities are changing. Moreover, the data are only available in this form for these three particular years and although we would prefer a much longer series. We consider this is sufficient to give some sense of how these methods might be developed when more data are available or for other cities where such data are available over longer periods and where the speed of development is much less than that in Singapore.

To anticipate the ultimate outcomes of our analysis, we show the emergence of subcenters and communities for Singapore based on the data for 2010, 2011, and 2012 in Figure 1. In this figure, we show three regionalizations or partitions of the Singapore city-state taken from our network analysis of communities based on Rosvall and Bergstrom's (2008) method. Their method detects communities using random walks to identify the hierarchy of the importance of links and hubs in the network graph, detecting communities at different levels using an entropy decomposition algorithm. From these decompositions, we also show how the eight clusters that are defined grow and morph with one another over the 3-year time series using another technique due to Rosvall *et al*.

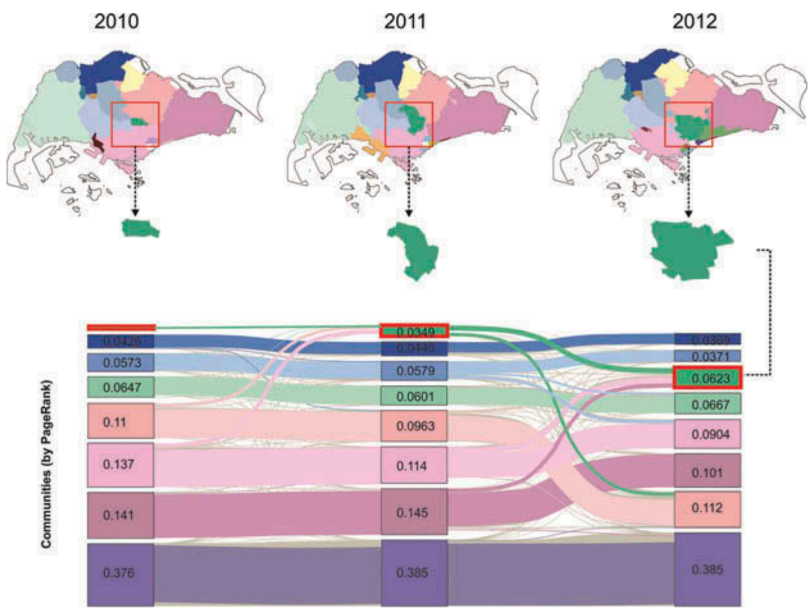

Figure 1. Changing communities and borders detected from daily transportation in Singapore from 2010 to 2012.

Notes: Singapore has been partitioned into smaller neighborhoods emerging from urban movements (top row). A representative emerging new neighborhood is highlighted in the center row. The overall partition of the space and the emerging new neighborhoods over the 3-year time series reveals rapidly changing spatial structure following a polycentric urban transformation. The partitioning of the urban space is conducted by a community detection method applied to a network of urban movements. The alluvial diagram (bottom) shows the changing values of network attributes in terms of significant communities with the highest PageRank (values shown in rectangles), as well as the changing organization among these communities (interchanging flows). All this is explained in detail in the sequel.

(2010), a flow diagram of how these clusters change through time, which they call an 'alluvial' diagram. The rest of this paper explains how we come to these conclusions.

We consider that our analysis is novel and useful in the following sense. First, we provide a quantitative method for detecting urban hubs, centers, and borders as well as changes in the overall spatial structure of urban movement using daily transportation data, and present our methods through an appropriate workflow of the way these techniques are applied. Second, we provide a systematic analysis of linking measured parameters with real urban phenomena, which is applicable to new methods of identifying communities based on mobility, and third, we validate our ideas from novel insights into the actual development of Singapore.

The rest of the paper is organized as follows. In Section 2, we review previous and related work arguing that the novelty of our approach depends on extracting hubs, centers, and borders from networks and flows, rather than from the attributes of locations or areas. In Section 3, we present our approach, the workflow, the explanations of the individual steps involved, and the technical basis of the methods for decomposing large networks such as these into distinct network-based communities. In Section 4, we introduce the case study area, the data set, and we then discuss our results and findings. We conclude in the fifth and final section where we point to open question and unresolved problems.

2. Related and previous research

2.1. Spatial analysis of movement data

Although our work is motivated by relatively recent issues of urban transformation, it addresses essentially a classical spatial aggregation problem. Over the last half century, various urban researchers have developed techniques for decomposing spatially aggregate data and aggregating individual data to spatial areas or zones. These techniques are usually based on meeting some criterion such as homogeneity with respect to various spatial attributes based on demographic, economic, and ethnographic structures, using a variety of multivariate methods that have emerged from social area analysis, transport studies, and geographical analysis. Formal tools for deriving mutually exclusive or overlapping hierarchies in such data were pioneered in the 1960s (Berry 1967) and these were extended to interaction data where densities, volumes, and directions of flows formed the basis for aggregation to specific partitions of the spatial system (Slater 1976, Batty 1978). In fact, although rudimentary network analyses were developed for spatial partitions using graph theory (Nystuen and Dacey 1961), it was not until the rise of network science in the last two decades that a new wave of methods such as the ones we will use here linking flows to planar and topological graphs were developed.

There is now an increasing momentum in the analysis of networks and flows in cities using data that are collected routinely from digital sensors that pertain to the movements of travelers. These kinds of 'big' data are being explored with respect to the new insights it can give into the dynamics of human movement. Many new methods are being applied to new problems. For instance, statistical analysis of human travel behavior using these types of transportation data is being conducted in many cities (Park *et al*. 2008, Liang *et al.* 2009, Munizaga and Palma 2012). In established cities such as London, these patterns of activity and movement are being used to identify the urban spatial structure at a fine scale (Roth *et al*. 2011), where real-time smart card (the 'Oyster-card' in London) data of individual person movements are analyzed to identify the polycentric structure and organization of the central city. In the case of Singapore, stochastic models are being developed to estimate dynamic workplace capacities (Ordóñez Medina and Erath 2013), and to identify urban activities from a synthesis of smart card and survey data (Chakirov and Erath 2012). New centrality measurement has been proposed to identity functional centers (Zhong *et al.* 2013). In other cases, machine learning methods are being introduced to infer land use from mobile phone activity records and zoning regulations (Toole *et al.* 2012) while data mining methods are being explored for discovering patterns (Jiang *et al*. 2012). The availability of large data sets now enables us to discover and verify these various patterns and laws (Song *et al*. 2010, Noulas *et al.* 2012, Simini *et al*. 2012). The detection of urban spatial structure emerging from urban movement in this way is clearly central to estimating the social, environmental, and economic impact of changed activity and movement patterns, and there is now an effort to develop spatial statistics to analyze the spatial impacts of such urban processes, for example, as that reviewed by (Páez and Scott 2005).

In this work, we aim to develop an integrated method based on a synthesis of network science and spatial analysis to detect the changing structure of urban space by making use of new data sources such as smart card data.

2.2. The spatial network approach

Research using network and flow theory with smart card data analysis does not have a very long history, largely because network science has only very recently been extended to deal with spatial networks (Barthélemy and Flammini 2008) and smart card data pertaining to travel on such networks has only just become available. Here, we will give a short review. Conventional research typically applies network analysis to street layouts in terms of their urban topology (Cardillo *et al.* 2006). However, correlation between the accessibility of a street layout with human movements is controversial and remains an open question (Hillier and Iida 2005). Many methods based solely on network topology such as those in space syntax tend to ignore flows, explain urban space, and form primarily in terms of simple concepts of accessibility based on network properties. Recent research however has begun to extend this kind of analysis to incorporate weighted measures, which pertain to human movement data as flows on networks. (Soh *et al*. 2010) used the same data source as we do here, but focus their analysis on the transit system *per se*, not on urban space that is associated with this. In terms of understanding urban space, relevant work using network analysis to find geographical borders between human movements at the regional scale have used GPS tracked vehicle data (Rinzivillo *et al*. 2012), telephone data sets at national scales (Ratti *et al*. 2010), and air transportation data at national and global scales (Guimerà *et al.* 2005, Thiemann *et al.* 2010). Theses 'border' effects were demonstrated by in (Szell *et al*. 2012) as a mechanism behind human movement using data from a multiplayer online game. There is other work regarding regularity of human mobility with network approaches, like in (Sun *et al*. 2013) where a time-resolved in-vehicle social encounter network on the public bus was constructed to discover the hidden encounter small-world in 'familiar strangers' daily life.

This work is motivated by the need to elaborate how properties and configurations of spatial networks can be used to interpret structures of urban movement with specific reference to strong urban planning strategies. Extending previous work, we combine network analysis and spatial statistics and apply the approach to data of different years. Using such combination, changes in urban structure can be effectively detected from the changing distributions of network and spatial properties. In addition, we use smart card data, which is comparatively new but being rapidly introduced for the biggest transit

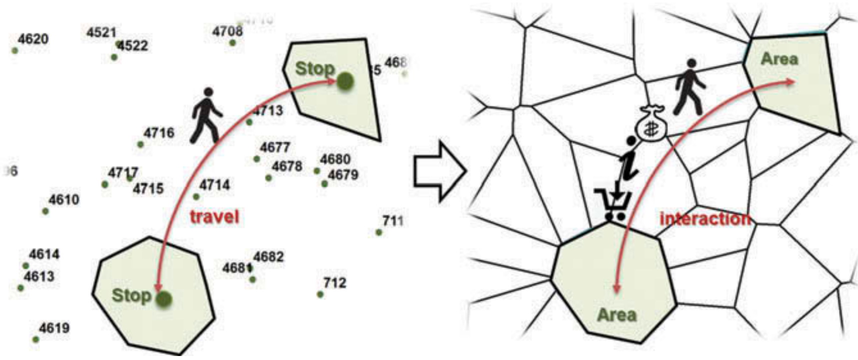

Figure 2. A Voronoi map defining urban spaces generated from the stop locations on the travel network.

Note: People traveling between stops create the physical interactions between any two areas, and this human movement is a proxy for the transfer of urban stocks such as materials products, money, people, information, diseases, and so on.

systems worldwide, generating millions of records per day. In Figure 2, we show a schematic of how the flows of materials, money, people, information, and so on between origins and destinations that we treat as stops on the travel network can be generalized to activities within an urban space around the points in question. We aggregate trips between stops to provide a measure of activity that we compute similarities for in the analysis that follows, and from this, we are able to define the borders around the urban spaces in question. This provides us with proxies for the physical transfer of urban stocks between places and although these are a crude simplification of the homogeneity and heterogeneity of well-defined urban spaces, they represent a first attempt at defining such places with respect to flow networks, linking ideas about regionalization from the 1960s to contemporary network approaches.

3. The method of analysis: urban spaces from network flow data

We propose a method that integrates network and spatial analysis based on a workflow that combines the three stages shown in Figure 3. This starts out with the smart card data obtained from automatic fare collection systems as the input data set to the pipeline. From

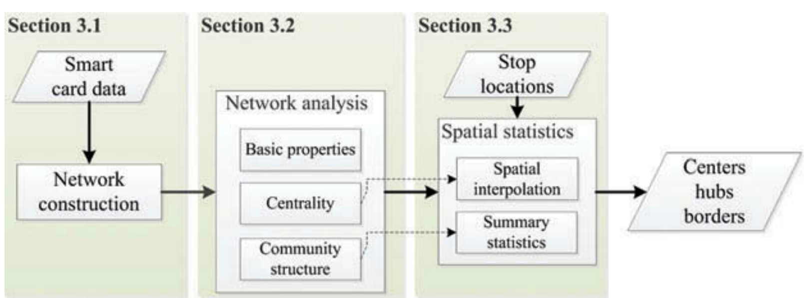

Figure 3. Workflow of the proposed method of analysis.

these data sets, we construct a weighted directed graph that is central to the network analysis that we report below in three subsections. The network analysis includes deriving the basic graph properties that we use to measure various centralities (accessibilities) and from which we derive the different community structures that in turn are associated with the geographical partition of the city into hubs, centers, and their borders. These basic properties provide an overall view of travel demand and interactions in the city. Centrality is used to identify the relative positioning of local areas in the more global spatial structure defining hubs through 'betweenness centrality' measures, centers through the equivalent of 'page rank' analysis, while 'community detection' of network clusters is used to understand the spatial organization of these interaction patterns. The computed results are then mapped onto geographic space, not only for providing an immediate intuitive visualization, but also for further analysis of the spatial impacts using various spatial statistics. We then apply spatial interpolation based on various attributes associated with the stops – the origins and destinations – using the typical distance function to generate what we call the human movement landscape. From this landscape, hubs and centers appear. Summary statistics are finally used to group spatial units of any one community into neighborhoods, from which new borders defining the partition into a contiguous landscape of social–economic spaces are generated. From local scale to global scale, these spaces represent the structure generated by urban movements.

3.1. Network construction and representation

The recorded smart card data contains detailed information on each trip, including trip id, passenger id, age, boarding and alighting time, boarding and alighting location, distance, fare, and an index associated with transfer trips. We first construct an O–D (origin–destination) matrix of travel (trip volumes) between all areas, and then convert it to a weighted directed network. Rather than a travel network, we constructed a social network formed by physical urban activities.

Formally, we define a directed weighted graph as $G \equiv (N, L, W)$ that represents the overall travel on every pair of links in the city during an average workday. It consists of a set N of stops or nodes denoting areas around locations, a set L denoting travel between any two areas, such that L is a set of ordered pairs of elements of N, and a set W denoting the volume of travel – numbers of trips – between any two areas. Hence, $N = \{n_1, n_2, n_3, \ldots, n_I\}$ are the nodes of the graph G, and $L = \{l_1, l_2, l_3, \ldots, l_J\}$ are the J edges of graph G with associated weights $W = \{w_1, w_2, w_3, \ldots, w_J\}$.

3.2. Complex network analysis

'Network anatomy is important to characterize because structure affects function and vice-versa' (Doursat 2005). As indicated previously, we approach network analysis from three perspectives: its global properties, local information pertaining to city hubs and centers, and community detection to identify neighborhoods and their borders. We deal with these in turn.

3.2.1. Global properties

The basic topological and planar properties of a network can reveal important information on spatial interactions. This gives us an overall view of changing travel demands, in particular,

- the number I of *nodes* indicates how many areas are accessible in total, and the number J of *edges* indicates how many areas are directly connected to each other;
- the *degree* of each network node denotes how many areas are directly connected to an area from any other, in terms of their in-degrees – those which contain trip volumes that are destined for that area and out-degrees – those that originate from that area;
- the *strength* is the weighted degree that indicates intensity of travel – trip volumes – to and from one area;
- the *shortest path* refers the minimum network distance possible from one area to another area;
- the *clustering centrality* is an index that measures how 'close'/'cohesive' the areas are to one another in terms of their accessibility to shared neighbors; and
- the *closeness centrality* is an index that evaluates how fast information spreads in the whole area.

By comparing these properties, we can figure out if urban interactions within a city, which are the key elements in the associated weighted graph, are becoming more active or passive with respect to location, and the extent and degree to which these locations are changing or remaining stable.

3.2.2. *Local centrality*

Beyond global properties, we now define two kinds of centrality, the first, which is the well-known measure of *betweenness centrality*, which we use for our definition of a *hub* and second, the *PageRank* which is a measure of accessibility in the network taking account of all direct and indirect links, their weights and their directions. This is a measure we use to define the degree to which each node is a center.

The hub index. Betweenness centrality is an index that measures how well-connected an area is and is key to identifying city hubs. The betweenness centrality of a node k is the number of shortest paths connecting any two areas (nodes) i and j in the graph that pass through the node k. A node has a higher centrality $C_{betweenness}(k)$ the greater the number of shortest paths that traverse it and is defined as follows:

$$C_{betweennss}(k) = \sum_{ij} \delta_{ij}(k) / \delta_{ij} \tag{1}$$

where $\delta_{ij}(k)$ is the number of shortest paths between any two nodes i and j that pass through k and δ_{ij} is the total number of such paths between i and j. Sometimes, this measure is normalized with respect to the total number of nodes N but here we will use it in this basic form.

The center index. PageRank measures the role of a node or local area in attracting flows from all nodes in the network. The measure is a generic representation of the probability of any random walker on a network visiting a particular node and in this sense, it relates directly to a first-order (Markov) probability process that is the basis for many processes of social interaction. In this context, it was originally used for extracting information about Internet link structures and the measure we use here is based on (Rosvall and

Bergstrom 2008) method in which they determine the importance of nodes in a network in analogy to Google's *PageRank* (Brin and Page 1998). In fact, this measure is implicit in the community detection algorithm that is used below to determine community structures. The probability r_j of visiting any node j (or in Google's term, the 'page rank' that is represented as a probability between 0 and 1) is defined as follows:

$$r_j = [(1-\rho)/N] + \rho \sum_i r_i p_{ij} \tag{2}$$

where $1-\rho$ is the probability of the walker j making a random switch to any other node in the network and p_{ij} is the probability of making a switch from node i to j which is proportional to the trip weight on the link i to j, that is

$$p_{ij} = w_{ij} \Big/ \sum_k w_{ik}, \quad \text{and} \quad \sum_j p_{ij} = 1 \tag{3}$$

The steady state probability $\{r_j\}$ is computed by solving the linear simultaneous equations in Equation (2) using iteration, the power method, or the appropriate matrix inversion method. The parameter ρ is a damping factor, which can be set between 0 and 1, but usually is set to 0.85, which we use in this application. If $\rho = 1$, then for all nodes to have a positive probability (for all pages to have a rank), the matrix $\{p_{ij}\}$ must be strongly connected.

3.2.3. *Community structure*

Besides local information, the organization of components (i.e., the 'communities') of the network is crucial for understanding spatial structures. The borders, which subdivide the whole land area, which is covered by the network into smaller neighborhoods, are obtained by detecting what is called *community structure* in network science. The **Border Descriptor** is generated by partitioning the network into two levels where the nodes form modules, which are communities, and the divisions between the modules are the borders. In the case of the network we are dealing with here, we concerned that identifying communities based on the density and interactions of flows that within each community are stronger and in volume terms greater than those between communities: in short, we wish to partition the networks into mutually exclusive clusters that are communities.

Detecting communities in networks has been a fundamental problem in complex network analysis for many years. According to the comparative analysis of different methods (Lancichinetti and Fortunato 2009), the map equation approach called *Infomap* developed by (Rosvall and Bergstrom 2008) is one of the recent algorithms that has shown excellent performance in generating such a two-level hierarchy of clusters and in addition, it is one of the few algorithms suitable for weighted and directed networks. Moreover, *Infomap* considers not only pairwise-relationships, which most partitioning algorithms work with, but also flows between pairs of nodes. It uses the probability flows created from random walks on the graph and the probabilities of visiting a node at random (which is the same as the *PageRank* above) as a proxy for information flows in a real system. It then decomposes the network into clusters by compressing a description of the probability flow in such a way that the average description posed by the probabilities associated with each community and those of the nodes within each community are the most dense and have minimum entropy. In short, the algorithm divides the nodes of the graph into modules or communities that are highly structured, which implies a minimum in the entropy of the partitioned graph.

This entropy is essentially a subdivision of the total entropy of the system into entropy between the modules and a weighted entropy between the modules, these weights being related to the probabilities of the occurrence of each module. Rosvall and Bergstrom (2008) define this entropy as follows:

$$\left.\begin{aligned} Lg(M) &= H(P) + \sum_{i=1}^{m} P_i H(p)_i \\ &= -p\sum_{i}^{m} P_i \log P_i - \sum_{i=1}^{m} P_i \sum_{k=1}^{M_i} \frac{p_k}{P_i} \log \frac{p_k}{P_i} \end{aligned}\right\}, \qquad P_i = \sum_k p_k \tag{4}$$

where P_i is the probability of the module m being visited and p_k/P_i is the probability of the node k which is part of module M_i being visited. These probabilities are not the actual page ranks but the page ranks modified by appropriate exit probabilities as defined in detail by Rosvall and Bergstrom (2008). The way the algorithm works is by first setting each node in its own module and then at each step identifying the node that can be added to a module that decreases the overall entropy in Equation (4). This process continues until no further reduction in entropy can take place and at this point, the number of modules provides a distribution of nodes within communities that is the most organized. Note that M_i is a module, which contains a series of nodes $k \in M_i$ that become stable when the algorithm has converged to minimum entropy. Like all such iterative optimization procedures, simulated annealing or a related procedure is used to ensure that the likelihood that the true optimum has been reached is maximized. This then gives the distribution of nodes, or stops in this case, within each community and this distribution is then mapped to geographical locations.

3.3. Geographical mapping and interpolation

Besides the basic geospatial operations such as geo-referencing, intersection, and mapping, spatial interpolation and summary statistics are the tools we use to transform the discrete network nodes into structured regions.

3.3.1. Spatial interpolation

Spatial interpolation is first applied to generate a movement landscape based on both centrality indices. Such a landscape portrays the properties of each area according to some sample points that in our case are the stops surrounding the area in question where we assume that people choose the nearest stop to their destinations. We thus apply interpolation to the nearest neighbors of each stop. Although there are many variants of interpolation, we use inverse distance weighting (IDW) where each measured point has a local influence that diminishes with distance. The method weights the points closer to the particular location more highly than those further away, and the weights are defined generically for each point as follows:

$$W_i(x,y) = 1 \Big/ d_{ij}(x,y)^{\lambda} \tag{5}$$

where $W_i(x,y)$ is the weight of the location around the point i at coordinates x,y which are nearest neighbor points to j and $d_{ij}(x,y)$ is the distance at x,y from point i towards the nearest neighbor point j. Note that the weights are normalized around a particular point to sum to 1, that is $\sum_{\forall x,y} W_i(x,y) = 1$, and λ is a parameter which is set here as 2, which implies an inverse square law.

3.3.2. Summary statistics

Summary statistics are then used to assign a community to individual spatial units based on the sampled points. The main problem here is to deal with noisy points, which refer to points that belong to a community in network space but are not geographically adjacent to the main cluster defining that community. This clearly emerges because the community detection algorithm is not constrained to achieve geographically contiguous areas, and thus, the communities that are initially detected may have non-contiguous parts in the two-dimensional space. This situation does not occur very often but when it does, it typically occurs in boundary areas where people have different travel preference to nearby centers. To remove these fuzzy boundaries and the noisy points, we essentially count the number of points in different communities and compute a page rank. We then assign boundary points to the nearest communities with the highest page ranks and in this way move boundary points to their geographically closest community where there is ambiguity at their borders. In this way, compact and geographically intact communities are produced which are geographically contiguous and exhaust the whole space. We will now use the pipeline of network processes in the analysis of the smart card data shown in Figure 3 to generate hubs, centers, and borders between the various communities that define the transportation flow patterns on the public transport system in Singapore.

4. Applications to Singapore

4.1. Salient urban characteristics

Singapore is an island city-state with an area of 710 km^2 and a current population of approximately 5.3 million, of whom about 62% (or 3.29 million) are residents, the rest being foreign workers or their dependents. Urban planning and transportation planning have a strong influence on each other and have visibly impacted on Singapore's urban development through a tight planning system that is particularly vigilant with respect to the location of housing and industry. From the 1970s, transportation planning has been a prominent tool in shaping the structure of the city. In 1987, the first line in the Singapore Mass Rapid Transit (MRT) system was opened and the system now covers 102 subway stations, with particularly fast development of the system during the last 5 years with several new lines opening. Today, the land-based public transportation system in Singapore comprises two networks: the MRT and the bus system and more than half of the population are now using public transportation as their main transport mode (Cheong and Toh 2010).

The collected tap-in/tap-out events offer a huge data set, with around 5 million daily travel records, which we have been able to access as smart card data provided by the Singapore Land Transport Authority. This study was conducted using the available smart card records over three sets of workdays in September 2010, April 2011, and September

2012. With data sets in different years, we can begin to evaluate the feasibility of our method in exploring emerging spatial structure in Singapore. A period of 3 years given by three cross sections is a fairly short time to observe changes of urban structure. However, a yearly population growth of nearly 3% per year in Singapore over the last decade may contributes to significant changes even for a short time.

4.2. Mapping travel in Singapore

As indicated previously, an O–D (origin–destination) trip volume matrix is constructed from the original smart card data. Each node in this network denotes an area with one stop inside. From a preliminary analysis of five working days of data, we found that the overall travel activity in Singapore using public transportation system reveals a very regular pattern with the usual morning and evening peaks. The peak hours appear almost exactly at the same time every working day in the same areas and the overall distribution curves are similarly shaped to one another. Therefore, the network we constructed represents urban movement on an average working day and covers the whole array of daily activity types. After data processing, there are 621,731 edges linking 4638 nodes from the 2010 data, 702,803 edges linking 4716 nodes from 2011 data, and 730,885 edges linking 4727 nodes from the 2012 data. Network properties and indices were computed using the *i-graph* package on the *R* platform (http://igraph.sourceforge.net/). Community structure was generated using the tool *Map Equation* (http://www.mapequation.org/). Spatial analysis was conducted on the *ArcGIS* platform (http://www.arcgis.com/).

Figure 4 illustrates of two types of mapping. The left image shows the network mapping at an early stage in the workflow and highlights structure but neglects geographical information so that local changes cannot be detected. On the other hand, the image on the right shows a traditional geographical mapping from which structures can be barely identified, but local relevance is clearly visible. Thus, we attempted to combine the two representations in order to obtain the missing information.

4.3. An overall view of urban movement

Table 1 shows the global network properties for the years 2010, 2011, and 2012, and from the table, some changes can easily be recognized: the number of edges has increased meaning that more areas in Singapore are connected due to increases in the bus and MRT system infrastructures. Strength in terms of trip volume has increased both in total and on average, meaning that the demand for public transportation has been increasing. The lengths of shortest paths have decreased slightly indicating closer connections among areas. The increasing average degree means that each BUS/MRT stop has more connections to other stops/stations, though the total number of stops and stations did not increase from 2010 to 2011. What has led to this increase is probably due to the newly added bus lines and more active human behavior due to an increase in economic and related demand. We can thus say that Singapore is becoming increasingly connected. Though traffic jams still exist, increasing clustering centrality and decreasing closeness centrality shows that transferring between lines and modes in Singapore has gradually become more convenient and efficient. Such observations are consistent with other works such as (Cheong and Toh 2010), who have used travel surveys from different years for comparison. These surveys are conducted every 4–5 years, but cover only 1% of households in Singapore giving about 100,000 records. According to these statistics, there are more than 2 million people using the two transit systems, which generate some 5 million daily records. This clearly

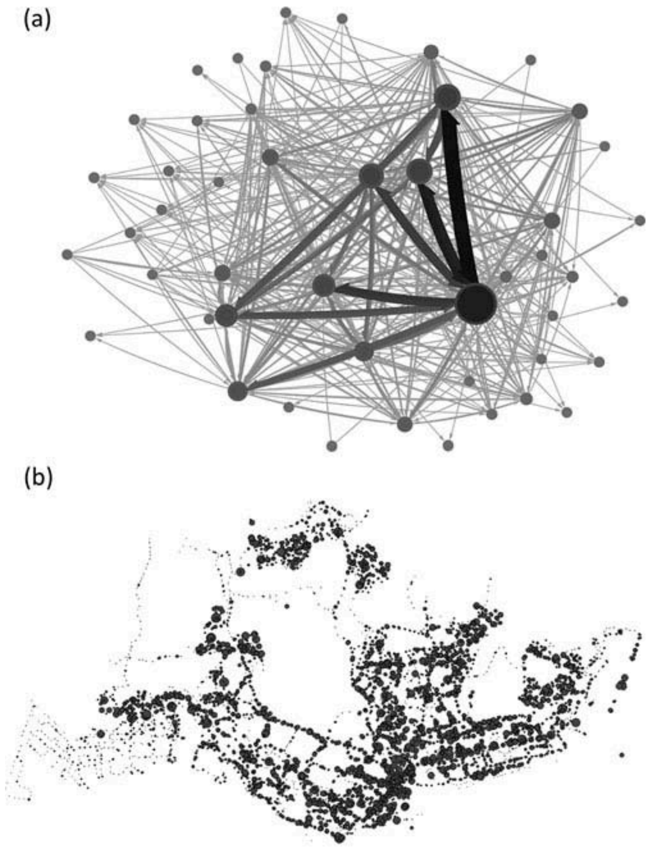

Figure 4. Two varieties of network mapping.

(a) The weighted directed graph constructed from smart card data; nodes represent the module it belongs to, and the larger the nodes, the higher the total PageRank of its module. (b) Nodes mapped into geographical space in proportion to analyzed property values, in this case by node degree that is mapped to node size.

means that extracting smart card data provides more information with respect to making travel behavior more efficient. In addition, traditional surveys consume much more manpower and time.

In the next two sections, we link the various generated parameters from this analysis to certain urban phenomena. We compare the data from different years to show urban change, which is the main objective of this paper. To verify our explanations, we have compared our results with other results in the related literature and to the various urban plans produced for Singapore, and this enables us to draw both qualitative and quantitative conclusions from this analysis.

4.4. City hubs and centers – anomalous centrality

Figure 5 shows a plot of the degree and average trip strength for the years 2010, 2011, and 2012. In the constructed network of human movement, there are a limited number of areas

Table 1. A comparison of basic network analysis parameters with data from 2010, 2011, and 2012.

Year of smart card data	2010	2011	2012
Number of nodes	BUS: 4599 MRT: 107	BUS: 4599 MRT: 107	BUS: 4599 MRT: 117
Number of edges	621,730	702,052	725,046
Average degree	131.8342	148.866	153.4164
Average trip volume by weighted edges	645.5789	788.577	801.2078
Average shortest path length by edges	2.229015	2.196655	2.185142
Clustering centrality	0.2116035	0.2238426	0.2268748
Closeness centrality	1.161199e-06	1.170022e-06	1.085218e-06

that have very high and intense connections to the other areas. Together with the relative short length of the shortest paths in the network, this is indicative of the 'small world' phenomena in the network over each of the years. We must however be cautious of drawing too strong a conclusion in this regard because we are dealing with spatial graphs, which tend to be planar and in their pure form, do not demonstrate small worlds. In Figure 6, we compare the distribution of degrees in 2010, 2011, and 2012 and find that this distribution is becoming slightly more even over time. In other words, it appears that travelers have more diverse location choices for their activities and their average activity spaces are becoming larger.

To gain a deeper understanding of the distributions of these locations, we have projected the various indices into geographical space to determine the locations of hubs and centers. Shown in Figures 7 and 8 are two interpolated maps of our computed centrality index, namely betweenness centrality and *PageRank*. By comparing these two maps, an anomalous distribution appears in those city hubs that are most efficiently connected, but are not necessarily the most central areas. This is a finding that is implicit in our observations even though it tends to fight against our intuition about the role of centrality and accessibility in cities, which traditionally have been monocentric. More specifically in Figure 8, the PageRank map shows that the central area is one of the most visited and most significant places, but also shows that the most efficiently connected areas are not only found in the city center, but in many other areas across the whole island.

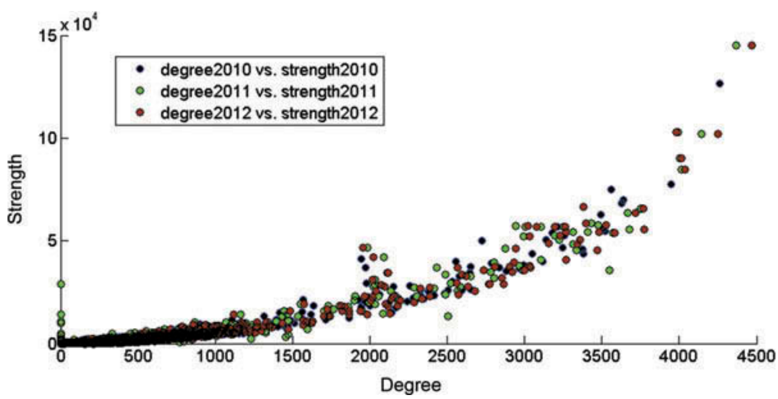

Figure 5. Degree and average trip strength distribution in 2010, 2011, and 2012.

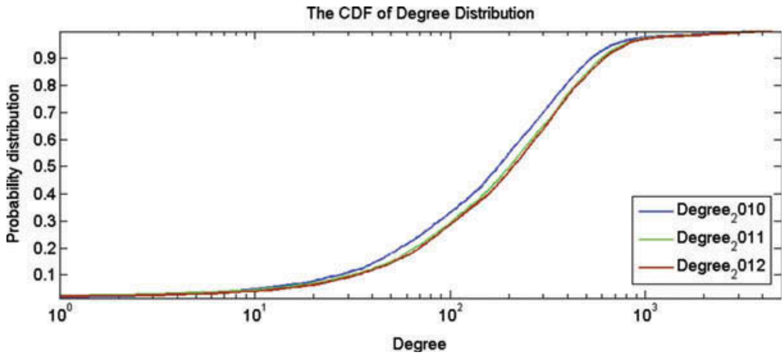

Figure 6. Changing degree distributions in 2010, 2011, and 2012 with the overall distribution becoming slightly more even.

Note: There are few nodes with a very high degree, which results in a very broad tail of the degree distribution. For a better view, we show degrees <1200, shown in a magnified figure (top right).

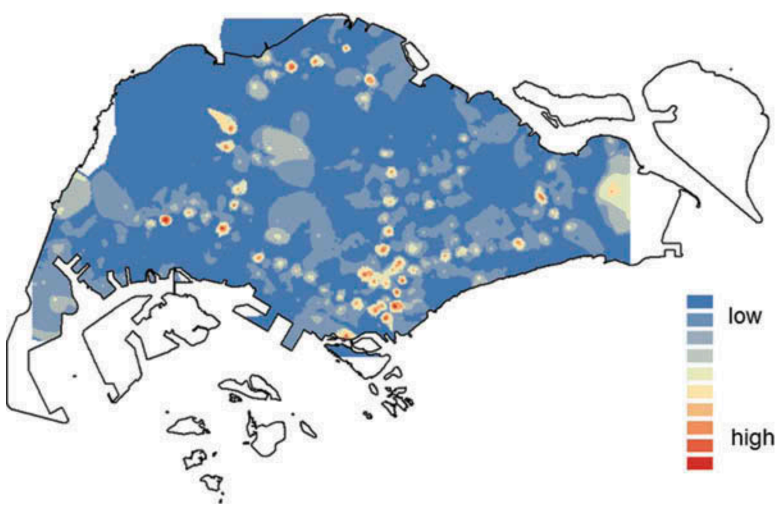

Figure 7. Interpolated betweenness centrality landscape for 2011.

Note: The areas in red are detected hubs that are consistent with locations of the MRT stations.

Indeed, we find that these hub locations are almost perfect matches with key points defined by the MRT lines. This means that the MRT lines have a significant position and serve as the wider skeleton linking all regions of the city-state together. In fact, this finding is consistent with Singapore's physical concept plans. Back in the 1970s, transportation was prominently considered in shaping the structure of the city. According to the various concept plans, high-density public housing areas were planned along high-capacity public transportation lines, near to industrial areas and to other employment. And to an extent, this is now borne out in the patterns of accessibility and transport usage revealed from the smart card data.

The network landscapes are also changing like natural landscapes but these are driven by multiple forces, including new development in the city, advances in the infrastructure of the transportation system, and the way peoples' individual choices have been

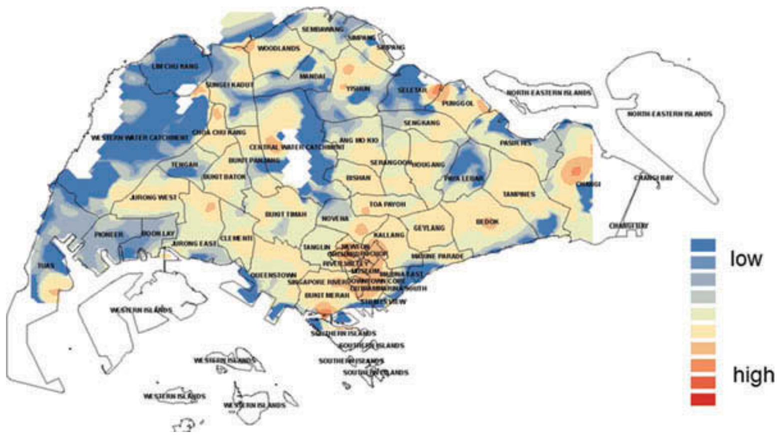

Figure 8. Interpolated PageRank landscape of Singapore for 2011.

Note: The areas in red are detected centers.

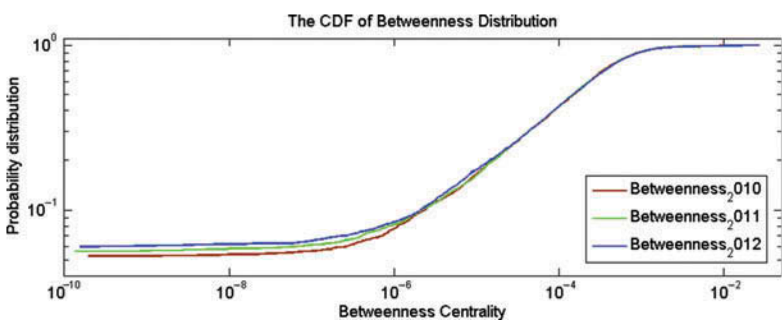

Figure 9. Changing distributions of Betweenness Centrality in 2010, 2011, and 2012.

Notes: The overall distribution becomes more concentrated. Higher Betweenness Centrality is associated with fewer areas.

augmented. From Figure 9, we can see that the number of areas with lower betweenness centrality have slightly decreased, while the number of areas with higher betweenness centrality have increased. This indicates that the most connected areas (the city hubs) largely coincide with MRT stations and these are likely to function more intensively. It also means that the development of the MRT promotes longer distance travel because the population can easily travel to areas that are more central from anywhere in the system. From Figure 10, we have found only slight changes in structure from the *PageRank* distribution. In general, if the number of highly centered areas has deceased while the number of secondary centered areas has increased, this implies a polycentric urban transformation where the influence of strong center areas has gradually relaxed, their centrality increasingly shared with emerging subcenters. However, the slight changes in Figure 10 as well as Figure 9 do not provide us with very strong evidence of urban transformation. As a supplementary analysis, we can reinforce this interpretation from the generated borders of urban movement within different communities that we describe in the next section.

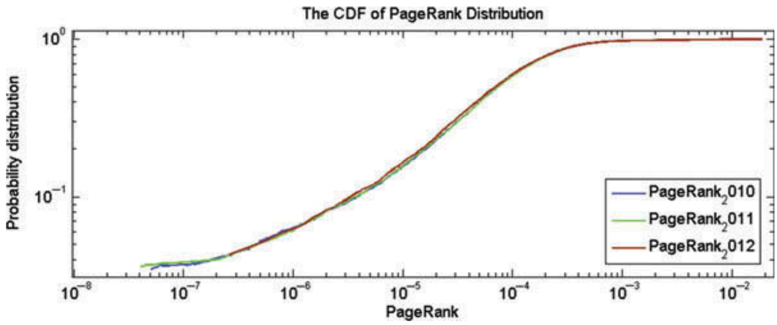

Figure 10. Changing distributions of PageRank in 2010, 2011, and 2012.

Note: The overall distribution shows slight changes while the number of highly centered areas slightly decreases.

4.5. Borders and new neighborhoods – entangled community structure

Borders are important elements that subdivide the entire space into smaller communities. These serve as an important reference for measuring and analyzing the urban data in terms of the original urban structure, the administrative borders, and older city centers, which were planned throughout the twentieth century. They are historical markers that represent past human interactions during the last 100 years. In this section, we generate the geographical borders based on community structure detection and compare these to Singapore's concept plan. The changing communities in terms of volume of flows, number of communities, and their sequences were previously shown in Figure 1, using the concept of the alluvial diagram due to (Rosvall *et al*. 2010) based on data taken from the different community clusters at the three points in time 2010, 2011, and 2012, which we outlined in the previous section. Since our focus is on detecting change, we use only first layer of community clusters. In the case of Singapore, only this layer of communities generates clear geographical partitions of neighborhoods. At lower spatial levels, the neighborhoods are entangled, which indicates a random distribution of peoples' activities in smaller spatial areas.

Zooming into the results for 2012 shown in Figure 11, Singapore can be subdivided into nine small regions that are the most significant communities detected from the network analysis. To clean up the noise in these results, we have aggregated them into subzones equivalent to the smallest levels of geographical subdivision used in Singapore's national statistics. More specifically, we use subzones as the basic spatial unit and then assign them to the most significant communities, which cover the subzones in question. Summing the PageRanks determines the most significant community. The original results before data cleaning can be found in Figure 12.

As introduced earlier, the actual network contains no geographic information *per se*. The community structure is generated from the natural patterns within the network itself. However, after several iterations of the detection algorithm, a clear territorial subdivision emerges, which is consistent with many groupings in cities where people do divide into groups along lines of interest, occupation, age, and so on (Newman 2003). In urban movement, areas can also be divided into groups classified by factors involving the economy and its function. These results show that spatial impact is the most prominent factor that influences people movement in cities and their interaction. When comparing the generated borders of human movement in 2012 that we have extracted and focused

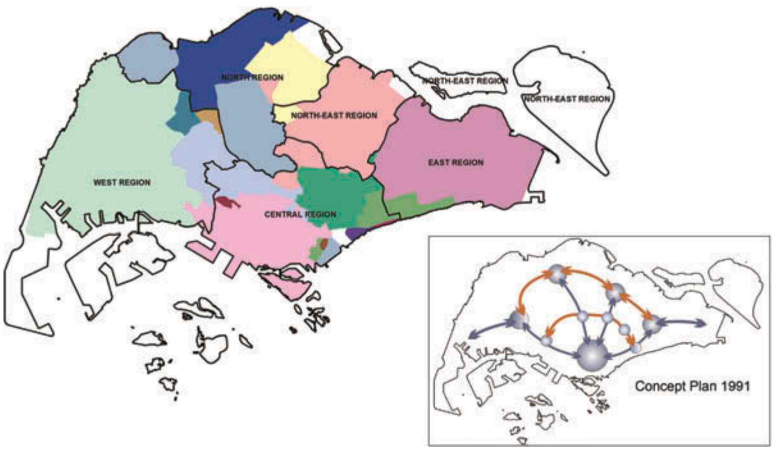

Figure 11. Borders defining communities of urban movement in 2012.

Notes: Community structure detected from smart card data using Infomap marked in different colors. The black boundaries indicate the original administrative borders. In the right corner, planned decentralization of urban form is drawn based on the 1991 concept plan, which is quite in line with the overall structure of urban movements.

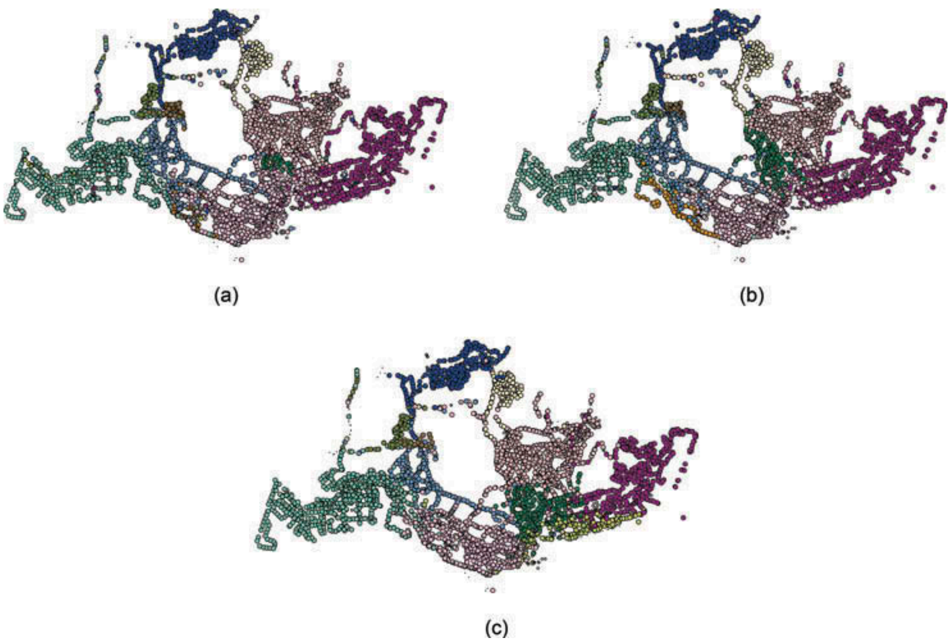

Figure 12. Changing communities from (a) 2010 (b) 2011 (c) 2012.

Note: Nodes denote stops and colors indicate which community they belong to.

upon to administrative borders, it is clear that these borders have shifted a little bit west because of the development of new centers such as the Jurong East area in the west.

At a larger scale, this phenomenon also matches the planned 'decentralization of urban form' which was part of the revised concept plan of 1991 where the emphasis was on facilitating sustainable economic growth through the idea of decentralization. The city was then planned to be surrounded by four regional centers, located in the west, north, northeast, and east, several subcenters and fringe centers, as we show in the inset in Figure 11. This decentralization is part of a top-down panning process that will likely take decades to realize as some sub-centers are still under development. Detecting these trends of change does indeed provide deeper information for planners and designers to evaluate their plans or to link these plans to their actual realization on the ground.

We attempt here to track the path of changes by comparing the analyzed results of the data in 2010, 2011, and 2012 as shown originally in Figure 1. We found that though there are some significant changes in flows between communities, the most important communities remain the same, with only a few changes in their sequence with respect to their summed *PageRanks*. An obvious and gradual change from 2010 to 2011 shows there is an emerging new community. When mapping the nodes as shown in Figure 12, we found that all the nodes in this new community are located in one area, the Bishan, Toa Payoh, and east Novena area. If we refer to the concept plan of new centers shown in Figure 11, the emerging subcommunity consists of one of the subcenters and this suggests that Singapore is slowly becoming more polycentric. Moreover, the emergence of this new community has occurred within only 1 year, illustrating the rapidity of the urban development process in Singapore. The fact that this is only a snapshot of change means that we cannot be certain that these patterns imply the ultimate outcome of these development processes in Singapore.

When comparing these results from 2010 to 2011, we found certain differences with respect to the flows. The difference of the *PageRank* among communities even out a little, which means, the share of flows to each community becomes more balanced. From the geographic perspective, we can see that the areal sizes of communities also becomes even. In addition, an interesting finding is that the south-west area, which is an isolated area in 2011, disappears and is dissolved in adjacent neighborhoods in 2012. The reason for this change is likely to be because of the extension of the MRT lines, which started operation across this area in early 2012, making this region much more accessible to the rest of the network. Even over this short period, our results show how quickly and how strong the transit system influences the pattern of urban movement and the communities that define it. In summary, all these insights from the analysis reveal that the Singapore urban system is becoming ever more polycentric and diverse as developments spread throughout the city-state.

5. Conclusions and future work

Our analysis has revealed an alternative approach to the study of urban dynamics from the traditional more macro analysis of urban structure, and this is primarily due to the availability of new data sources and techniques. Changes in movement, which are not easy to spot in the original data mapping, can be detected through changing centrality and community detection. A qualitative interpretation of the various quantitative indices is also given here and this enriches the analysis with a semantic interpretation that is meaningful to urban planning applications. The results show that even at the urban scale, collective movement still can shape geographic communities as happens in social networks. The methods which we have used can also be applied to other forms of urban movement analysis such as food chain analysis, package delivery, and other systems that involve flow data such as migration, trade, various materials, and of course information between different spatial locations.

This research is only one example of using big data to infer the changes in urban processes and as such this is a prototype. There is still much potential for developing this research further in particular to focus on the periods of time over which we are able to assess change for which we are currently limited by the availability of data. The data sets that we have used have only become available recently and it is likely that we will need a much longer series and a higher temporal resolution before we are able to definitively demonstrate the ultimate pattern of diversity and polycentricity that Singapore is becoming. Moreover, our analysis is limited by the fact that we only use public transportation data here but the public transit system does cover the geographical extent of the city, and as over half of the population uses public transport, we have some confidence that our analysis is revealing relevant trends. The resolution of the data is also good enough to represent the whole urban area, but if we were able to fuse this with other data sources such as taxi and private car data, the analysis could certainly be more accurate. The limitations to our analysis posed by only 3 years of data implies it is clearly not sufficient to depict the path of urban processes and the detailed sequence of urban change but in time we should be able to resolve this. As new data becomes available each year, this type of analysis should be deepened and updated. This prospect has not been possible before but in the case of this research, we stand at the threshold of mounting a long-term project for this kind of travel analysis. Moreover, as follow up work to this paper, further analysis will be done, for instance, using a node-based community detection method to uncover the overlapping and hierarchical neighborhoods; comparing differences in movements between weekdays and weekends; and finding out the causes and sequences of change by adding other thematic data sets using appropriate multivariate methods.

In sum, there is still much to do by focusing on integrated techniques using multiple data sources for studying urban processes. This will undoubtedly contribute to a better understanding of urban dynamics, in terms of human behavior, movements, and urban processes, and we believe the template we have established here shows the direction in which we should go.

Acknowledgments

This work was established at the Singapore-ETH Centre for Global Environmental Sustainability (SEC), co-funded by the Singapore National Research Foundation (NRF) and ETH Zurich. The authors would like to express their sincere gratitude to the Singapore Land Transport Authority for supporting this research and providing the required data.

References

Anas, A., Arnott, R., and Small, K.A., 1998. Urban spatial structure. *Journal of Economic Literature*, 36 (3), 1426–1464.

Barthélemy, M. and Flammini, A., 2008. Modeling urban street patterns. *Physical Review Letters*, 100 (13), 138702. doi:10.1103/PhysRevLett.100.138702.

Batty, M., 1978. Speculations on an information theoretic approach to spatial representation. *In*: I. Masser and P.B. Brown, eds. *Spatial representation and spatial interaction*. Leiden: Martinus Nijhoff Social Sciences Division, 115–147.

Berry, B.J., 1967. Grouping and regionalizing: an approach to the problem using multivariate analysis. *In*: W.L. Garrison and D.F. Marble, eds. *Quantitative geography part 1: economic and cultural topics*. Chicago: Northwestern University, 219–251.

Brin, S. and Page, L., 1998. The anatomy of a large-scale hypertextual web search engine. *Computer Networks and ISDN Systems*, 30 (1–7), 107–117. doi:10.1016/S0169-7552(98)00110-X.

Cardillo, A., *et al.*, 2006. Structural properties of planar graphs of urban street patterns. *Physical Review E*, 73 (6), 066107. doi:10.1103/PhysRevE.73.066107.

Chakirov, A. and Erath, A., 2012. *Activity identification and primary location modelling based on smart card payment data for public transport*. Zurich: Eidgenössische Technische Hochschule Zürich, IVT, Institute for Transport Planning and Systems.

Cheong, C.C. and Toh, R., 2010. *Household interview surveys from 1997 to 2008 – a decade of changing travel behaviours* [online]. Singapore Land Transport Authority. Available from: http://www.ltaacademy.gov.sg/LA-01Journeys_May10_files/J10May-p52Choy&Toh_HIS1997-2008.pdf [Accessed 25 April 2014].

Doursat, R., 2005. Topology and dynamics of complex networks. *In*: *CS 790R Seminar modeling & simulation.* Department of Computer Science & Engineering University of Nevada, Reno, Spring.

Guimerà, R., *et al*., 2005. The worldwide air transportation network: anomalous centrality, community structure, and cities' global roles. *Proceedings of the National Academy of Sciences*, 102, 7794–7799.

Hillier, B. and Iida, S., 2005. Network and psychological effects in urban movement. *In*: A. Cohn and D. Mark, eds. *Spatial information theory.* Berlin: Springer, 475–490.

Jiang, S., Joseph Ferreira, J., and Gonzalez, M.C., 2012. Discovering urban spatial-temporal structure from human activity patterns. *In*: *Proceedings of the ACM SIGKDD international workshop on urban computing,* New York. New York: ACM, 95–102.

Lancichinetti, A. and Fortunato, S., 2009. Community detection algorithms: a comparative analysis. *Physical Review E*, 80 (5), 056117. doi:10.1103/PhysRevE.80.056117.

Liang, L., *et al*., 2009. Understanding individual and collective mobility patterns from smart card records: a case study in Shenzhen. *In*: *Proceedings of the 12th international IEEE conference on Intelligent Transportation Systems*, 3–7 October, St. Louis, MO, 1–6.

Munizaga, M.A. and Palma, C., 2012. Estimation of a disaggregate multimodal public transport origin–destination matrix from passive smartcard data from Santiago, Chile. *Transportation Research Part C: Emerging Technologies*, 24, 9–18. doi:10.1016/j.trc.2012.01.007.

Newman, M.E., 2003. The structure and function of complex networks. *SIAM Review*, 45 (2), 167–256. doi:10.1137/S003614450342480.

Noulas, A. and Añel, J.A., 2012. A tale of many cities: universal patterns in human urban mobility. *PloS One*, 7 (5), e37027. doi:10.1371/journal.pone.0037027.

Nystuen, J.D. and Dacey, M.F., 1961. A graph theory interpretation of nodal regions. *Papers of the Regional Science Association*, 7 (1), 29–42.

Ordóñez Medina, A. and Erath, A., 2013. Estimating dynamic workplace capacities using public transport smart card data and a household travel survey. *Paper presented at Transportation Research Board (TRB) 92nd Annual Meeting*, 2344, 20–30.

Páez, A. and Scott, D.M., 2005. Spatial statistics for urban analysis: a review of techniques with examples. *GeoJournal*, 61 (1), 53–67. doi:10.1007/s10708-005-0877-5.

Park, J.Y., Kim, D.J., and Lim, Y., 2008. Use of smart card data to define public transit use in Seoul, South Korea. *Transportation Research Record: Journal of the Transportation Research Board*, 2063 (–1), 3–9. doi:10.3141/2063-01.

Ratti, C., *et al*., 2010. Redrawing the map of Great Britain from a network of human interactions. *PloS One*, 5 (12), e14248. doi:10.1371/journal.pone.0014248.

Rinzivillo, S., *et al*., 2012. Discovering the geographical borders of human mobility. *KI-Künstliche Intelligenz*, 26 (3), 253–260. doi:10.1007/s13218-012-0181-8.

Rodrigue, J.-P., Comtois, C., and Slack, B., 2009. *The geography of transport systems*. London: Routledge.

Rosvall, M. and Bergstrom, C.T., 2008. Maps of random walks on complex networks reveal community structure. *Proceedings of the National Academy of Sciences*, 105 (4), 1118–1123. doi:10.1073/pnas.0706851105.

Rosvall, M., Bergstrom, C.T., and Rapallo, F., 2010. Mapping change in large networks. *PloS One*, 5 (1), e8694. doi:10.1371/journal.pone.0008694.

Roth, C., *et al.*, 2011. Structure of urban movements: polycentric activity and entangled hierarchical flows. *PloS One*, 6 (1), e15923. doi:10.1371/journal.pone.0015923.

Simini, F., *et al*., 2012. A universal model for mobility and migration patterns. *Nature*, 484 (7392), 96–100. doi:10.1038/nature10856

Slater, P.B., 1976. A hierarchical regionalization of Japanese prefectures using 1972 interprefectural migration flows. *Regional Studies*, 10 (1), 123–132. doi:10.1080/09595237600185121.

Soh, H., *et al.*, 2010. Weighted complex network analysis of travel routes on the Singapore public transportation system. *Physica A: Statistical Mechanics and its Applications*, 389 (24), 5852–5863. doi:10.1016/j.physa.2010.08.015.

Song, C., *et al.*, 2010. Modelling the scaling properties of human mobility. *Nature Physics*, 6 (10), 818–823. doi:10.1038/nphys1760

Sun, L., *et al.*, 2013. Understanding metropolitan patterns of daily encounters. *Proceedings of the National Academy of Sciences*, 110 (34), 13774–13779. doi:10.1073/pnas.1306440110

Szell, M., *et al.*, 2012. Understanding mobility in a social petri dish. *Scientific Reports*, 2, 1–6. doi:10.1038/srep00457

Thiemann, C., *et al.*, 2010. The structure of borders in a small world. *PloS One*, 5 (11), e15422.

Toole, J.L., *et al.*, 2012. Inferring land use from mobile phone activity. *In*: *Proceedings of the ACM SIGKDD International Workshop on Urban Computing*. Beijing: ACM, 1–8.

Zhong, C., *et al.*, 2013. Identifying spatial structure of urban functional centers using travel survey data: a case study of Singapore. *In*: *Proceedings of The First ACM SIGSPATIAL International Workshop on Computational Models of Place*, Orlando FL. New York: ACM, 28–33.