Working Paper

Measuring polycentricity via network flows, spatial interaction, and percolation

Somwrita Sarkar[1], Hao Wu[1], and David M. Levinson[1]

[1]The University of Sydney

{*somwrita.sarkar,hao.wu,david.levinson*}*@sydney.edu.au*

September 2018

Abstract

Polycentricity is most commonly measured by location-based metrics (e.g. employment density or total number of workers, above a threshold, used to count the number of centres). While these metrics are good indicators of location 'centricity', the results are sensitive to threshold-choice. We consider here the alternate idea that a centre's status depends on which other locations it is connected to in terms of trip inflows and outflows: this is inherently a network rather than a location idea. A set of flow and network-based centricity metrics for measuring metropolitan area polycentricity using Journey-To-Work (JTW) data are presented: (a) trip-based, (b) density-based, and, (c) accessibility-based. Using these measures, polycentricity is computed and rank-centricity distributions are plotted to test whether these distributions follow Zipf-like or Chirstaller-like distributions. Further, a percolation theory framework is proposed for the full origin-destination (OD) matrix, where trip flows are used as a thresholding parameter to count the number of sub-centres. It is found that trip flows prove to be an effective measure to count and hierarchically organise metropolitan area sub-centres, and provide one way of dealing with the arbitrariness of

defining a threshold on numbers of employed persons, employment density, or centricities to count sub-centres. These measures demonstrated on data from the Greater Sydney region show that the trip flow-based threshold and network centricities help to characterize polycentricity more robustly than the traditional number or density-based thresholds alone and provide unexpected insights into the connections between land use, transport, and urban structure.

Keywords: Polycentricity, Journey-to-work, origin-destination flows, networks, accessibility, percolation

Introduction

The spatial structure of employment and residential locations shapes accessibility. The idea of *polycentricity*, a city or urban region with multiple centres of employment, economic, or social activity that act as daily attractors for its residential population, has a long history in urban economics, urban planning, and regional science (Anas et al. 1997, Hall & Pain 2006). As the world witnesses accelerating urbanization, the nature of polycentric development is becoming increasingly critical to understand, especially because the locations of residences and employment shape travel demand and travel behavior. Internal city structures may move from monocentric to polycentric or dispersed organization as cities grow in size and density. Concurrently, multiple cities or towns may merge in a process of conurbation, leaving multiple centres in a metropolitan region. Thus, polycentric or monocentric organization shapes the fundamental socio-spatial interaction patterns between transport and land use. Measuring this feature accurately and robustly can help to determine the various efficiencies/inefficiencies of existing urban structure.

Scientific/urban modelling sees polycentricity as a positive concept; it is something to be measured, in order to empirically and theoretically characterize urban structure: through what mechanisms do polycentric agglomerations arise (Fujita & Ogawa 1982, Louf & Barthelemy 2014), and how to accurately measure the numbers and spatial distributions of centres (at spatial scales ranging from metropolitan areas, to regions, to countries) (Green 2007, Meijers 2008, Roth et al. 2011, Vasanen 2012, Liu et al. 2016, Barthelemy 2016).

Urban planning sees polycentricity as a *normative* concept; regions and metropolitan areas

must actively aim for creating, planning, or discouraging polycentricity as an efficiency response to transport challenges posed by growing size and density of urban areas (Meijers 2008, Vasanen 2012, Liu et al. 2016, Green 2007).

Existing measures of polycentricity

The empirical measurement of polycentricity, that affects both the positive and the normative ends has remained an open question (Meijers 2008), despite a wide-ranging variety of approaches (Green 2007, Meijers 2008, Vasanen 2012, Liu et al. 2016, Barthelemy 2016, Giuliano & Kenneth 1991, McMillen 2001, McMillen & Stefani 2001, Louail et al. 2014). Most of these measures are *location and activity-based* rather than *spatial interaction/flow-based*. Some approaches consider flow connectivity from a qualitative perspective (Vasanen 2012). When network structure is formulated mathematically, full flow matrices are summarised into nodal degrees (of locations) and link densities (of connections) (Green 2007). The existing approaches point to the potential of looking at the detailed organisation of network flows in defining polycentricity.

The literature differentiates between *morphological* and *functional* polycentricity, broadly capturing the activity and location-based identification of centres, and the connectivity and flow-based identification of centres, respectively. However, a bridge between the functional and morphological is noted in literature as a gap area needing to be addressed (Meijers 2008).

Several studies note a substantial limitation: the arbitrary nature of defining thresholds on employment density and total numbers of employed persons. The resulting number of centres is extremely sensitive to these thresholds (McMillen 2001, Anas et al. 1997), and non-parametric estimates, while proposed, work primarily with location-based measures and do not consider the network flows in the system. This brings us back full circle to the starting point of turning the focus onto network, flow, and spatial interaction-based measures.

In a related direction, we propose that measurement of polycentricity itself, instead of being measured and analysed in isolation, should be related to accessibility and the spatial structure of jobs and residences. Instead of defining polycentricity as an abstract normative goal, relating accessibility analysis to polycentricity provides more pragmatic and implementable goals for land use and transport planning, or locational planning in terms of employment and residential growth. The measurement of polycentricity through the lens of accessibility will enable the morphological-

structural view to be integrated with the functional-flow view of polycentricity.

In summary, polycentricity is traditionally defined as a location in the city showing more employment/activity concentration compared to some threshold average (Giuliano & Kenneth 1991, McMillen 2001, McMillen & Stefani 2001, Louail et al. 2014). However, the spatial interactions and flows that shape such activities at locations should also be explicitly accounted for. A location should be defined as a 'hotspot' or 'centre' not only when it attracts more people than a threshold, but when *it attracts more 'net inflow' in relative comparison to all the other places it is connected to, and attracts flow from.* Thus, a location-based metric, such as employment density, measures absolute concentrations of workers. But the net-inflow-based measures would measure relative concentrations of workers: a place could be a sink attracting large numbers of workers, but also acting as a source for other sinks, which is distinct from a place which is a sink in both absolute and relative terms. Thus, a good measure is needed for measuring the relative destination-ness or origin-ness of a location. As we show in this work, the relative centre-ness of a location produces hierarchical organizations in space defining differing reaches into source regions, and could be used to measure a number of different normative criteria, such as spatial mismatches between workers and jobs, or planning of the transport network for better accessibility.

The number of centres identified is related to ideas of overall spatial structural organization of a city and to accessibility. We discuss these ideas next, in order to establish the need for developing network and flow-based measures of polycentricity.

Christaller's Central Places and Zipf's Law

Central place theory, developed by Christaller in the 1930s (Christaller 1966), and picked up by the quantitative geography movement in the 1950s/1960s (Berry & Garrison 1958*b*,*a*), has been applied to within metropolitan areas. In Christaller's formulation, central places will be hierarchically organised: larger centres will be the most central in relation to smaller centres organised around them. If a size distribution of these are plotted in a rank size plot (centre size versus rank) then one would see a staircase like structure, with the largest centres having the smallest ranks, and vice-versa. In contrast, a Zipf-like distribution would show a power law hierarchy: smaller centre sizes are some relative proportion of the largest centre, and their sizes progressively go down, producing a more or less continuous distribution. Intriguingly, urban

structure seems to show combinations of both forms. For instance Johnston (Johnston 1966), studying Melbourne in 1966 finds that "although a city-wide hierarchical structure exists ..., within the city wide variations occur in the relative proportions of each order of centre present." So the key issue is how regular does the hierarchy need to be before we judge the hierarchy to be a continuum? The continuum suggests a distribution like Zipf's may be more appropriate for the feature-full topography than the staircase Christaller argued for on a featureless plain. But we also need to consider dynamics (Allen & Sanglier 1979), as technologies change and cities grow, equilibration may be in the process of occurring, but not yet have occurred.

This paper postulates that the theoretically possible range of relationships between the equilibrium-approaching Zipf distribution versus the at-equilibrium Christaller distribution. In particular, we show that although rank-size plots of centres show continuous Zipf or log-normal type distributions, using flows in the system as a threshold and studying percolation of connected clusters at various flows, the hierarchical distribution of centres proposed in the Christaller distribution may be revealed. Thus, a city can be seen to be a Zipf-like system that is always approaching a Christaller-like equilibrium, but is constantly facing reorganisations of land use - transport relationships and novel technologies.

Paper summary

The paper presents three metrics for measuring polycentricity all of which consider the flow/spatial interaction aspects, treating the origin-destination (OD) matrix as a network. Then, areas are ranked as centres based on these three measures, and their rank-size distribution is studied. The approach is demonstrated on journey-to-work (JTW) data for the Greater Sydney metropolitan region. Comparisons are made with the employment density measure. Finally, due to the continuity of centre ranks observed in the rank-size distributions, a percolation-based thresholding framework based on flows is presented. In particular, the full OD matrix is thresholded at different flows: centres emerge hierarchically. We postulate that the higher the position of a centre in the system, the longer the range of trip flows and higher the range of trip volumes over which it stays connected to the system. Lower order centres will both get disconnected faster, and also support a lower volume of flows. The approach provides a heuristic to identify Christaller-like sub-centres in a hierarchy. This heuristic is used to count the number of hierarchically organised

centres. For the specific case of Sydney, it is seen that beyond the Level 1 centre (the traditional Central Business District (CBD)) and 5 other Level 2 centres at the next hierarchical level, the distribution quickly turns into a continuous one, suggesting that beyond these 5-6 centres (covering about 25% of the employed workers, with Sydney CBD accounting for 15%), most of the employment in Sydney is dispersed rather than polycentric. Though in this paper we work with the latest 2016 Australian Census data, these findings corroborate with what was reported from the earlier census periods (Pfister et al. 2000), which suggests that despite planning authorities pushing for polycentric urban development, most of the on-ground employment and JTW patterns are divided between a few top-level centres and a large number of dispersed locations.

Methods

A graph $G(V, E)$ is defined, where the vertex set V represents the smallest areas at which flow data is available following a consistent definition of *origins* and *destinations* (e.g. Statistical Areas Level 2 (SA2) for Australia, Census Tracts for the US, etc.), and the edge set E represents the flows of people entering and exiting the nodes. Each node in the graph is an origin as well as a destination. The graph G is a full graph, in the sense, that potentially all possible edges exist. The graph G is represented by its adjacency matrix T, where T_{ij} represents the count of the number of people going from origin i to destination j. In transport terminology, T is the *trip distribution origin-destination flow matrix.* Thus, G is a weighted, asymmetric graph, since usually the number of people going from i to j differs from the number of people going from j to i.

Figure 1 shows three chord diagrams at a coarse spatial scale demonstrating the basic structure of such a graph. A node i should only be defined as a *centre relative to another node* j if the *net flow* into it is higher as compared with another node j; that is, the flow from j to i is higher than the flow from i to j. Thus, in Figure 1, analysing the nature of a particular node i (e.g., Sydney CBD, Parramatta, and the Eastern Suburbs), a link for this node i is only colored orange when the flow from another node j into i is higher than the flow from i into j. Using this logic, it is easy to observe visually that Sydney CBD is a global centre for the metropolitan region, Parramatta is a local centre for the western and southern suburbs, while the Eastern Suburbs are not a centre

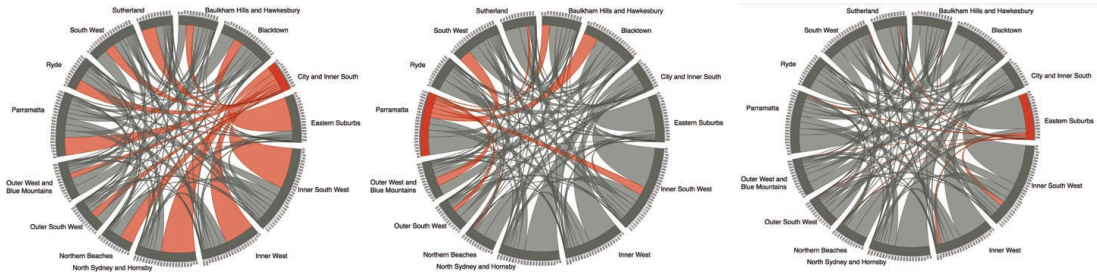

Figure 1: Journey-to-work data from 2016 Census, Australia. Flows entering and exiting three locations in Sydney are shown: Sydney CBD (left), Parramatta (middle), and the Eastern Suburbs (right). The length of the circle arc represents the number of people resident in an area. The width of the link at the base shows the number of people going from i to j. A link is colored orange for a focus node i (with the resident base also in orange) when the number of people flowing into i from j is higher than the number of people flowing into j from i: in this case, i is a centre relative to j.

(at this coarse area definition).

Trip-based Centricity Index

In order to formally capture this notion of the relative centre based on network flows and spatial interaction between locations, we propose the following measure of 'centricity'. Each node (k) has a 'trip-based centricity index' (C_t), which is defined as follows:

$$C_{t,k} = \frac{T_{D,k} - T_{O,k}}{\sum_{i=1}^{I} \sum_{j=1}^{I} T_{ij}}, \tag{1}$$

where, $T_{O,i}$ is the number of people originating in i,

$$T_{O,i} = \sum_{j=1}^{J} T_{ij}, \tag{2}$$

and $T_{D,j}$ is the number of people with destination j,

$$T_{D,j} = \sum_{i=1}^{I} T_{ij}. \tag{3}$$

We now let T represent the total number of employed people in the system,

$$T = \sum_{i=1}^{I} \sum_{j=1}^{I} T_{ij}, \tag{4}$$

getting

$$C_{t,k} = \frac{T_{D,k} - T_{O,k}}{T}. \tag{5}$$

If all T people travel to one particular destination zone, then $C_{t,k} = 1$ for this zone, and negative for all other zones. If all T people travel from one particular origin zone, then $C_{t,k} = -1$ for this zone, and positive for all other zones. Overall, the value of $C_{t,k}$ will vary from -1 to 1, and the more positive values of $C_{t,k}$ imply more of 'destination-ness' associated with a zone, thereby marking a hierarchy of 'centres' or 'sinks' in the system. In contrast, the more negative values of $C_{t,k}$ imply more of 'origin-ness' associated with a zone, thereby designating a hierarchy of 'sources' in the system.

While this measure scales from -1 to +1, we would like to enable comparisons between cities, as well as rankings of centres over time in the same city. Thus, for the purpose of comparisons (also with other centricity measures proposed later), we perform a standard normalization as follows yielding values between 0 and 1:

$$C_{T,k} = \frac{C_{t,k} - min(C_t)}{max(C_t) - min(C_t)}. \tag{6}$$

Density-based Centricity Index

Instead of dividing by the total number of employed persons or the total number of trips in the system (which are a link property), the centricity index can also be defined by the employment density at a location (which is a node property). The basic idea would then be that a location is a centre when the *net flow density* into it is high. Each node has a 'density-based centricity index' (C_d), which is defined as follows:

$$C_{d,k} = \frac{T_{D,k} - T_{O,k}}{r_k}, \tag{7}$$

where, as before, $T_{O,k}$ is the number of people originating in k, and $T_{D,k}$ is the number of

people with destination k. We let r_k represent the total area of the location k. Then, $C_{d,k}$ is a net flow density-based centricity measure. The values of $C_{d,k}$ will vary from negative to positive, and as before, to enable comparisons between cities, and to enable ranking change comparisons for the same city over time, we normalize them to lie between 0 and 1:

$$C_{D,k} = \frac{C_{d,k} - min(C_d)}{max(C_d) - min(C_d)}. \tag{8}$$

Accessibility-based Centricity Index

Metropolitan regions have defined labour markets, where a labour market implies the spatial spread that is taken into consideration living and daily commutes to work for labour. As cities grow in spatial size, with only a few centres of employment and the older centres being the largest in the hierarchy, the distance to jobs will, in general, grow. Here we define an accessibility-based centricity (C_a) index by tying the idea of polycentricity to the idea of accessibility.

We count the total number of jobs available within a particular time/cost threshold from a location k and call this Employment Accessibility $A_{e,k}$. In data terms, this is computed as the sum of the column sums of the JTW matrix, filtered by the time/cost threshold function c_{ik}, which is the cost of travel distance, time or money cost between locations i and k, and $f(c_{ik})$ representing a function of this cost.

$$A_{e,k} = \sum_{i=1}^{I} T_{D,i} f(c_{ik}). \tag{9}$$

We now count the total number of workers (labour) available within a particular time/cost threshold to a location k and call this Labour Accessibility $A_{l,k}$. In data terms, this is computed as the sum of the row sums of the JTW matrix, filtered by the time/cost threshold:

$$A_{l,k} = \sum_{i=1}^{I} T_{O,i} f(c_{ik}), \tag{10}$$

We define the cumulative opportunity threshold as follows (Wickstrom 1971, Wachs & Kumagai 1973, Ingram 1971):

$$f(c_{ik}) = \begin{cases} 1 & if\ c_{ik} \leq W_c \\ 0 & if\ c_{ik} > W_c \end{cases} \tag{11}$$

If the Employment Accessibility at location k is higher than Labour Accessibility from the location k, then this location k has positive Accessibility Centricity, implying that more jobs can be accessed from this location than workers, given a particular time/cost threshold. Thus, Accessibility Centricity is defined as:

$$C_{a,k} = \frac{A_{e,k} - A_{l,k}}{T}, \tag{12}$$

For example, setting the threshold (W_c) to 30 minutes of travel time, $f(c_{ik}) = 1$, when the travel time between i and k is within or up to 30 minutes, and 0 otherwise. Thus, $A_{l,k}$ will give us the number of workers to k who can access k from within 30 minutes. Similarly, $A_{e,k}$ will give us the number of jobs at all locations that are accessible within 30 minutes from k.

The accessibility-based centricity measure can range from negative to positive, and we normalize and scale it from 0 to 1 using:

$$C_{A,k} = \frac{C_{a,k} - min(C_a)}{max(C_a) - min(C_a)}. \tag{13}$$

The time/cost threshold can now be parametrically varied (e.g., 15, 30, 45, 60 minutes, etc.) and accessibility-based centricity can be studied for different time-thresholds.

Percolation Analysis on the Full O-D Flow Matrix to count sub-centres

The three proposed measures above are computed for SA2 regions in Sydney and compared against the traditional employment density measures. It is found that they show a continuous rank-size distribution (unimodal for trip and density centricities, bimodal for accessibility centricity). Thus, the issue of counting sub-centres again rests on deciding a particular cut-off threshold that we discuss in the introduction as suffering from arbitrariness: for all locations showing positive centricity by any of the above proposed measures, at what rank or level does a location become a centre as opposed to a non-centre?

To tackle this issue, we propose a percolation-based framework. Starting with the full O-D

matrix of flows, we first produce a symmetric version that counts all the in-flows as well as the out-flows occurring from a location:

$$T = T + T'. \tag{14}$$

Then, extract all the unique flows, i.e., we put all the entries T_{ij} into a single vector, arrange them in descending order, and then extract the unique flow values. Then, we use each of these unique flows as thresholds to produce a series of O-D matrices, progressively thresholded at each unique flow. Thus, the flows act as the percolation paramaters, above which flow is permitted in the system of nodes (with possible all-to-all connectivity) and below which flow is not permitted. If there are $K = 1 \ldots k$ unique flows f_k, then we have a series of matrices:

$$T_{k,(i,j)} = \begin{cases} T_{ij} & if\ T_{ij} > f_k \\ 0 & otherwise \end{cases} \tag{15}$$

Thus, the row or column sums of these thresholded O-D matrices T_k will give us at each flow threshold f_k the total number of trips associated with a particular location (node) in the system. A binary connectivity version of this matrix can be produced, following (Batty 2013), where

$$T_{k,(i,j)} = \begin{cases} 1 & if\ T_{ij} > f_k \\ 0 & otherwise \end{cases} \tag{16}$$

Thus, in the above matrix, if a particular T_{ij} falls under the threshold f_k, it will turn into a 0. At the f_k where a full row or column of zeros is produced, it implies that the node is disconnected from the system at this flow threshold. Thus, the row or column sums of the thresholded O-D matrices T_k will give us, at each flow threshold f_k, the number of nodes that are connected or disconnected.

When the number of connected nodes (from Eqn 16) is plotted against flows, the results show that there is critical threshold before which almost all nodes are connected, and after which there is rapid disconnection of nodes. Further, if the total number of trips per location is plotted (computed from Eqn 15 against flows, we see that sub-centres emerge in a clear hierarchical structure, where for higher order centres high volume trips are maintained for a much longer span of flow volumes, and the fall offs are slower. Thus, a heuristic is provided to count the number of

centres emerging hierarchically at different thresholds.

We note that the ordering produced by this process differs from the ordering produced if the locations were simply ranked by their gross in-flows or out-flows. Instead, the analysis proposed here reveals both (a) which locations remain connected for the longest span of trip flows in the system, and (b) which locations support the highest trip volumes at each flow threshold. Put together, this helps us identify the centres. A location can have overall low flows, but be connected to very few other nodes via very large flows. In contrast, a location can have overall high flow volumes, but get disconnected faster. Thus, the threshold flow acts like a parametric 'knob' specific to a region, but generated automatically by the underlying data, that helps to differentiate between the hierarchy of centres and also decides whether a region is highly centralised (mono or poly) or whether is a combination of central place(s) and dispersed lower order centres.

Data

Journey-to-work data from the 2016 Australian Census is acquired for the Greater Sydney metropolitan region. The Australian Bureau of Statistics (ABS) defines a statistical geography for the whole of Australia. Statistical Areas Level 2 (SA2s) are the smallest areas for which journey-to-work data are available with a consistent definition of origins and destinations. A finer level of data is available for smaller areas, namely the number of workers residing in Statistical Areas Level 1 (SA1s) and the number of workers working in Destination Zones (DZs). However, this smaller geography cannot be used, since the SA1s and the TDZs overlap spatially, and no correspondence map is available from the ABS that allows the re-mapping of either data set to the other.

To compute the accessibility, a full travel time matrix was created for the 282 SA2s in the greater Sydney metropolitan region by querying the Google Distance Matrix API. The queries were based on a typical weekday profile (Wednesday) and a typical time profile (8 am peak hour) of travel. Transit travel time incorporates all stages of travel, including walk access to, and egress from, stations, in-vehicle time, and transfer time, and is replaced by walking time to the destination when that is shorter. Automobile travel time estimates are based on historical traffic conditions.

Results

The proposed measures are demonstrated and evaluated on JTW-data for the Greater Sydney metropolitan region. The measures are used to determine the hierarchical organization of centres, and Christaller's Central Place Theory and Zipf formulations are tested empirically. Finally, the number of centres are counted by using the percolation framework, and the ranks from centricities and counts from the percolation framework are compared.

Trip and density-based measures of polycentricity

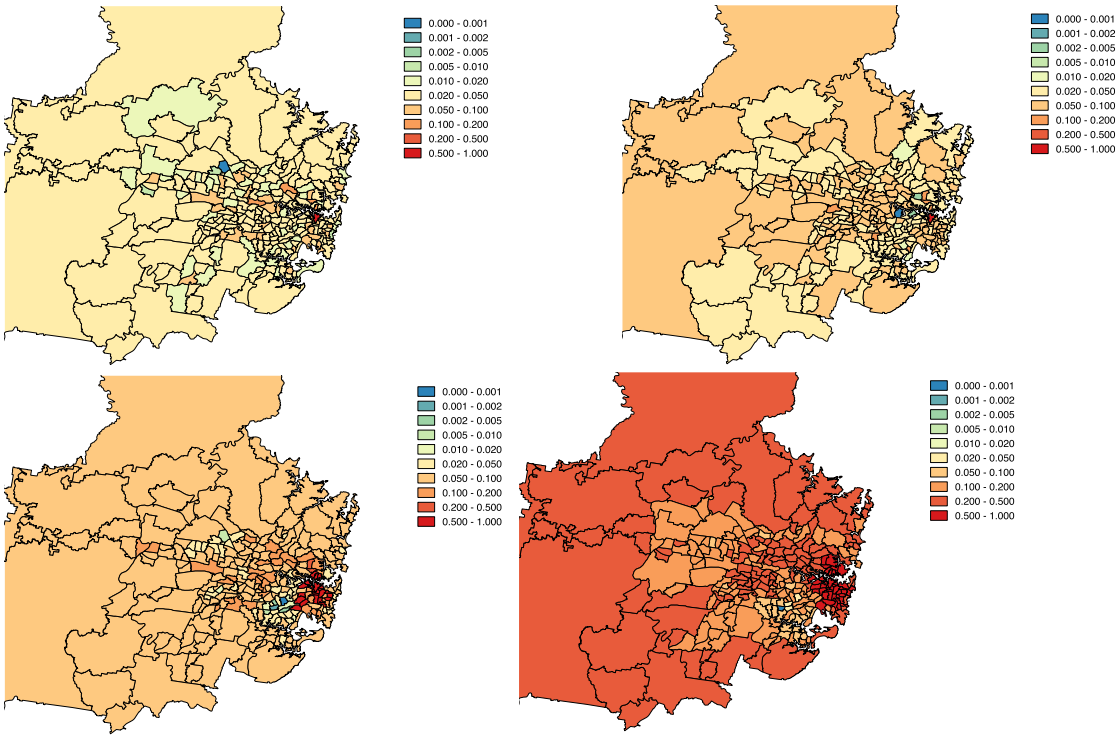

Figure 2: Trip (top-left), Density (top-right), Transit Accessibility (bottom-left) and Auto Accessibility (bottom-right) Centricity Index

Figure 2 demonstrates the centres identified through the Trip (top-left) and Density-based (top-right) Centricities $C_{T,k}$ and $C_{D,k}$, respectively. For both measures, Sydney Central Business District (SA2 Sydney-Haymarket-The Rocks) emerges as a clear outlier: it lies far removed from the bulk distribution. The other centres that are identified by the Trip Centricity metric are close to the CBD, but Parramatta - Rosehill is more towards the west, and is currently proposed as a

'second city' in the 'three cities' polycentric plan proposed by the Greater Sydney Commission (GSC) (Greater Sydney Commission 2018). Parramatta-Rosehill only appears as a second order centre, even though it is proposed as one of the primary centres in the GSC plans. For Parramatta to become one of the primary centres similar to the Sydney CBD, it has to attract a much higher proportion of trips, raising its trip-based and density-based centricities. On the other hand, some smaller centres (SA2s) surrounding the Sydney CBD, such as Lavender Bay and North Sydney show much higher trip and density-based centricities than Parramatta. Therefore, the asymmetric spatial position and the dominant role of the Sydney CBD retains (it certainly is much larger than twice the second largest centre), with other centres weakly arranged into a more continuous distribution (Figure 3).

Accessibility-based measure of polycentricity

Figure 2 shows the accessibility-based centricities (bottom-left: transit, bottom-right: auto) computed at the SA2 level for the Greater Sydney region. Sydney CBD and its surrounding SA2s clearly show very high accessibility-based centricity as compared to any other SA2. No other part of the entire Greater Sydney region shows accessibility-based centricity that comes close to the CBD and its surrounding locations. This reiterates the still leading, dominant and monocentric role of the Sydney CBD in the metropolitan region. Figure 2 also shows another signature: the SA2s that comprise the outer Western and outer Southern suburbs show lower accessibility-based centricity, like a middle ring band separating the inner city areas and the absolute outer fringe.

The policy takeaway is that if polycentricity is to be truly realised in this region, as proposed by the GSC, then accessibility to employment from SA2 areas in the western and southern suburbs must be raised. The analysis shows that at the moment, even with Parramatta and Liverpool proposed as centres, and even with Parramatta functioning as a second-order centre, the western and southern suburbs continue to act as residential sources for the entire region. In particular, the job-housing imbalance between where jobs are located and where people live may actually be exacerbated since very high numbers and densities of residential dwellings are currently proposed for these western and southern areas. In contrast, in the eastern and northern suburbs, that do have significantly higher accessibility centricity and are much closer to the Sydney CBD, proposed residential numbers and densities are relatively lower when compared to the west and

south. However, the results imply that raising only residential densities without raising their job accessibility-based centricity may actually exacerbate the current spatial imbalances and the current inefficiencies further; suggesting accessibility to employment should be explicitly considered as a framework for implementing the normative goal of polycentricity.

Rank Size Distributions: Zipf and Christaller

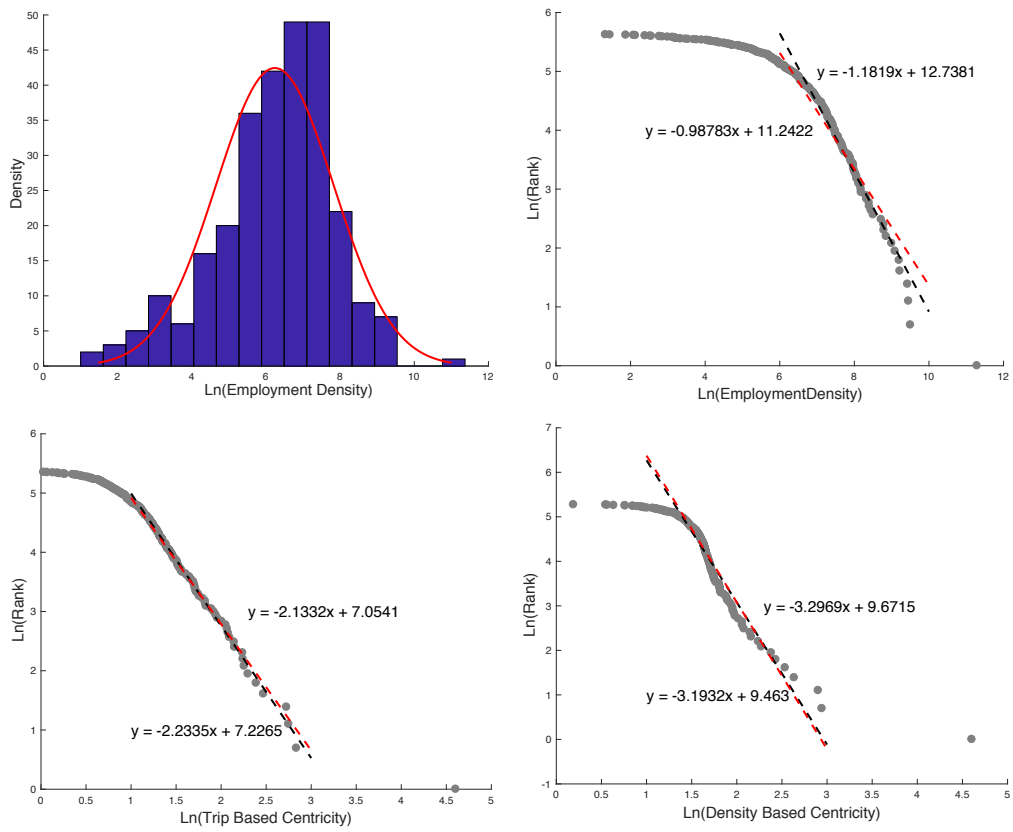

Figure 3: Employment Density, Trip and Density-based Centricities: Rank Size Distribution Analysis. Red line shows fits for the top 200 centres. Black line shows fit for the top 150 centres.

The spatial organization of social and economic activity in space generates regularities of structure. Cities, for example, are said to follow Zipf's law, where the probability that the size of the population is greater than some S is proportional to $1/S$. We test the same idea for employment centres: the probability that the centricity (as a measure of size) of a centre is greater than some C is proportional to $1/C$,

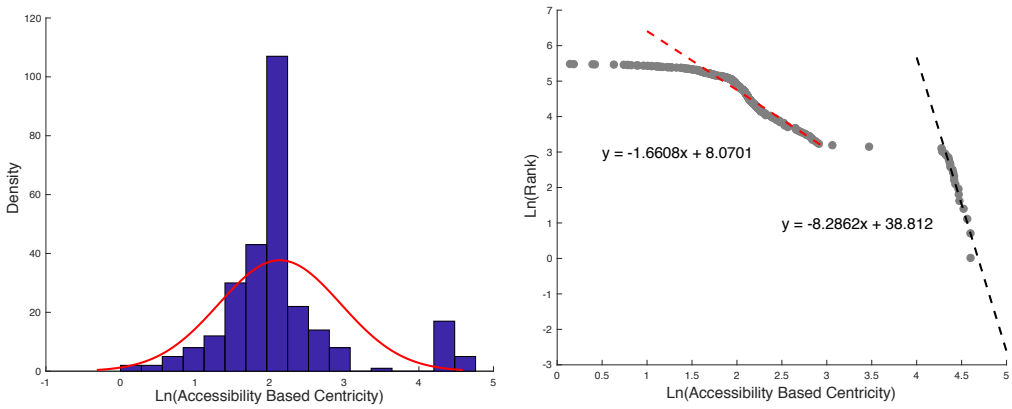

Figure 4: Transit Accessibility-based Centricity: Rank Size Distribution Analysis

$$P(Centricity > C) = \frac{a}{C^{\alpha}}. \tag{17}$$

When the value of the exponent α is equal to 1, the distribution is said to follow Zipf's Law. In the case of cities, especially in the US, the value of α comes almost exactly to 1, leading to the idea that city size distributions follow Zipf's law (Gabaix 1999). The largets cities are ranked in descending order by their populations and log(populations) are plotted against log(ranks), with the result that the slope of the distribution is -1. However, it was also shown that when cities of all sizes across the spectrum were considered, including smaller and medium sized cities, the distribution is log normal (Eeckhout 2004).

Similar distributions of trip, density and accessibility centricities are shown in Figures 3 and 4. In each case, the SA2s have been ranked by their centricities, with the largest centricity locations receiving the lowest rank.

Employment density as well as the centricities appear as log-normally distributed (at least at this spatial scale)(Figures 3, top left, and 4, left). One could argue that not all the SA2 areas are centres, and so only the distributions of centres should be measured. However, as a map of employment densities and trip and density-based centricities show, at least in the case of Sydney, apart from Sydney CBD and the surrounding area, there is no other dominant centre of the same order: the primate CBD dominates. However, the proportion of the total number of workers with Sydney - Haymarket - The Rocks as a destination is only about 15% (2016 census data).

Most employment in the region is actually dispersed, rather than agglomerated in other centres. Parramatta, and the areas of Strathfield and Homebush, may appear as second order centres that are spatially distinct from the Sydney CBD, but these emerge as having lower centricities than the SA2s that surround Sydney CBD. Thus, in terms of centre identification, a set of contiguous SA2s around the CBD (Figures 2[bottom-left] and [bottom-right]) dominate, with the rest of the employment dispersed. Thus, we plot the centricities of all the SA2s. Moreover, with the centricities computed as they are, the centres emerge from the bottom up, instead of being pre-defined from the top down (which should be a necessity for any rigorous definition of a centre).

Figure 3 shows the rank size distributions for employment density (top right), trip-based centricity (bottom left), and density-based centricity (bottom right). Curve fitting routines in Matlab were used to compute the straight-line fits for the upper tails. This required the choice of a cut-off point. A systematic range of cut-off points were chosen and the fit performed. It can be seen that the more centres chosen, the lower the slope of the line, and choosing a lower number of centres raises the slope. Further, the analysis was performed by first including and then leaving out the Sydney - Haymarket - The Rocks SA2, which is a clear outlier in all cases. Because this SA2 is a clear outlier, it has the capacity to change the resulting slopes of the lines. The results shown visually are for when it was left out of the analysis, and the fit was computed for the second centre up.

Employment density exponent values are close to Zipf ($\alpha = 0.98$ for the top 200 centres, $\alpha = 1.19$ for the top 150 centres). However, the exponent values are much higher for the trip and density-based measures: $\alpha = 2.1 - 2.2$ for trip-based and $\alpha = 3.2 - 3.3$ for density-based. This shows clearly that when net in-flow measures are considered as opposed to density-based measures, the fall-off of centricities is higher.

Figure 4 shows the density plot (left) and the rank size distribution (right) for accessibility-based centricity, leading to a surprising and unexpected result: accessibility centricity appears bi-modal, with one set of centres around Sydney CBD followed by a large gap in the distribution, followed by all the other centres. The slope of the line is very high for the first set of centres ($\alpha = 8.2$), establishing the primacy of the CBD. The slope of the line for the rest of the distribution is much lower ($\alpha = 1.7$). This plot also points to a connection between the Christaller distribution idea of centres being organised like stairs with gaps between hierarchies) versus power-law or Zipf

like distributions, which are more continuous.

One conjecture the results point to is that there are Christaller-like hierarchies for centres, within which there are more continuous Zipf-like distributions. Of interest would be the underlying processes that could lead to such bi- or multi-modal distributions. If we think of Christaller-like hierarchies as the resulting equilibrium distribution of centres in a static environment, but a dynamic economy where technology and organizational changes continuously disrupt that equilibrium, perhaps the Zipf characterizes the ever shifting dynamics, allowing centres to lie more or less on a more continuous size distribution. Since transport and spatial technologies and economic activity organization are changing, the equilibrium is not reached, resulting in more continuous Zipf-like forms. Nonetheless, these findings and related conjectures need more empirical testing for different cities, and across different geography definitions (e.g. Census Tracts in the US, Output Area hierarchies for the UK, etc.)

Number of centres via percolation analysis

Figure 5 shows the results of the percolation analysis. The top two parts show that there is a definite critical threshold of flows: above this threshold all the nodes are connected, below this threshold, there is rapid disconnection of nodes. The top centres will be those that support highest flows and remain connected to the very end. The bottom two parts show two snapshots, identifying hierarchically what may be called Level 1 and Level 2 centres. With total number of trips at each flow threshold plotted for each location, Sydney-Haymarket-TheRocks emerges as the top level centre, completely separated from the others (Figure 5, bottom-left). We then remove this centre and replot the same graph. At the next hierarchical level, we now see 5 more clear centres emerge, that are separated from those below (Figure 5, bottom-right). These are, in order: Parramatta-Rosehill, Macquarie Park-Marsfield, North Sydney-Lavender Bay, Pyrmont-Ultimo, Chatswood (East)-Artarmon. 2 of these are adjacent to Sydney-Haymarket-The Rocks (North Sydney-Lavender Bay, Pyrmont-Ultimo), with Parramatta-Rosehill the farthest, and with Macquarie Park-Marsfield and Chatswood (East)-Artarmon at medium level distance from Sydney-Haymarket-The Rocks. If this process were repeatedly performed, e.g. removing these 5 centres and replotting the graph, the next level of hierarchy emerges. But, as the hierarchical levels increase, the distribution again approaches continuity, as the gaps between the higher and lower

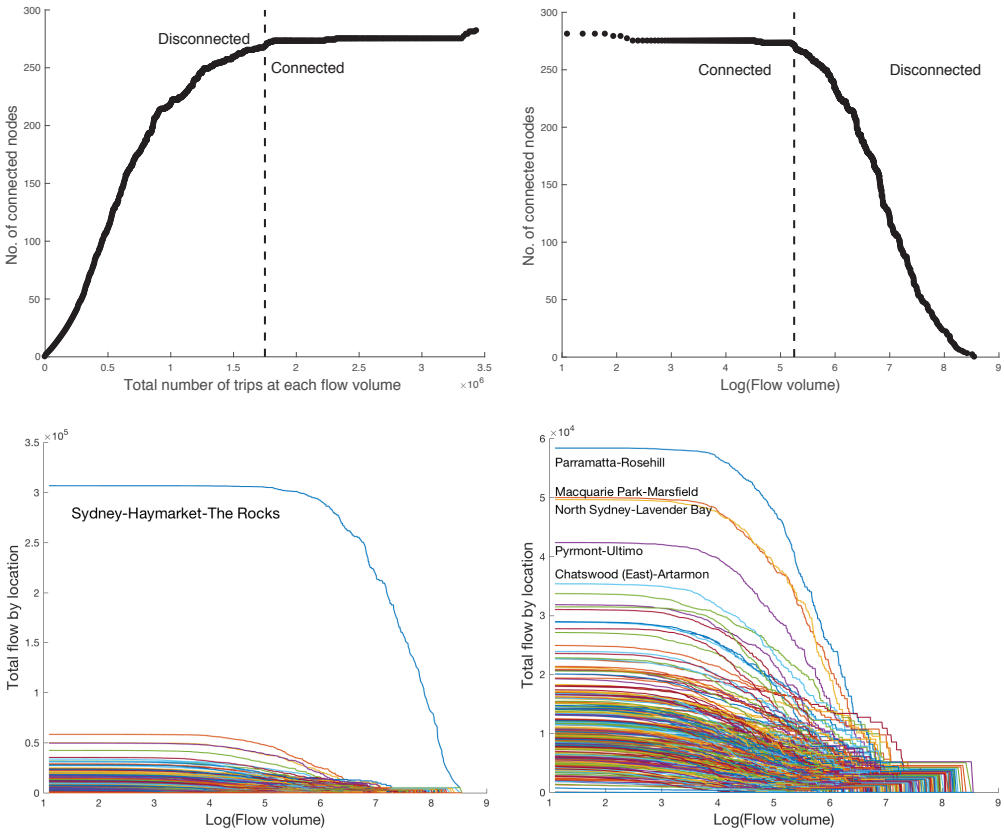

Figure 5: Identifying centres via percolation and flows: (a) Top-left: Total trips at each flow against number of connected SA2s, (b) Top-right: Flows against number of connected SA2s, (c) Bottom-left: Flows by total number of trips at each location, All SA2s, (d) Bottom-right: Flows by total number of trips at each location, All SA2s but top one, Sydney-Haymarket-The Rocks.

centres begins to merge.

Findings and Comparative Analysis

Table 1 shows the top 35 centres identified by flows. They are in descending order by the total number of trips they account for at the flow thresholds discussed above. In parallel, %-employed workers, employment density, centricity-based ranks, and the hierarchical level identified through the percolation analysis are also shown. We discuss a few of the important findings of our analysis.

Absolute versus relative centres. While employment density and %-employed workers per area measure absolute centricity, the net in-flow-based metrics measure relative centricity. An absolute centre attracts overall high numbers of workers, but does not lose a significant number of workers to other locations. Sydney-Haymarket-The Rocks and Parramatta-Rosehill are absolute centres, even though Parramatta-Rosehill is much smaller by any measure of size. On the other hand, a relative centre attracts a large number of workers, but also loses workers to other locations. Examples are Chatswood (East) - Artarmon, Baulkham Hills - Bella Vista, Erskineville - Alexandria, or Concord West - North Strathfield, that have relatively high % of employed workers but slide down the trip centricity rank showing that a significant proportion of workers residing in these locations go to work at other places. Thus, some of these locations become centres at a lower hierarchical level (in the Christaller sense) if they are only relative rather than absolute, and one policy response could be to look at options of increasing their absolute centre-ness by increasing the number of jobs available at these locations, if doing so can reduce overall trip inefficiencies in the system.

Different types of flow polycentricity. Some identified centres are support high flows over a long range of thresholds, but get disconnected from the system at the very highest flows. In contrast, some support lower flows over a long range of flow thresholds, but remain connected to the end. Parramatta-Rosehill is an example of the former, where the number of trips is constantly high, but it gets diconnected faster than say Pyrmont-Ultimo that constantly shows a lower number of trips, but remains connected almost to the very end. This observation differentiates between highly connected versus high volume centres.

Not a polycentric Sydney yet. Apart from the top 5-6 centres, no other dominant centres are identified. Sydney-Haymarket-TheRocks is clearly the top level centre by all measures, but

	% of employed workers	ED Rank	TBC Rank	DBC Rank	ABC_{Auto} Rank	$ABC_{Transit}$ Rank	Percolation Analysis Sub-centre level
Sydney-Haymarket-The Rocks	15.24	1	1	1	15	4	Level 1
Parramatta-Rosehill	2.39	45	4	18	174	34	Level 2
Macquarie Park-Marsfield	2.30	244	3	68	79	28	Level 2
North Sydney-Lavender Bay	2.33	3	2	2	31	2	Level 2
Pyrmont-Ultimo	1.75	202	5	58	21	10	Level 2
Chatswood (East) - Artarmon	1.31	6	13	6	27	24	Level 2
Surry Hills	1.42	24	7	14	20	3	Level 3
Freshwater-Brookvale	0.77	17	34	22	157	201	Level 3
Newtown-Camperdown-Darlington	0.98	145	26	57	30	26	Level 3
Penrith	0.96	49	11	20	164	59	Level 3
Cambelltown-Woodbine	0.87	14	18	11	213	77	Level 3
St Leonards-Naremburn	1.40	2	6	3	12	23	Level 3
Waterloo-Beaconsfield	0.69	15	137	196	25	7	Level 3
Baulkham Hills (West)-Bella Vista	1.12	30	14	16	122	67	Level 3
Erskineville-Alexandria	1.12	116	16	38	32	27	Level 3
Mascot-Eastlakes	1.06	20	20	17	16	96	Level 3
Blacktown (East)-Kings Park	0.96	43	15	21	204	253	Level 3
Warriewood-Mona Vale	0.61	31	60	46	151	79	Level 3
French's Forest-Belrose	0.67	36	51	40	173	172	Level 3
Redfern-Chippendale	0.69	54	45	44	23	1	Level 3
Dee Why-North Curl Curl	0.27	112	269	242	149	233	Level 3
Concord West-North Strathfield	0.94	7	23	9	108	226	Level 3
Manly-Fairlight	0.41	11	164	273	159	129	Level 3
Potts Point-Woolloomooloo	0.71	65	46	47	8	8	Level 3
Mount Annan-Currans Hill	0.45	74	229	208	202	197	Level 3
Lane Cove-Greewich	0.59	10	109	198	49	135	Level 3
Homebush Bay-Silverwater	1.26	23	10	12	210	40	Level 3
Marrickville	0.63	25	87	121	184	22	Level 3
Bondi Junction - Waverly	0.73	44	25	26	6	20	Level 3
Leichhardt-Annandale	0.50	18	174	247	26	128	Level 3
Mosman	0.42	99	220	179	7	236	Level 3
Crows Nest-Waverton	0.42	55	136	171	28	5	Level 3
Sutherland-Kirrawee	0.50	41	92	128	261	189	Level 3
Double Bay-Bellevue Hill	0.45	40	160	199	11	162	Level 3
Liverpool	0.93	47	17	23	214	73	Level 3

Table 1: Comparative findings from centricity and percolation analysis

accounts for only 15.25% of the employed workers in the metropolitan region. The next 5 centres together with the CBD account for about 25% of the employed workers. Thus, most of the employment is dispersed across a range of smaller centres, the largest of which are shown in Table 1, but have disagreements between trip, density and accessibility-based centricities: some are low-density centres, and some are low-accessibility centres.

Spatial mismatches between centricities and accessibility to jobs. Positive correlations exist between %-employed workers, trip centricity and the sub-centre level. But, significant disagreements exist between these and accessibility centricities, showing the spatial mismatches for commute lengths in the system. A sub-centre with high trip-centricity, employing a high percentage of workers, but relatively lower auto and transit-based accessibility centricity, implies that even though a significant percentage of the population comes to this location to work, the access of jobs from this location in the time threshold of 30 minutes is low. A policy response would be to increase the accessibility of jobs from this location, as it already serves as a centre. This situation is particularly clear for the case of Parramatta-Rosehill and Macquarie Park-Marsfield.

Conclusions and Future Work

We presented trip, density and accessibility flow measures to determine the centricity potentials of small areas comprising a city, extending traditional views on morphological and functional polycentricity, and proposing (a) that centre-identification should rest on explicit considerations of networks, flows, and spatial interactions, (b) a distinction between absolute and relative centres and resulting central place hierarchies, and (c) that the idea of accessibility of places should be tied into the measurement of polycentricity. Since accessibility encodes the desirability and ease of reaching centres and is a measurable spatial characteristic, the implementation of polycentricity as a normative goal can be informed by the pragmatic aim of making places more accessible. Planning for increased access to centres would naturally aid in building a polycentric city.

Polycentricity thus computed was used to measure size distributions to test whether the centres follow Zipf-like or Chirstaller-like distributions. A percolation theory framework was presented, where flows are used as a thresholding parameter to count the number of sub-centres. Flows prove to be an effective measure to count and hierarchically organise metropolitan area sub-centres.

An existing empirical limitation is the specific geography definition. While the approach presented is general in application, the Sydney SA2 geography is too coarse. Furthur, SA2 areas near the CBD are very small, and grow progressively larger near the fringe areas of the city. Thus, it is difficult to develop a reliable basis to say that a particular SA2 is a centre, and another one isn't. Larger and coarser area definitions will also likely suffer from the Modifiable Area Unit Problem (MAUP). Thus, in future work, the approach presented here will be separately applied to finer census tract level data when available.

Nonetheless, given the constraint that the SA2 is the smallest area definition for which JTW data for Sydney is available (and, from a pragmatic perspective, the smallest area definition at which planning bodies operate), the analysis provided a reliable framework to enable a comparison between monocentricity versus polycentricity for Sydney. It demonstrated that if Sydney is to move from dispersed to polycentric, the low accessibility-based centricity to second order centres must be raised if a truly polycentric Sydney is to emerge.

References

Allen, P. M. & Sanglier, M. (1979), 'A dynamic model of growth in a central place system', *Geographical Analysis* **11**(3), 256–272.

Anas, A., Arnott, R. & Small, K. A. (1997), 'Urban Spatial Structure', *University of California Transportation Center* **working paper**(UTUC-357).

Barthelemy, M. (2016), *The Structure and Dynamics of Cities: Urban Data Analysis and Theoretical Modelling*, Cambridge University Press.

Batty, M. (2013), *The New Science of Cities*, MIT Press, Massachussetts.

Berry, B. J. & Garrison, W. L. (1958*a*), 'A note on central place theory and the range of a good', *Economic Geography* **34**(4), 304–311.

Berry, B. J. & Garrison, W. L. (1958*b*), 'Recent developments of central place theory', *Papers in Regional Science* **4**(1), 107–120.

Christaller, W. (1966), *Central places in southern Germany*, Prentice-Hall.

Eeckhout, J. (2004), 'Gibrat's Law for (All) Cities', *American Economic Review* **94**(5), 1429–1451.

Fujita, M. & Ogawa, H. (1982), 'Multiple equilibria and structual transition of non-monocentric urban configurations', *Regional Science and Urban Economics* **12**, 161–196.

Gabaix, X. (1999), 'Zipf's Law for Cities: An explanation', *Quarterly Journal of Economics* **August 1999**(August 1999).

Giuliano, G. & Kenneth, S. (1991), 'Subcenters in the los angeles region', *Regional Science and Urban Economics* **21**(2), 163–182.

Greater Sydney Commission (2018), *A Metropolis of Three Cities*, Greater Sydney Commission.

Green, N. (2007), 'Functional polycentricity: A formal definition in terms of social network analysis', *Urban Studies* **44**(11), 2077–2103.

Hall, P. & Pain, K. (2006), *The Polycentric Metropolis: learning from mega-city regions in Europe*, Earthscan, London.

Ingram, D. R. (1971), 'The concept of accessibility: A search for an operational form', *Regional Studies* **5**(2), 101–107.

Johnston, R. J. (1966), 'The distribution of an intra-metropolitan central place hierarchy', *Geographical Research* **4**(1), 19–33.

Liu, X., Derudder, B. & Wu, K. (2016), 'Measuring polycentric urban development in china: An intercity transportation network perspective', *Regional Studies* **50**(8), 1302–1315.

Louail, T., Lenormand, M., Ros, O., Picornell, M., Herranz, R., Frias-Martinez, E., Ramasco, J. & Barthelemy, M. (2014), 'From mobile phone data to the spatial structure of cities', *Nature Scientific Reports* **4**, srep05276.

Louf, T. & Barthelemy, M. (2014), 'Modeling the polycentric transition of cities', *Physical Review Letters* **111**, 198702.

McMillen, D. (2001), 'Non-parametric employment subcenter identification', *Journal of Urban Economics* **50**(3), 448–473.

McMillen, D. & Stefani, S. (2001), 'The number of subcenters in large urban areas', *Journal of Urban Economics* **55**(3), 321–338.

Meijers, E. (2008), 'Measuring polycentricity and its promises', *European Planning Studies* **16**(9), 1313–1323.

Pfister, N., Freestone, R. & Murphy, P. (2000), 'Polycentricity or Dispersion? Changes in Center Employment in Metropolitan Sydney, 1981-1996', *Urban Geography* **21**(5), 428–442.

Roth, C., Kang, S., M., B. & Barthelemy, M. (2011), 'Structure of urban movements: polycentric activity and entangled hierarchical flows', *PLOS ONE* **6**(1), e15923.

Vasanen, A. (2012), 'Functional polycentricity: Examining metropolitan spatial structure through the connectivity of urban sub-centres', *Urban Studies* **49**(16), 3627–3644.

Wachs, M. & Kumagai, T. G. (1973), 'Physical accessibility as a social indicator', *Socio-economic Planning Science* **7**(5), 437–456.

Wickstrom, G. V. (1971), 'Defining balanced transportation-a question of opportunity', *Traffic Quarterly* **25**(3).