Street network analysis for understanding typology in cities: Case study on Sydney CBD and suburbs

Somwrita Sarkar[1]
[1]Design Lab, Faculty of Architecture, Design, and Planning, The University of Sydney

Abstract: Cities are networks of streets, places, people, systems, and their interactions. Network analysis has been extensively used in quantitative and economic geography, and transportation research. Methods such as the space syntax approach have also been used in urban design. More recent research in modeling cities as complex physical systems, and more specifically street networks as spatial networks, has shown that the network metaphor can provide global insight into city structure, the patterns of its growth, and how social and economic structures interact with spatial structures. Surprisingly, there exists almost no literature on mapping and analyzing the structure of Australian cities using the complex systems perspective. In this paper, we report on the network theoretic analysis street networks of Sydney CBD and suburbs. We generate primal graph representations of the fine scale structure of street networks, and use network metrics to measure and characterize the formal aspects of these networks. Through the results, we show that a surprising diversity of form exists even within what is understood as the same historical typology (say, grid layouts, or the garden city cul-de-sac). This study represents a first step in modeling an Australian city as a complex network. We relate the historical qualitative approaches of analysis of city structure to a network based quantitative approach, providing a mechanism for studying typology classification in city structure.

Introduction

The formal study of city structure and street networks has historically being amongst the richest areas of inquiry in urban design and planning. Traditional approaches have been largely qualitative, and have derived from historical and cultural studies (Kostof, 1991; Rossi, 1984). The understanding of formal “type” has been defined on the basis of design movements, specific stylistic contributions of designers and cultural or art specific eras, socio-political forces, and geographic factors. Other prevalent approaches have been based on quantitative modeling. From Christopher Alexander’s Pattern Language (Alexander, 1977) and comparison of tree and semi-lattice structures (Alexander, 1966), to the Space Syntax approach (Hillier and Hanson, 1984), to Batty’s work on modeling the fractal structure of cities (Batty and Longley, 1994; Batty, 2007), to more traditional approaches (Haggett and Chorley, 1969) in quantitative and economic geography, economics, and transportation research, all quantitative modeling approaches attempt to derive pattern based and mathematical first-principles based ways to characterize urban structure.

In this paper, we focus on a complex systems perspective. As Jane Jacobs pointed out in her seminal book (Jacobs, 1961), cities are problems of organized complexity. In recent years, tremendous progress has been made in understanding the structure and dynamics of a wide range of natural and artificial complex systems (Newman, 2010). Protein interaction networks in cells, the network of synapses and neurons in the brain, the structure of the Internet and the World Wide Web, metabolic and genetic networks, social networks, online social media networks, networks of collaboration between scientists, citation networks – these are all examples of a diverse range of complex systems that are now being understood using a single representational and analytical mechanism (Newman, 2010). All of these complex systems can be represented as networks, and recent research in physics has revealed that surprisingly diverse and different systems can share universal organizational properties. Cities and city systems, such as street networks, transportation networks, or infrastructure networks, are all spatially embedded networks (Barthelemy, 2011). Research in modeling cities as complex physical systems (Barthelemy, 2011; Porta et al., 2006, 2009) has shown that graph representations can provide global insight into city structure, the patterns of its growth, and how social and economic structures interact with spatial structures. Such understanding is important not only from a scientific perspective; it can also inform policy, design, and management of cities (Barthelemy, 2011; Porta et al., 2006, 2009).

Surprisingly, there exists almost no literature on mapping and analyzing the structure of Australian cities using the complex systems perspective. In this paper, we report on the network theoretic analysis of street networks of Sydney CBD and suburbs. We generate primal graph representations (Barthelemy, 2011; Porta et al., 2006, 2009) of the fine scale structure of street networks, and use network metrics to measure and characterize the formal aspects of these networks. Through the results, we show that a surprising amount of diversity and range of form exists even within what is understood as the same historical typology (say, grid-iron layouts, or the garden city cul-de-sac), and that this diversity can be captured successfully by network based metrics. This study represents a first step in modeling an Australian city as a complex network to relate the historical qualitative approaches of analysis of city structure to a network based quantitative approach, providing a mechanism for studying typology classification in city structure.

Network data and primal graph representation

How many defined typologies of street network types characterize a city's structure? Can this understanding be expressed by quantitative metrics that derive from studies of the deep structure of street networks, as opposed to simple historical or stylistic classifications? Historical typologies and classifications often serve a useful purpose in characterizing how a particular city or suburb was formed. For example, the history of urban design and planning in Sydney is well-documented in various historical and modern sources (Dictionary of Sydney, 2013). These sources show a discussion of typology largely focused around the grid layout versus the garden city like cul-de-sac layouts. However, such a view of typology is limited, as it is often not helpful in characterizing morphological changes that happen over time (either by centralized planning action or via decentralized socio-economic-political action), or in capturing the diverse range of structural properties that may characterize even street plans falling under the same type.

The aim of this paper is to present a framework of analyzing street networks at a fine-scale structure, identify principal structural features that characterize the plan, and come out with a basis to measure typology in a quantitative sense using graph theoretic analysis. In this preliminary attempt, we choose Sydney as a case study, and apply the framework to study the typical structure of the CBD and some of the city's suburbs. In this preliminary study, we chose suburbs based on the diversity of plan forms, in particular, choosing "perfect grid" type layouts as well as "perfect cul-de-sac" type layouts.

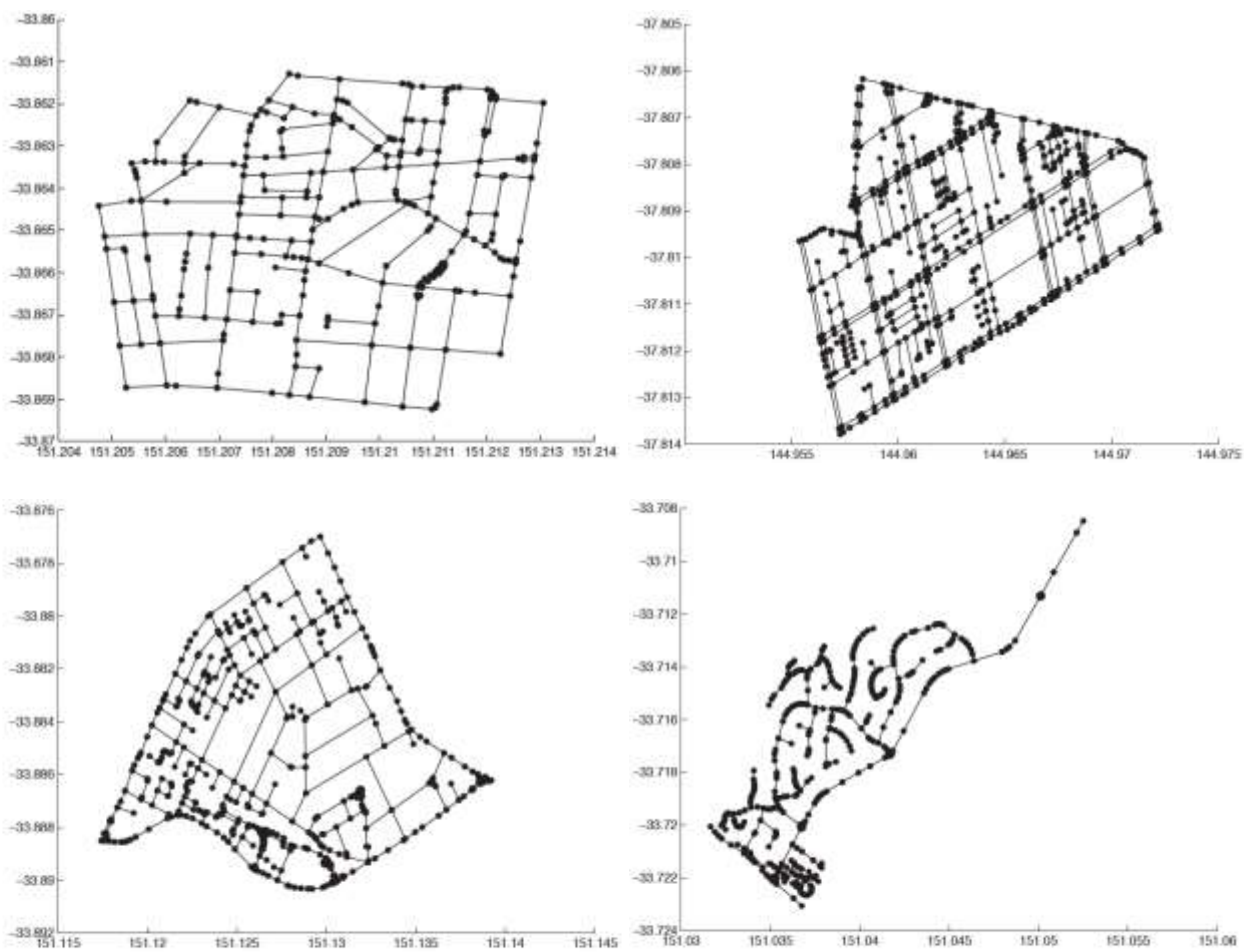
The street networks we consider here are essentially planar networks. Since we are interested essentially about the fine scale structure of local areas within the city, we do not consider non-planar features such as multi-level intersections. In formal terms, a street network is represented as a graph $G = (V, E)$, where the graph G is an abstract mathematical structure defined by the two sets V and E. The set V is a set of n *nodes* or *vertices*, and the set E is a set of m unordered pairs drawn from V, each pair representing a *link* or an *edge.* That is, the set V has n nodes, $1 \leq i \leq n$, and the set E has m edges, $1 \leq i \leq m$. This graph G is represented in matrix form by an *adjacency matrix* A, whose rows and columns represent the n nodes, and matrix entry $A_{ij} = 1$ if there is a link between nodes i and j, and $A_{ij} = 0$ otherwise. In a weighted graph, if a link exists between the nodes i and j the entry A_{ij} represents the length of the link between nodes i and j. That is, $A_{ij} = d_{ij}$, where d_{ij} is the metric distance between the nodes and measures the length of the link. Further, these street networks are undirected; i.e., the edges do not have a direction. Thus, the adjacency matrix A is symmetric, with $A_{ij} = A_{ji}$. This representation of a street network is referred to as the primal graph representation in research literature (Barthelemy, 2011; Porta et al., 2006, 2009), as opposed to the dual graph, where the streets are nodes, and share a link if they share an intersection.

We developed primal graph representations discussed above for Sydney CBD and several of Sydney's suburbs. We also generated a primal graph representation for Melbourne CBD, primarily for a preliminary comparison, and with the idea of extending this analysis framework to compare several Australian cities in future work. We note here that the primal graph representation is different from the way streets or ways are represented inside traditional Geographic Information Systems (GIS). In a GIS environment, a street, as a single entity, is defined as a series of nodes, whereas in the primal graph representation we use, the geometric and topological connectivity between every pair of nodes is defined as a separate link (even if several of them put together define a single street).

To generate the primal graphs, the Google Street Layer plugin in the QGIS software was used as the base for producing a digitized vector layer for the street networks. A standard road-centerline approach was used for digitizing. Each node, defined by a latitude-longitude pair, is either an intersection point where two or more streets meet, or it is a point where a road turns and changes direction. Each edge or link is a street portion or street, joining two nodes.

The digitized vector layer was exported from QGIS as a Keyhole Markup Language (KML) file, and imported into the matrix computation software Matlab. KML files contain information about each node and link, including their latitude and longitude positions. Inside Matlab, the KML file was used to extract the street network primal graph and its adjacency matrix. Figure 1 shows the primal graph representations for several representative CBD and suburban areas; The axes show actual latitude-longitude positions, and the networks show topology, as well as geometry of the street network structure.

Figure 1. Primal graph representations of Sydney CBD (left top), Melbourne CBD (right top), Ashfield (bottom left) and Cherrybrook (bottom right) using the KML file exported from the GIS environment into Matlab



Overall Network metrics and analysis results

We study the primal graphs generated for the following metrics: the *alpha*, *beta,* and *gamma* indices. These indices have been used in quantitative geography for a long time to study river networks, infrastructure networks, rail, road, or water based transportation networks (Haggett and Chorley, 1969). The most important characteristic of these indices are that they help to distinguish clearly between the tree-like and mesh-like (grid-like) structure of networks. As we will see, these metrics are specially important for characterizing dominant typologies that characterize suburbs, beyond the simplistic grid and mesh plan typologies versus the garden city type cul-de-sac plan forms.

The simplest index to measure the gross topological characteristics of the network is the ***beta index*** β, which is the ratio between the number of edges and the number of vertices,

$$\beta = \frac{m}{n}$$

In a fully connected graph with no cycles, i.e., a tree graph, the number of edges m is exactly 1 less than the number of nodes n, and therefore $\beta < 1$. When a graph contains cycles, i.e., lattice or mesh graphs, $m > n$, and $\beta > 1$. Thus, values of β less than 1 represent tree-like networks, whereas values of β more than 1 show the presence of cycles in the graph. The higher the value of β, the more the number of cycles or mesh-like / grid-like structure in the network.

The second measure we study is the ***cyclomatic number***

$$C = m - n + 1$$

This measures the number of cycles in the network. The cyclomatic number $C = 0$ when there are no cycles, because as noted above, for a fully connected acyclic graph, $n = m + 1$. If there are any edges over this number, then this shows the number of cycles in the network.

In our studies we found that the cul-de-sac type suburbs are extremely tree-like, with very few cycles. For example, Cherrybrook, with a network size of about 419 nodes has only 427 edges, with a $\beta = 1.01$ and $C = 9$. That is, the beta index is just above 1, and the number of cycles is just 9. In more grid-like or mesh-like layouts, the beta index and the cyclomatic number are much higher.

However, we also found that suburbs with an apparent grid-like layout, for example see the Ashfield network in Figure 1, have a lower β and C as compared to other grid layout areas (for example the CBD, Stanmore, etc.). The reason is that several areas with *apparent* grid or mesh-like structure at a coarse description level can still have tree-like structure embedded at a finer scale. This is a typological feature we observed in several suburbs – it is possible for tree-like structures to be embedded in a larger grid structure, leading to a mix of types. Table 1 shows the beta index and cyclomatic numbers for a selection of CBD and suburban areas discussed above.

The third index we study is the ***alpha index*** α. The alpha index is a measure of the cycle density, i.e., the actual number of cycles that exist in a network as compared to the maximum number that can exist. The street networks we have primarily considered are planar networks. For planar networks, the maximum number of cycles that can exist is known to be $2n - 5$. Thus, the alpha index is defined as

$$\alpha = \frac{C}{2n - 5}$$

Table 1 shows the alpha index for the various CBD and representative suburban areas. In our studies, the CBD areas shows the highest cycle density. It is interesting to note that Melbourne CBD shows higher α than Sydney CBD. Some suburbs that continue the strict grid layout, such as Stanmore, show comparably high α to the CBD areas. However, several others, such as Ashfield, show lower α, despite apparently following the same grid-like layout. The lowest α is recorded by tree-like suburbs, such as Cherrybrook.

Another feature of interest was that several suburbs, for example areas around the coast such as Bondi and Balgowlah, showed a mixed typology, with medium $\beta, C,$ and α. Typically, in these suburbs, there appeared to be a tree-like arrangement of streets around the coast line, and a more grid or mesh-like layout as one moves further inland. These border suburbs are particularly interesting, as they appear to lie on the border between the cul-de-sac type and grid-type, incorporating structural features from both types.

Table 1. Overall network metrics for representative CBD and suburb areas

Area	Nodes	Edges	Beta index	Cyclomatic number	Alpha index	Gamma index
Sydney City	257	329	1.2802	73	0.1434	0.0954
Melbourne City	535	698	1.3047	164	0.1540	0.1026
Ashfield	479	538	1.1232	60	0.0630	0.0419
Stanmore	115	147	1.2783	33	0.1467	0.0973
Cherrybrook	419	427	1.0191	9	0.0108	0.0072

The fourth index we study is the ***gamma index*** γ. The gamma index is a measure of the link density, i.e. the actual number of links that exist in a network as compared to the maximum number that can exist. The street networks we have primarily considered are planar networks. For planar networks, the maximum number of links that can exist is known to be $3n-6$. Thus, the gamma index is defined as

$$\gamma = \frac{m}{3n-6}$$

Table 1 shows the gamma index for the various CBD and representative suburban areas. Confirming the observations above, the lowest link density was recorded in tree-like suburbs, and higher link densities were recorded by mesh-like layouts, with the link density correlated positively with the number of cycles.

Shortest paths, centrality metrics and results

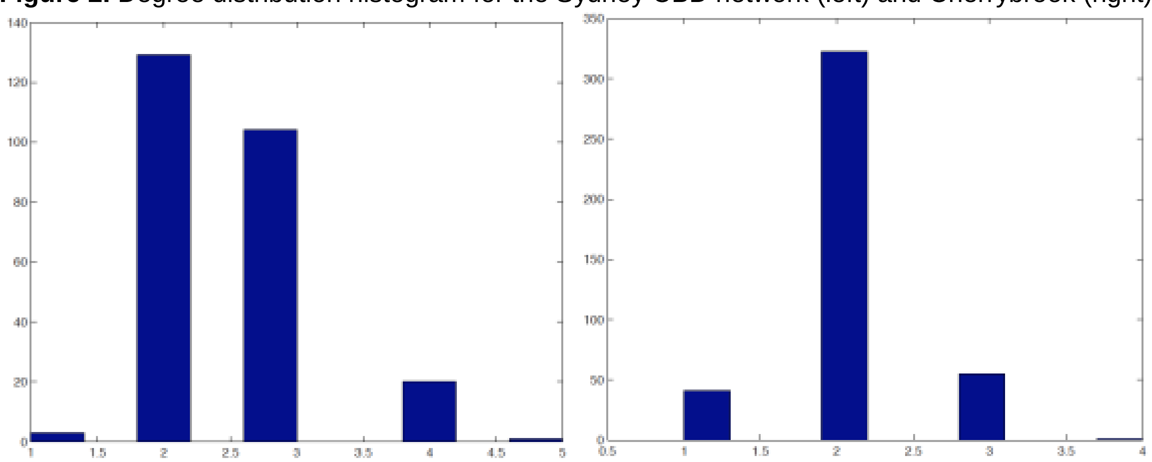
Another set of indices we study are centrality indices: degree, closeness, and betweenness centralities (Barthelemy, 2011; Porta et al., 2006, 2009). The concept of centrality measures how "near", "central", "accessible", or "connected" a node or an edge is from other nodes and edges in the rest of the network. Centrality measures tell us about the most central locations of a place. Understandably, there are edge effects in these measures, i.e., the area that is mapped for the network will decide to a large extent which nodes in the network appear as most central. However, even with the edge effects, it is still useful to have this information, since it can largely inform planning, design, and location of public facilities and spaces within a local area or suburb level.

Degree centrality is defined as simply the number of other nodes a single node is connected to. For undirected networks studied here, this is simply the sum of the rows (or columns) of the adjacency matrix,

$$\deg(i) = \sum_{j=1}^{n} A_{ij}$$

In planar street networks, the degree distribution of nodes is severely restricted because of spatial constraints. Every intersection can have, on an average, only 2-4 streets coming out. This characteristic differentiates planar and spatially embedded networks from other classes of networks. Therefore, degree centrality is not a significant measure for planar networks. For example, Figure 2 shows the histogram of degree distribution of the Sydney CBD and the Cherrybrook networks. It is obvious that most nodes have 2, 3 or 4 links. While the Sydney CBD has a dominance of 2- and 3-link nodes, with significant number of 4-link nodes, the Cherrybrook network has a dominance of 2-link nodes, with significant number of 1- and 3-link nodes (nodes with degree 1 are dead-ends), with almost no 4-link nodes.

Figure 2. Degree distribution histogram for the Sydney CBD network (left) and Cherrybrook (right)



Closeness centrality measures how "near" a node is to other nodes in the network, i.e., it measures the mean distance of a node to other nodes in the network. It is defined in terms of the *shortest path* or *geodesic distance* d_{ij} between two nodes. This d_{ij} is the smallest sum of edge lengths throughout all the possible paths in the graph from i to j in a weighted graph (or the minimum number of edges in a binary unweighted graph). If d_{ij} is the length of the shortest path in the network between nodes i and j, the mean shortest path distance from i to j, averaged over all nodes j in the network is given by

$$l_i = \frac{1}{n}\sum_{j=1}^{n} d_{ij}$$

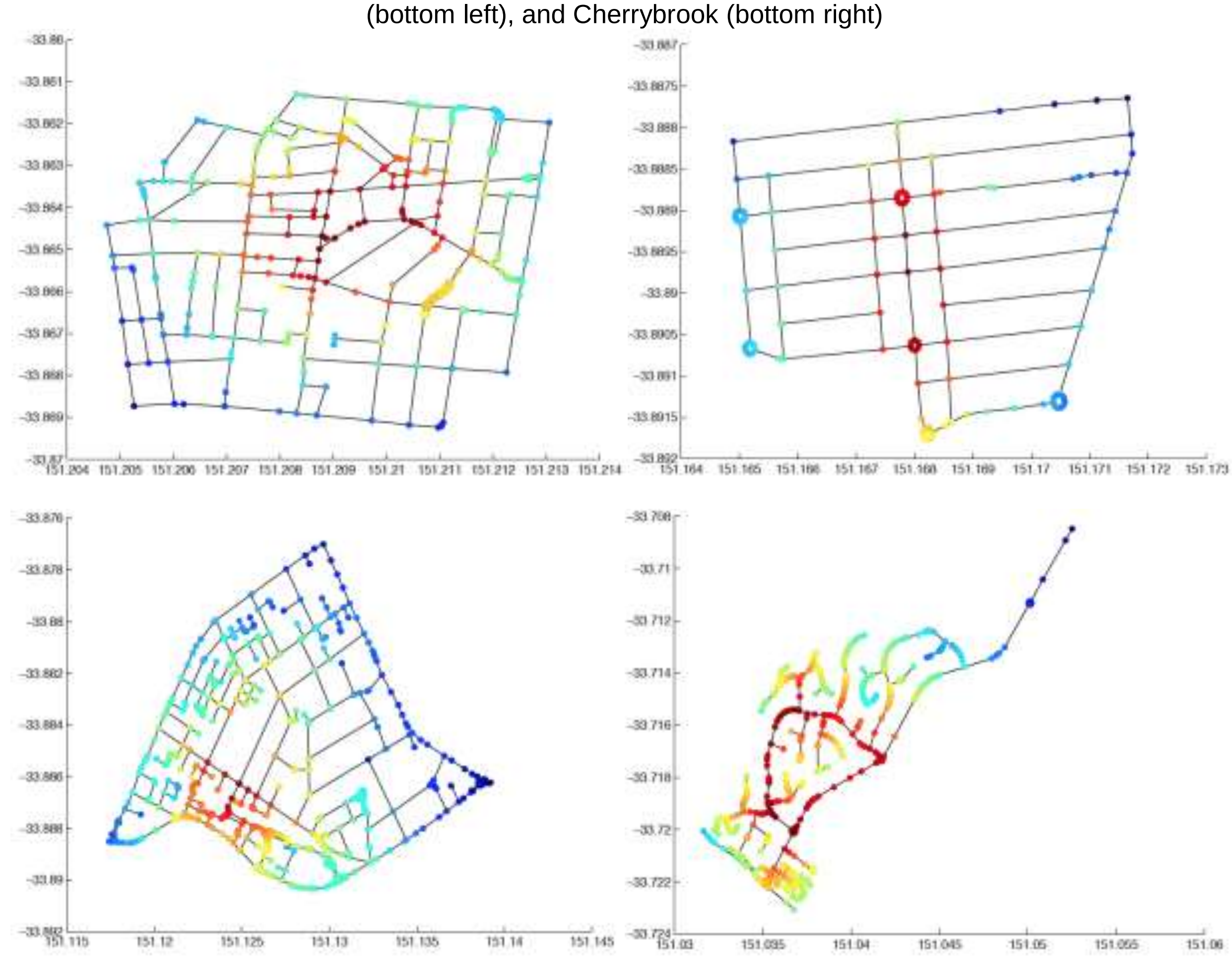
The inverse of this is the closeness centrality

$$c = \frac{1}{l_i}$$

This implies that the larger the distance between a node and all other nodes, the less "close" this node is to the network as a whole, and vice versa, the less the mean geodesic distance between them, the closer this node to the others.

Figure 3 shows the closeness centrality plots for all nodes of networks for Sydney CBD and selected suburbs. The most central nodes appear as red, with the blue end of the color scale showing the nodes with least closeness. It is interesting to note that for tree like suburbs with a typical cul-de-sac plan form, the nodes that form the coarsest-level largest cycle in the network shows highest closeness centrality. Edge effects are apparent. Nonetheless, it is clear, that it is not only the edge effects of defining the boundaries that affect the closeness centrality. For example, in the cases of the Sydney CBD and Ashfield, the link density and connectivity structure affects the closeness centrality, pushing it off the centre of the network. In Stanmore, the flat link density structure causes the most central nodes to sit exactly in the middle of the network.

Figure 3. Closeness centrality for Sydney CBD (top left) and suburbs Stanmore (top right), Ashfield (bottom left), and Cherrybrook (bottom right)

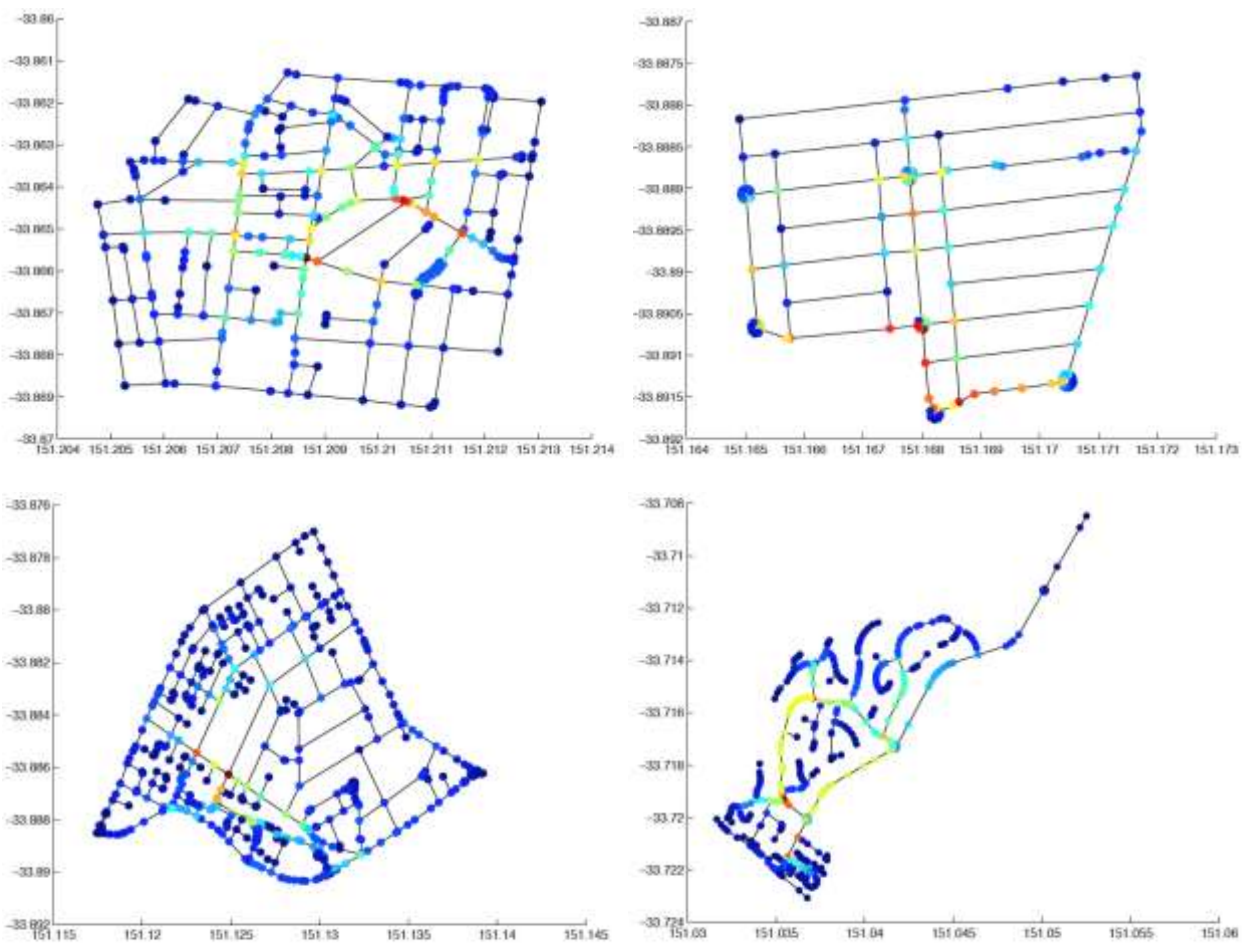


Betweenness centrality is a measure of the extent to which a node lies on the paths between other vertices. We imagine that quantities such as traffic, or information, "flow" between the nodes. Then, if a node lies on many shortest paths or geodesics (defined above) between multiple nodes, it will have a high betweenness. It need not be connected to many other nodes, but may fall on the path between many connected nodes. Nodes or moves with high betweenness may have considerable influence in a network because they can control the “flow” in the network. Thus, betweenness centrality captures the intuitive notion that interactions between two non-neighbouring nodes are affected by intermediate nodes that can have control over the communication. The assumption made is that the most efficient flow happens along the shortest geodesic paths. If n_{jk} is the number of shortest paths linking two nodes j and k, and of these $n_{jk}(i)$ is the number of nodes linking j and k that contain the node i, the betweenness centrality B is defined as

$$B = \frac{1}{(n-1)(n-2)} \sum_{j,k \in V, j \neq k, j,k \neq i} \frac{n_{jk}(i)}{n_{jk}}$$

The factor in front of the ratio is a normalizing factor that ensures that B is between 0 and 1. Figure 4 shows the betweenness centrality for the Sydney CBD and suburbs. The betweenness centrality captures the series of nodes forming skeletal paths that define the structure of the network, signifying that the most number of shortest paths in the network pass through these nodes.

Figure 4. Betweenness centrality for Sydney CBD (top left) and suburbs Stanmore (top right), Ashfield (bottom left), and Cherrybrook (bottom right)



Conclusions and Discussion

The network theoretic analysis presented above for the Sydney CBD and some representative suburbs brings out interesting bases for typological classification of the fine scale structure of cities. First, it moves traditional typological classification of plan forms or street networks in cities beyond the qualitative classifications established by historical research. It establishes a continuous spectrum of forms, with perfect tree-like and perfect mesh-like or grid-like forms defining two ends of the spectrum, based on the alpha, beta, and gamma indices, and the cyclomatic number. The fine scale structure of street networks of various suburbs in the city are then shown to sit somewhere on this continuous spectrum, with continuously varying levels of tree-like and mesh-like structure. As we have discussed, it is possible for tree-like and mesh-like behavior to mix and merge together at same or different hierarchical levels in the city structure.

The most important implication of such a lens is the opportunity that it provides to assess morphological changes in city structure. Even if a suburb begins as a pure grid layout or a cul-de-sac layout, structural changes over time may result in typological changes. Although we have presented the preliminary analysis here for static snapshots of plans, performing the same analysis over multiple snapshots in time and tracking a change in the values of the indices will permit analysis of how city structure evolves over time.

Second, we showed an analysis of centrality measures of local scale structures of cities, following the work in (Barthelemy, 2011; Porta et al., 2006, 2009). Centralities of different types can bring out the “central” and “important” nodes and edges in a plan, and can serve as a valuable global level design, planning, and analysis tool. Further, similar to the ideas discussed in the paragraph above, performing the same centrality analysis for the same part of the city over multiple points in time will permit analysis of how centrality evolves with changes to the physical structure of the city. Most importantly, in future work, the analysis of centrality measures will be particularly useful in bringing out the relationship between structure of cities, the dynamics of activity occurring on the spatial structure, and the control that structure and dynamics mutually exert on each other. A particularly interesting future direction will be to detect the embeddedness of social and economic community structures in spatial networks, and investigate the relationships between community structures, centrality measures, and spatial embeddedness. Community structure detection in networks is a very active area of research (Sarkar and Dong, 2011; Sarkar et al., 2013; Fortunato, 2010; Barthelemy, 2011; Newman, 2010), however, its relationship to spatial embeddedness is still not deeply understood, and presents an area of future potential research.

The ideas presented on centrality have been previously applied for different cities and large scale city-wide plans (Porta, 2006; Barthelemy, 2013). In making such comparisons, cities may be primarily classified as either centrally planned or naturally self-organized (Porta, 2006). In reality, however, each city is a mixture of the two organizing forces of central planning and self-organization (Barthelemy, 2013). In applying these analysis techniques to different parts of the same city, we have brought out the insight that often, the same city can contain a range of structural typologies, and that very different typologies can flow into each other or be embedded in each other. This paper presented preliminary analysis of the beginning stages of a research program and represents a first step in modeling an Australian city as a complex network to provide a mechanism for studying typology classification of fine scale city structure. In future work, we plan to extend this analysis framework to analyse Sydney in greater detail, capturing a larger range of suburbs, compare the structure of different Australian cities, and using similar graph theoretic mechanisms, investigate the relationship of socio-economic processes and activities and their embeddedness in and relationship to physical space.

References

Alexander, C. (1977) *A Pattern Language: Towns, Buildings, Construction* (Oxford University Press, USA).

Alexander, C. (1966) A city is not a tree, *Design, London: Council of Industrial design, No.206.*

Barthelemy, M. (2011) Spatial networks, *Physics Reports*, 499, pp.1-101.

Barthelemy, M., Bordin, P., Berestycki, H. and Gribaudi, M. (2013). Self-organization versus top-down planning in the evolution of a city, *Nature Scientific Reports,* 3, Art. Num. 2153, doi:10.1038/srep02153.

Batty, M. and Longley, P. (1994) *Fractal cities: A geometry of form and function*. (Academic Press, San Diego, CA and London).

Batty, M. (2007) *Cities and Complexity: Understanding Cities with Cellular Automata, Agent-Based Models, and Fractals,* (MIT Press, MIT).

Dictionary of Sydney (2013) http://www.dictionaryofsydney.org/entry/planning#page=all&ref=note1 , article on “Planning”, last accessed 26th August 2013.

Dictionary of Sydney (2013) http://dictionaryofsydney.org/entry/built_environment?zoom_highlight=grid , article on Built Enviroment, last accessed 26th August 2013.

Dictionary of Sydney (2013) http://dictionaryofsydney.org/entry/roads?zoom_highlight=road+networks, article on Roads, last accessed 26th August 2013.
Fortunato, S. (2010). Community detection in graphs. *Phys. Rep.*, 486:75–174.
Haggett, P. and Chorley, R.J. (1969) Network analysis in geography (Edward Arnold, London).
Hillier, B. and Hanson, J. (1984) *The social logic of space* (Cambridge University Press, Cambridge).
Jacobs, J. (1961) *The Death and Life of Great American Cities* (Vintage Books, Random House, New York).
Kostof, S. (1991) *The city shaped: urban patterns and meanings through history* (Thames and Hudson,
Newman, M. E. J. (2010) *Networks: An Introduction* (Oxford University Press).
Porta S, Crucitti P, and Latora V. (2006) The network analysis of urban streets: a primal approach, *Environment and Planning B: planning and design*, 33(5), pp.705-725.
Porta S, Strano E, Iacoviello V, Messora R, Latora V, Cardillo A, Wang F, Scellato S. (2009) Street centrality and densities of retail and services in Bologna, *Environment and Planning B: Planning and Design*, 36, pp.450-465.
Rossi, A. (1984) *The architecture of the city* (Opposition Books).
Sarkar, S. and Dong, A. (2011) Community detection in graphs using singular value decomposition, *Physical Review E*, 83(4), 046114.
Sarkar, S., Henderson, J.A., and Robinson, P.A. (2013) Spectral characterization of hierarchical network modularity and limits of modularity detection, *PLoS One*, 8(1), e54383.
London).