Topological Analysis of China Urban Streets

Zheng Ren
Department of Technology and Built Environment, Division of Geomatics
University of Gävle, 801 76 Gävle, Sweden
E-mail: tgm15zrn@student.hig.se

Abstract

In this study, topological analysis was adopted to large urban street networks in China. Top six cities in China were chosen as samples according to the rank of street junction numbers. The boundary of each city are delineated from the biggest patch of natural cities (Jiang and Miao, 2015) derived from road junctions. The three biggest patches also contain the most of number of the road junctions are called metropolises, they are Guangzhou, Shanghai and Beijing. The other three cities are more or less in similar scale, and they are Hangzhou, Chengdu and Wuhan. The topological analysis is based on named streets (Jiang and Claramunt, 2004) in all six cities to find the hierarchical patterns. The results conform the conclusion that 80% of streets are less connected (below the average), while 20% of streets are well connected (Jiang, 2009). The well-connected streets in six cities take a proportion from 21% to 26%. It was also found that the urban street networks for all cities show small-world property but not all of them show the scale-free properties. The betweenness distributions were also examined, most of them fit a power-law distribution. It was also found that the hierarchy of betweenness is less than degree by applying the head/tail breaks to all of the cities, and the levels of hierarchy is usually higher in metropolises than smaller cities. Furthermore, the streets with highest betweenness may imply that they are inherently bridges to large extent.

Keywords: Topological analysis, head/tail breaks, hierarchical levels, small-world, scale-free, ht-index, named streets, OSM data

1. Introduction

Geographical Information Systems (GIS) can be described as computer based decision making systems that are used to capture, store, manipulate, analyze, manage, and visualize all types of spatial or geographical data. The capability of spatial data analysis makes the GIS distinct from any other computerized decision making systems such Computer Aided Design systems (CAD). The spatial is special because of the spatial dependency or autocorrelation of spatial objects. The first law of geography (Tobler, 1970) reveals that "Everything is related to everything else, but near things are more related than distant things." Locally, the spatial objects are homogenous that everything is similar to each other. Globally, there is far more small things than large ones, the heterogeneity of the conventional space can be better described by the new term, scaling. Spatial analysis plays an important role in GIS due to the fact that it distill the knowledge from massive information so that we can capture the essence of spatial environment.

The Euclidean geometry and linear way of thinking have dominated the approach of spatial analysis in GIS for decades. However, there is a novel paradigm shifting that emphasizes the power of topology and scaling way of thinking especially in the field of spatial analysis. Topology refers to the relationship between spatial objects which is different from the Euclidean geometry using points, lines and angles to describe its axiomatic system. Scaling, also called scale-free can be simply comprehended as there are far more small things, and the hierarchy structure lies between them. More specifically, scaling of geographic space refers to the fact that for a large geographic area, its small parts are much more common than the large ones (Jiang 2010). Those two key points penetrate into every perspective of the spatial analysis. In this study, topological analysis for urban street networks are adopted to uncover the underlying hierarchical structure in six biggest cities in China.

Streets in an urban environment are not stand-alone objects, they are inter-connected and integrated to each other that forms a network topology (Jiang, 2009). Network analysis in conventional GIS is based on a graph view in which the intersections of linear features are regarded as nodes, and connections between pairs of nodes are represented as edges (Jiang and Claramunt, 2004). For example, the intersections of streets are the nodes and the street are edges so that we can calculation the shortest path. Another filed of topology used in traditional GIS is the Topologically Integrated Geographic Encoding and Referencing (TIGER) from U.S. Census Bureau. This kind of topology is also feature-based that it records the adjacent lines or polygons of each features.

Another aspect of topology is regarding each street as a node and the connections between streets are the links in a graph. Based on this graph, we can calculate the structural properties of each street (node). Such structural properties are within the scope of urban morphology. Space syntax (Hillier and Hanson, 1984) is a set of tools for urban morphology analysis based on graph theory which models the urban space using axial maps. In this study, 'named-streets' (Jiang and Claramunt, 2004) are used for topological analysis rather than axial lines. 'Named streets' retain more reality in a given street network. It not only concerns about the geometric connections, but also concerns the semantic and cultural connections of urban streets. In China, named streets may reflect a higher level of abstraction because the street are more related historically and culturally.

In this study, the node in a connectivity graph is entire named street instead of street segment. Such urban street network is hierarchically organized, it turns out that 80% of streets are less connected (below the average), while 20% of streets are well connected (Jiang, 2009). The scaling pattern of the connectivity is the signature of the complexity. As a result, some properties of complex networks such as small-world and scale-free properties are also examined in 6 biggest China cities.

The remainder of this paper is organized as follows. In section 2, some basic conceptions are introduced. In chapter 3 is about the data and methods used for the topological analysis. The methods are respectively about extracting data from Open Street Map (OSM) in ArcGIS, measuring the average path length and clustering coefficient of small-world in pajek and examining power-law degree distribution in Matlab. Results of hierarchical patterns of 6 cities, measurements of small-world and scale-free properties are included in section 4. In section 5, the study is discussed and in last section, I draw some conclusions.

2. Theoretical concepts

In this study, some theoretical concepts, such as power distribution and head/tail breaks and natural cities are introduced in this section. Power law distribution is universal in terms of probability and frequency. Head/tail breaks provides a novel way to uncover the hierarchal level of an urban street networks. And natural cities conform that scaling pattern of geographical phenomenon.

2.1 Power law distribution

Power law distribution indicates that given a variable of interest x, the probability of occurrence of the variable value y follows the formula that $y = cx^{-a}$ where a is the power law exponent and c is a constant. Many world phenomena fit power laws such as the word frequency in our languages (Zipf, 1949) that only few words are always used and most of the words are not frequently used in English. The histogram of power law distribution is a right-skewed curve and it is a straight line if we take logarithm at both sides of the equation.

2.2 Head/tail breaks

Head-tail breaks is a novel classification schema for the data and it is used for the long tail distribution. A long tail distribution consists of a head that the values are extremely high and the tail part is very low but long. The imbalanced structure can be expressed as "there are far more small things than larger ones" (Jiang, 2012). And this imbalance is recursive in the different hierarchies, which contributes to the basic idea of head/tail breaks. Jiang (2012) introduced that "the head/tail breaks method partitions the data values into two parts around the arithmetic mean and continues the partitioning for values above the mean iteratively until the head part values are no longer heavy-tailed distributed. " We can use this method to capture the underlying hierarchical structure of street networks. This structure is very realistic and natural because the number of hierarchy and the content in each hierarchy is determined by the data itself.

2.3 Natural cities

The term 'natural cities' refers to the spatially clustered geographic events, such as agglomerated patches from individual social media users' location (Jiang and Miao, 2015). Head/tail breaks (Jiang, 2012) is applied to generate the natural cities based on the scaling pattern of the earth surface in terms of density. The natural cities are a good representation of real world because the process to derive the natural cities in natural and very close to the reality. In this study, it is used to define the boundaries of each city based on the street junction points in order to make the result more realistic.

3. Data and methods

The China street data is the OpenStreetMap (OSM) roads data downloaded from a German free server called Geofabrik. It extracts the OSM data and re-organizes the data everyday and we can easily extract roads from the datasets. In addition, the natural cities have been generated before this project, the method to generate the natural cities can be obtained from the appendix of the paper by Jiang and Miao (2015). The six biggest cities in terms of road junction scale were chosen as samples, they are Guangzhou, Shanghai, Beijing, Hangzhou, Chengdu and Wuhan. The first three cities are much more bigger than the others in size.

3.1 Extracting OSM data

The first step is to extract the wanted study areas from the OSM roads data of whole China. The boundaries of chosen cities are determined by the largest patched of the natural cities derived from the roads junctions. The natural cities are the clustered geographical events that show a scaling pattern. Namely, there are far more small patches than large patches. But in terms of importance, the very few large patches usually stand for the most important objects. Unlike the real cities, the natural cities covers a larger areas. In figure 1, the natural city of Beijing actually covers the Beijing and Tianjin in the real world. This is because the road is in high density in this area, and the natural cities are based on the fact that there are far more low-density locations than high-density locations on the earth surface (Jiang and Miao, 2015). In this way, the boundaries of the cities are more accurate and nearer to the reality. In ArcGIS, we use the clipping function to extract the roads in each city. Once the clipped roads of each city are derived, we can later build up topology to those roads. This was conducted based on Data Interoperability Tool within ArcGIS. The ArcInfo Coverage file is then generated with the topological information, and the final result should be in segments as shown in the figure 2.

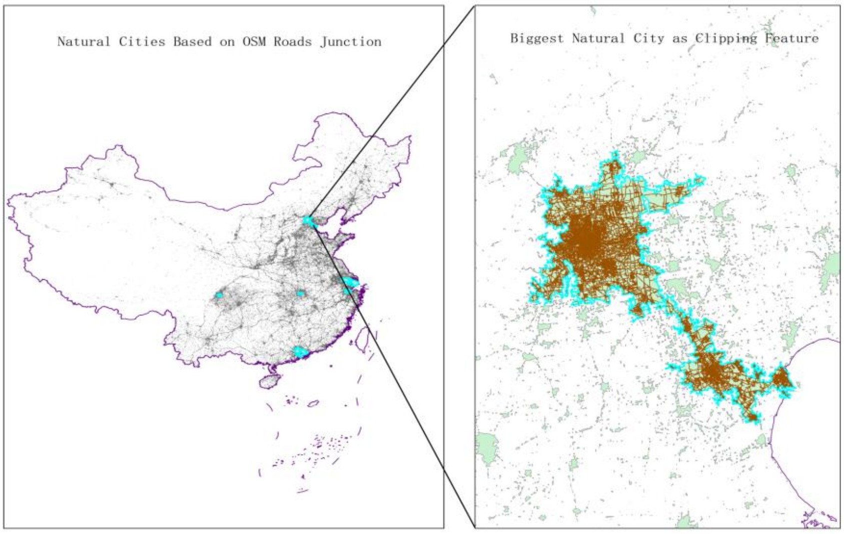


Figure 1. Natural city derived from roads junctions as a clipping boundary

(a) Streets before building up topology (b) Streets after building up topology

Figure 2. Importance of chopping streets at each junction

3.2 Finding the hierarchical patterns

After building up topology, the data is ready for the analysis of hierarchical patterns in terms of connectivity. This is the close to the urban morphology analysis by using the space syntax (Hillier and Hanson, 1984). The general principle of space syntax is spaces can be broken down into components and adopt topological analysis in term of graphs and networks as a representation. Any space can be perceived as large-scale space and small-scale space respectively. The small-scale space refers to the continuous spaces around us and can be perceived easily from where we are. Those infinite continuous small-scale spaces are the components of a large scale space. Consequently, every decomposed small-scale space can be a part of the connectivity graph enabling us to take morphological analysis based on topological point of view. There are several morphological measures such as connectivity, control value, global and local integration can be used to describe the urban structure both locally and globally. We can easily calculate those parameter easily with a GIS plug-in software called Axwoman (Jiang, 1999) for urban morphological analysis based on GIS. Moreover, The integration of space syntax into geographical information system (GIS) offers new perspectives to the development of urban morphology studies (Jiang & Claramunt, 2002).

In Axwoman, space syntax parameter such as connectivity can be calculated from a natural roads. The natural roads are those segments connected to their adjacent one at the junction with Gestalt principle for good continuation (Jiang, Zhao & Yin, 2008). The default threshold of generating natural roads from segments (also called tracking strokes) is 45 degree, which means if the angel between segments are within 45 degree, they will be join into one natural road. Since this study is based on the named streets, an extra step called merging named streets should be performed before tracking strokes. This can be automatically done in the Axwoman, just by clicking the button of join named streets. Consequently, the segments with same name are merged and those segments without name are left to do the regular tracking strokes. In this way, both geometric and cultural characteristics of streets to be analyzed are considered. Finally, the connectivity value of each named street can be calculated.

In order to see the hierarchical patterns of the connectivity, head/tail breaks was applied to each city. First, the mean value of the connectivity should be calculated. The data can be divided into two parts, the head part is those values above the mean, and tail part is lower than the mean value. Applying this division iteratively to the head part until there are no longer head and tail parts. The hierarchical patterns were finally derived after the head/tail breaks (Table 1). The Axwoman also has a function called HT mapping, it can define the constraint manually to stop the classification, the default setting is 40%. In this study, I set the parameter in 46% (highest percentage in all levels of all cities) after manually classification in order to generate the symbolized maps according to the head/tail breaks.

Table 1. Head/tail breaks for Beijing named streets (Note: # = the number, % = percentage)

#Named	# head	% head	# tail	%tail	Mean value
30136	7929	26%	22207	74%	4.86
7929	1646	20%	6283	80%	12.05
1646	418	25%	1228	75%	31.5
418	122	29%	296	71%	69.75
122	33	27%	89	73%	131.82
33	12	36%	21	64%	235.45

3.3 Measuring small-world and scale-free properties

A small-world network is a kind of complex network with the small separation globally and high clustering locally. Mathematically, such degree of separation can be measured by average path length. In a graph, the average path length of one node is the average distance between all of other nodes in the network. The distance here is in topological perspective, i.e. intermediate steps rather than lengths in Euclidean geometry. In graph theory, a clustering coefficient are used to measure the degree to which nodes in a graph tend to cluster together, it is the likelihood of the real links out of all possible links of one nodes in a graph (Watts and Strogatz, 1998). The small-world networks combine common properties between regular and random graph with short average path length and high value of clustering coefficient. For comparison, these two measures were calculated for a random graph which has the same nodes. For random graph, $L_{random} = \ln n / \ln m$, and $C_{random} = m / n$. The n is total number of vertices and m is the mean value of connectivity (Watts and Strogatz, 1998). If the average path length calculated in Pajek is short and the clustering coefficient is higher than the random graph, we can say that the network is a small-world network.

Scale-free properties illustrate that nodes connectivity within a large network follows a power-law distribution (Barabasi and Albert, 1999) which means there are only few well-connected nodes but large number of less connected nodes. The networks have also to be examined to see if they are scale-free networks. There are three kind of centralities can be calculated in Pajek. They are degree centrality, closeness and betweenness centrality. The degree centrality is same as the connectivity and the closeness is used to describe the distance from one node to all other nodes and betweenness shows the number of times a node is used for any other two nodes as bridge along the shortest paths. The degree and betweenness in this study are used to see if they fit the power law distribution with help of Matlab. There are three main steps in Matlab, first is fitting a dataset to a power law, calculating the alpha and minimum value of the dataset. Second is to examine the p value, if it is higher than zero, the power law is a plausible hypothesis for the data. Third, we use pl-plot function to draw the plot for the data (Clauset, 2009).

4. Results

4.1 Hierarchical levels of top 6 urban street networks

After performing the head/tail breaks on the connectivity or degree in 6 chosen cities. The results of the hierarchical structures were derived. The hierarchical levels of 6 cities are various . This level is also called ht-index, which is defined as one plus the recurring times of far more small things than large ones at different scales (Jiang and Yin, 2014). The higher the ht-index, the more complex the network structure. Accordingly, the ht-index of three metropolises (i.e. Guangzhou, Shanghai, Beijing) are respectively 10, 9, and 7. Guangzhou who owns the most number of named streets (41412 in total) has the highest hierarchical levels, which indicates that the street network is very complex in terms of underlying structure (Table 1). This complexity was visualized in ArcGIS with the help of HT mapping in Axwoman, the highest connectivity streets were represented in red color and the lowest connectivity is blue, the whole color scheme follows a rainbow like combination (Figure 3). The maximum percentage of head part is 45% in the seventh division of Shanghai (Appendix A, Table 2). The rule of head/tail breaks was set to 46% so that all the datasets follow the same head/tail breaks rule. In other words, the automatically symbolization will generate the same results as the symbolization manually.

Table 2. Head/tail breaks for metropolis (Guangzhou) (Note: # = the number, % = percentage)

#Named	# head	% head	# tail	%tail	Mean value
41412	10056	24%	31356	76%	4.46
10056	2266	22%	7790	78%	11.22
2266	643	28%	1623	72%	26.54
643	195	30%	448	70%	51.25
195	72	36%	123	64%	84.98
72	24	33%	48	67%	121.43
24	8	33%	16	67%	168.42
8	3	37%	5	63%	214.38
3	1	33%	2	67%	250

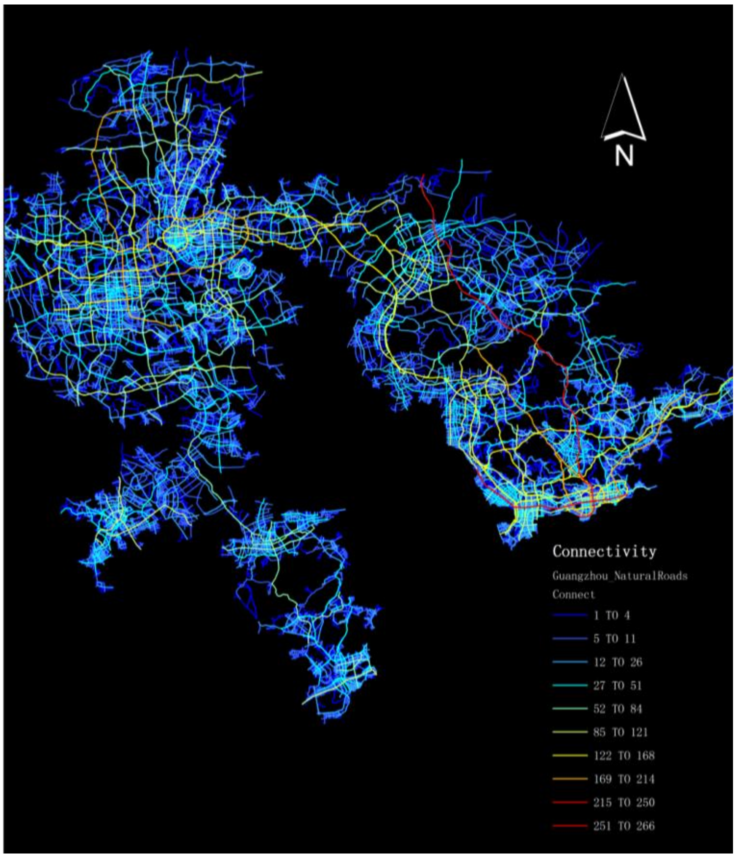


Figure 3. Hierarchical pattern of Guangzhou named streets

Based on the similar scales of the named streets major cities, the ht-index of three are derived. They are respectively 7 in Hangzhou, 6 in Chengdu and 8 in Wuhan. The ht-index is fewer than the metropolis city on average. For example, Chengdu is 6 but Guangzhou is 10 (Table 2,3). We can also found that some of the ht-index is similar to some cities both for metropolis and major city, e.g. Beijing and Hangzhou are both 7. However, the difference between metropolises and major cities are quite big. For example, in metropolises, they are more than 200 while they are less than 100 in major cities (Appendix A Table 1 - 6).

In all six cities, only less than 26% named streets are well connected. In Shanghai, there are only 21% of the named streets are well connected. This result conforms that urban streets demonstrate a scaling law (Zipf 1949) and can be characterized by the 80/20 principle (Jiang, 2008). The minority street with greatest connectivity are often the most important streets, the red streets in all six cities are always long and they tend to form the main frame of the city networks (Appendix B Figure 1-6). It has been proved that the 20% of well connected streets accommodate 80% of traffic flow (Jiang, 2008). Accordingly, the red ring roads in Chengdu (Figure 4), may face the risk of traffic congestion in rush hours.

Table 3. Head/tail breaks for major city (Chengdu) (Note: # = the number, % = percentage)

#Named	# head	% head	# tail	%tail	Mean value
4105	961	23%	3144	77%	5.3
961	254	26%	707	74%	13.97
254	68	26%	186	74%	29.86
68	23	33%	45	67%	60.31
23	8	34%	15	66%	100.39

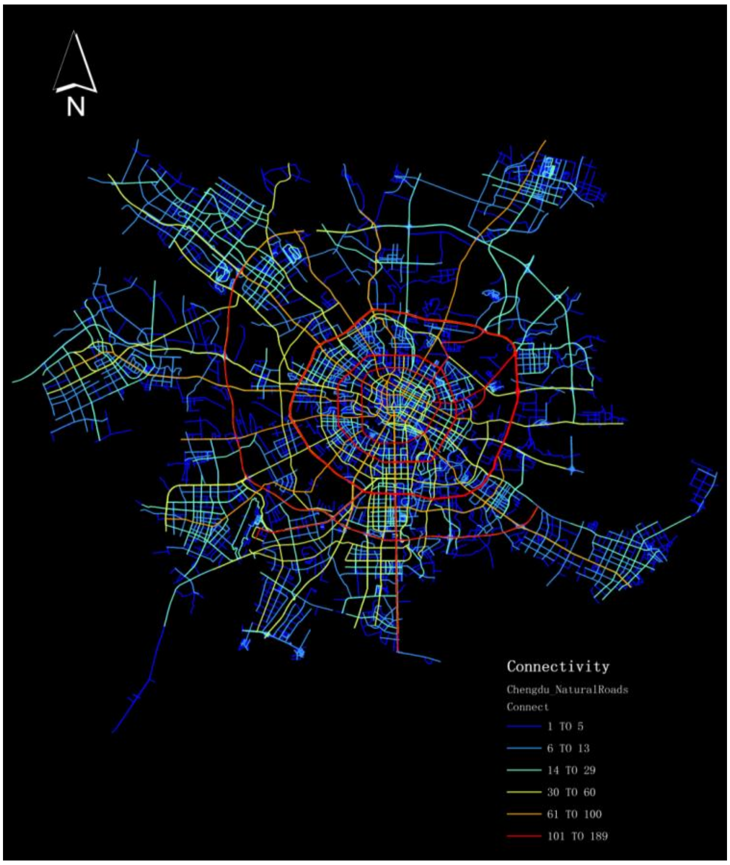


Figure 4. Hierarchical pattern of Chengdu named streets

4.2 Small-world property of urban street networks

In second part of study, small-world behaviors of the named street networks were evaluated. Average path length and clustering coefficient of 6 cities in China were calculated together in Pajek (Table 4.) These measurements were also derived in the random network with same number of named streets in each city.

Table 4. Small World Properties for Top 6 China Cities.

City	N	M	L	Lr	CC	CCr
Guangzhou	41412	4.46	10.01	7.11	0.33	0.00011
Shanghai	33275	5.09	7.75	6.40	0.29	0.00015
Beijing	30136	4.86	9.12	6.52	0.31	0.00016
Hangzhou	5934	5.14	7.76	5.31	0.30	0.00087
Chengdu	4105	5.30	5.59	4.99	0.31	0.00129
Wuhan	6279	4.53	6.77	5.79	0.36	0.00072

Note:

N = the number of named streets *M = the average connectivity or degree centrality*

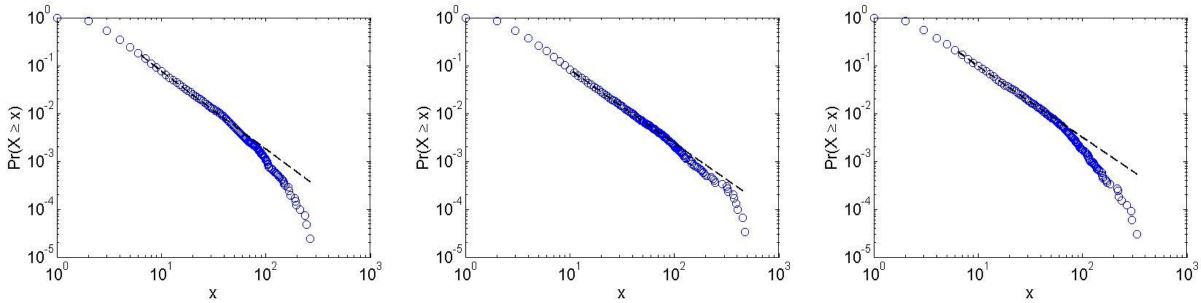
L = the average path length *Lr = L for random graph with same number of nodes*

CC = clustering coefficient *CCr = CC for random graph with same number of nodes*

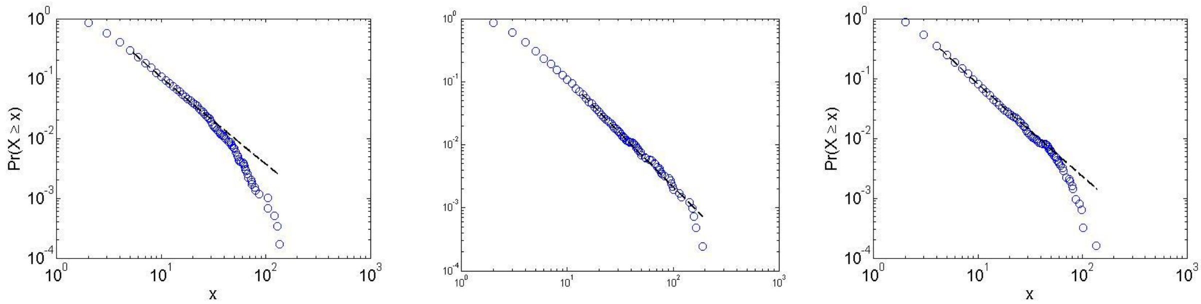
In general, the calculated results showed that the six urban named streets networks have small degree of separation. In Guangzhou and Beijing, the average path lengths are 10 and 9 which are relatively high. For other street networks, the L value is smaller than 8. In Shanghai, although the number of the streets are big, the separation is quite small both in reality and random graph. All cases displayed that CC >> CCr which means all the networks were highly clustered. We can draw the conclusion that all these named street networks had small-world property. Furthermore, Shanghai exhibits most obvious small-world property among six chosen cities because it has most number of streets but relatively small separation as other smaller cities.

4.3 Distribution of street degree

As for scale-free property, the degree distributions of six cities were calculated in Matlab using the three functions by Clauset, et al. (2009). The plfit function is for calculating the alpha and minimum values for datasets in power law. The plpva function is used for examining how well the data fit a power law, the indicator is p value, the higher it is, the more plausible the result is. The plplot is used to draw the logarithm plot against x and y axes for the data. The results are shown in figure 5. The data that fit a power law distribution tend to be straight lines in the logarithm plots. The p value of Shanghai and Wuhan is zero, so that the degree distributions of these cities are not power law distribution (Table 5). Shanghai and Wuhan did not have scale-free properties of their degree of named street networks. Guangzhou, Beijing, Hangzhou and Chengdu have the scale-free named street networks.



a) Guangzhou b) Beijing c) Shanghai



d) Hangzhou e) Chengdu f) Wuhan

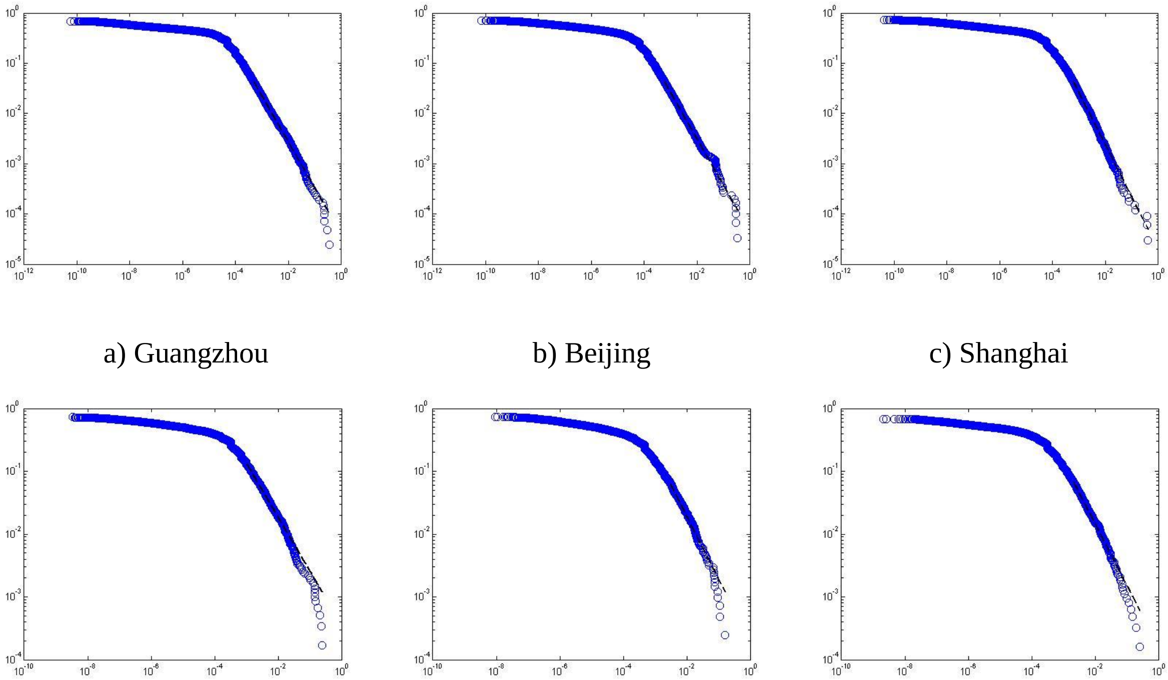
Figure 5. Power law distribution of degree centrality

Table 5. Statistical results of power law distribution of degree

City	Xmin	Alpha	P value
Guangzhou	6	2.60	0.03
Shanghai	6	2.46	0
Beijing	10	2.51	0.13
Hangzhou	5	2.42	0.01
Chengdu	13	2.68	0.67
Wuhan	4	2.51	0

4.4 Distribution of street Betweenness

The betweenness is also a centrality measurement of a network. The same procedure was taken to calculate the statistical result of power-law distribution. The results showed different case from the degree distribution. Except for Hangzhou whose p value is zero, the betweenness distribution of other 5 cities follow the power law distribution. Since the betweenness centrality indicates the number of shortest paths of all nodes to all others passing through certain node, it is used to describe the importance of a node as a bridge. Interesting is that in Beijing, Guangzhou and Shanghai, the highest betweenness streets are always those who act as a bridges to connect two parts or communities of the networks. They are the red color streets that can be easily found in Appendix B, Figure 1, 3, 4, 5 and 6.



a) Guangzhou b) Beijing c) Shanghai

d) Hangzhou e) Chengdu f) Wuhan

Figure 6. Power law distribution of betweenness centrality

Table 6. Statistical results of power law distribution of betweenness

City	Xmin	Alpha	P value
Guangzhou	0.00042	1.92	0.50
Shanghai	0.00053	2.04	0.08
Beijing	0.00051	1.92	0.36
Hangzhou	0.00091	1.86	0
Chengdu	0.00274	2.00	0.074
Wuhan	0.00191	1.98	0.098

5. Discussions

In this study, I used the natural cities to define city boundaries. There are two reasons, first the natural cities are derived based on the head/tail breaks and location density, they are good proxy of real cities, in the sense of understanding underlying patterns (Jiang and Miao, 2015). Second, the project aims to study the street networks, the natural cities are based on street junctions, they are inter-connected and originally from same data source, which can emphasize that the street network is self-organized and it is more natural than real cities. The metropolises such as Guangzhou, Beijing and Shanghai cover larger areas than the real cities. Because the pattern is largely depend on the streets density, there are few external factors to affect the data.

Topological analysis are based on named streets rather than axial lines. The named street is kind of culture dependent, the street segments are joined together not only based on the good continuity in terms of geometry, but also depend on the name of the streets. Considering the generality, a named street that is separated into two or more parts (e.g. North Ring Road and South Ring Road) are semantically aggregated (Jiang and Claramunt, 2004). Named streets make more sense in China, some of them are rather historical and culture dependent. Further work may consider the difference of the hierarchy levels between named streets from different culture backgrounds.

By studying the results of hierarchical levels, one of the metropolises like Beijing has same ht-index as a major city, Hangzhou. The number of named streets in is much more in Beijing than in Hangzhou, but the hierarchical level are the same. This can be explained that the ht-index captures the complexity of the geographical features well. The higher the ht-index, the more complex the geographic features (Jiang and Yin, 2014). The street network structure in Beijing is not as complex as other metropolises like Guangzhou and Shanghai, the regular shaped structure in Beijing, is decided by the traditional layout in a long history. The hierarchical levels have weak relationship with the number of streets, we can alternatively classify the geographical features according to their complexity instead of scales. However the maximum connectivity of the streets depends on the number of named streets in this case. The maximum connections of metropolises are over 230 while the major cities are only 100 on average. Applied once head/breaks, there are only less than 24% streets with higher connectivity than the mean values in all six cities, which proves that 80/20 principle well (Jiang, 2009). These minority of streets are fairly important that they account for a majority of traffic flow. These streets are extremely clear in Figure B 1 and Figure B 2, the red ring roads in Beijing and Cheng will be very vital to take the responsibility for the heavy traffic flows. These kinds of ring roads should be designed wider and carefully supervised by the related department in rush hours.

This study also examined the betweenness centrality in six cities, it has been found that the distribution of betweenness also fits a power law distribution in most cases. After comparison, the difference between the measurements of centrality can be easily seen in the symbolized maps by HT mapping (Appendix B). The patterns of Beijing and Guangzhou reflect the definition of the betweenness centrality well. The red streets that with highest betweenness are always the streets that links two clustered parts of networks, they are located in the narrow 'neck' of the networks of Beijing and Guangzhou (Figure B 1 and Figure B 3). These streets are not only act as bridges, but they are originally bridges themselves in the Hangzhou and Wuhan with rivers flowing through the city. The most famous Yangtze River Bridge in Wuhan is one of the two red streets with highest betweenness crossing the Yangtze River (Figure B 6). we can conjecture that betweenness as an indicator, can also be used to judge the functions of streets.

6. Conclusions

Topological way of thinking lying through all the study process. With the topological analysis of streets networks, the scaling (Jiang, 2010) patterns and hierarchical levels of named street networks are uncovered. It has been found that there are far more well connected streets than less connected streets and those well connected streets are also the most important in streets. All of the streets exhibits the small-world property and some of them are scale-free networks and the betweenness centrality distribution in most of the street networks are also power law distribution. Armed with topological approach, we can excavate the essence of scaling in terms of topology and semantics. Urban morphology, complex networks and graph theories are all based on topology. Spatial relationship and connectivity are the essence of topology, spatial is special because of the space is auto-correlated, it is homogenous locally and heterogeneous global. As a result, topology can be a very powerful approach in the spatial analysis.

Acknowledgement

I would like to thank Prof. Bin Jiang and Dr. Ding Ma for their guidance and help for this study. I also want to express my appreciation to those who wrote very detailed tutorials of generating natural cities and converting shape to net file. It is really a good way to learn such skills efficiently.

References

Barabási, AL. & Albert, R. (1999). Emergence of scaling in random networks, *Science,* 286, 509-512.

Clauset, A., Shalizi, C. R. and Newman, M. E. J. (2009). Power-Law Distributions in Empirical Data, *SIAM Review,* 51:4, 661–703.

Hillier, B. & Hanson, J. (1984). *The Social Logic of Space.* Cambridge University Press: Cambridge.

Jiang, B. & Miao, Y. (2015). The Evolution of Natural Cities from the Perspective of Location-Based Social Media, *The Professional Geographer*, 67:2, 295-306, DOI: 10.1080/00330124.2014.968886.

Jiang, B. (2012). Head/tail breaks: a new classification scheme for data with a heavy-tailed distribution. *The Professional Geographer,* 65:3, 482-494, DOI: 10.1080/00330124.2012.700499.

Jiang, B. & Claramunt, C. (2002). Integration of space syntax into GIS: new perspectives for urban morphology, *Transactions in GIS*, 6(3): 295-309.

Jiang, B. & Claramunt, C. (2004). Topological analysis of urban street networks, *Environment and Planning B: Planning and Design,* 31, 151-162.

Jiang, B. & Yin, J. (2014). Ht-Index for Quantifying the Fractal or Scaling Structure of Geographic Features, *Annals of the Association of American Geographers*, 104:3, 530-540.

Jiang, B. (2009). Street hierarchies: a minority of streets account for a majority of traffic flow. *International Journal of Geographical Information Science,* 23(8), 1033-1048.

Jiang, B., Zhao, S. & Yin, J. (2008). Self-organized Natural Roads for Predicting Traffic Flow: a Sensitivity Study. *Journal of Statistical Mechanics: Theory and Experiment*, July, 2008.

Jiang, B. (2010). Scaling of Geographic Space and its Implications. A position paper presented at *Las Navas 20th Anniversary Meeting on Cognitive and Linguistic Aspects of Geographic Space*, Las Navas del

Marques, Avila, Spain, July 5-9, 2010.

Jiang, B. (1999), SimPed: Simulating Pedestrian Flows in a Virtual Urban Environment, *Journal of Geographic Information and Decision Analysis*, 3(1), 21 - 30

Tobler W. (1970), A computer movie simulating urban growth in the Detroit region. *Economic Geography*, 46(2): 234-240.

Watts, D. J. & Strogatz, S. H. (1998). Collective dynamics of “small-world” networks, *Nature,* 393, 440-442.

Appendix A

Different scale of cities have different number of hierarchical levels. The results of head-tail breaks for the degree of top 6 cities are showed below. Note that # = the number, % = the percentage.

Table A 1. Metropolis (Guangzhou)

#Named	# head	% head	# tail	%tail	Mean value
41412	10056	24%	31356	76%	4.46
10056	2266	22%	7790	78%	11.22
2266	643	28%	1623	72%	26.54
643	195	30%	448	70%	51.25
195	72	36%	123	64%	84.98
72	24	33%	48	67%	121.43
24	8	33%	16	67%	168.42
8	3	37%	5	63%	214.38
3	1	33%	2	67%	250

Table A 2. Metropolis (Shanghai)

#Named	# head	% head	# tail	%tail	Mean value
33275	7067	21%	26208	79%	5.09
7067	1728	24%	5339	76%	14.65
1728	520	30%	1208	70%	33.9
520	181	34%	339	66%	63.76
181	57	31%	124	69%	99.74
57	20	35%	37	65%	148.81
20	9	45%	11	55%	205.6
9	4	44%	5	56%	256.44

Table A 3. Metropolis (Beijing)

#Named	# head	% head	# tail	%tail	Mean value
30136	7929	26%	22207	74%	4.86
7929	1646	20%	6283	80%	12.05
1646	418	25%	1228	75%	31.5
418	122	29%	296	71%	69.75
122	33	27%	89	73%	131.82
33	12	36%	21	64%	235.45

Table A 4. Major city (Hangzhou)

#Named	# head	% head	# tail	%tail	Mean value
5934	1351	22%	4583	78%	5.14
1351	392	29%	959	71%	13.83
392	131	33%	261	67%	27.7
131	50	38%	81	62%	45.36
50	17	34%	33	66%	64.78
17	6	35%	11	65%	88

Table A 5. Major city (Chengdu)

#Named	# head	% head	# tail	%tail	Mean value
4105	961	23%	3144	77%	5.3
961	254	26%	707	74%	13.97
254	68	26%	186	74%	29.86
68	23	33%	45	67%	60.31
23	8	34%	15	66%	100.39

Table A 6. Major city (Wuhan)

#Named	# head	% head	# tail	%tail	Mean value
6279	1560	24%	4719	76%	4.53
1560	372	23%	1188	77%	11.08
372	118	31%	254	69%	24.78
118	49	41%	69	59%	44.25
49	20	40%	29	60%	64.1
20	6	30%	14	70%	81.35
6	1	16%	5	84%	102.17

Appendix B

The hierarchical structures of 6 samples cities are shown in below. Head/tail breaks are applied based on two different centrality measurements, one is the connectivity, the other one is betweenness.

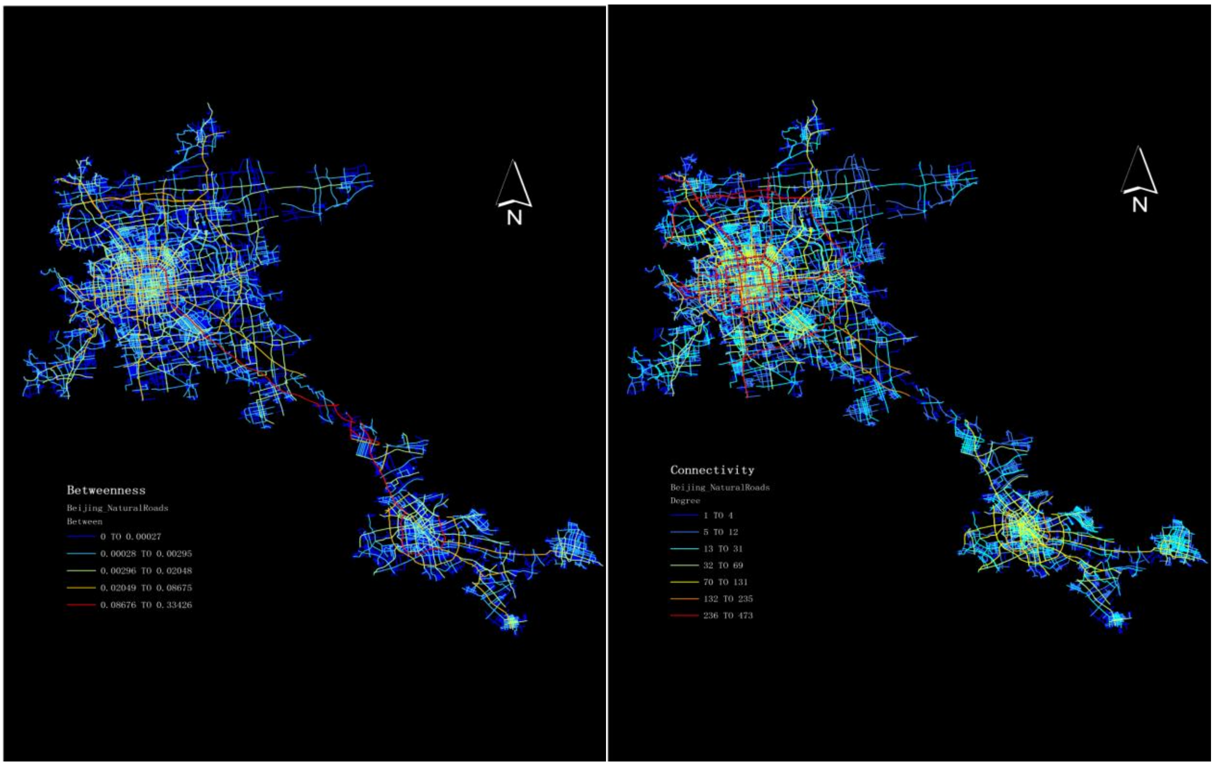


Figure B 1. Hierarchical pattern of Beijing named streets based on betweenness and connectivity

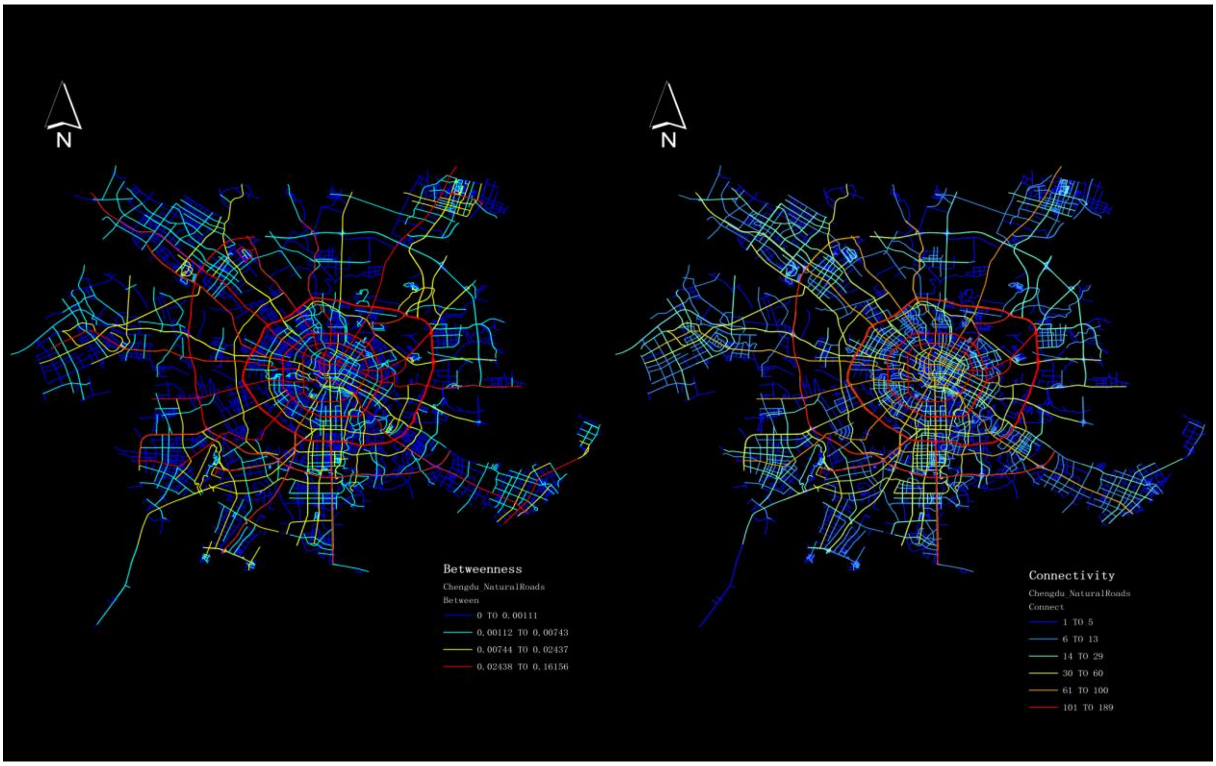


Figure B 2. Hierarchical pattern of Chengdu named streets based on betweenness and connectivity

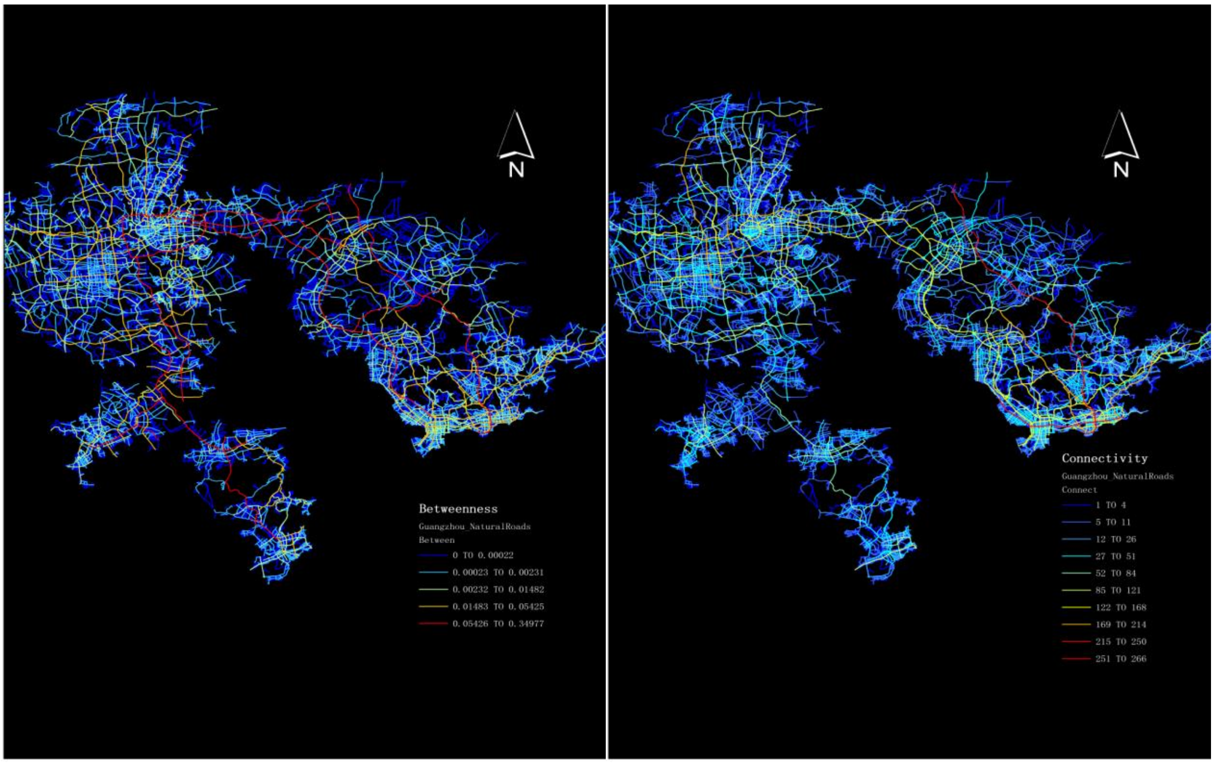


Figure B 3. Hierarchical pattern of Guangzhou named streets based on betweenness and connectivity

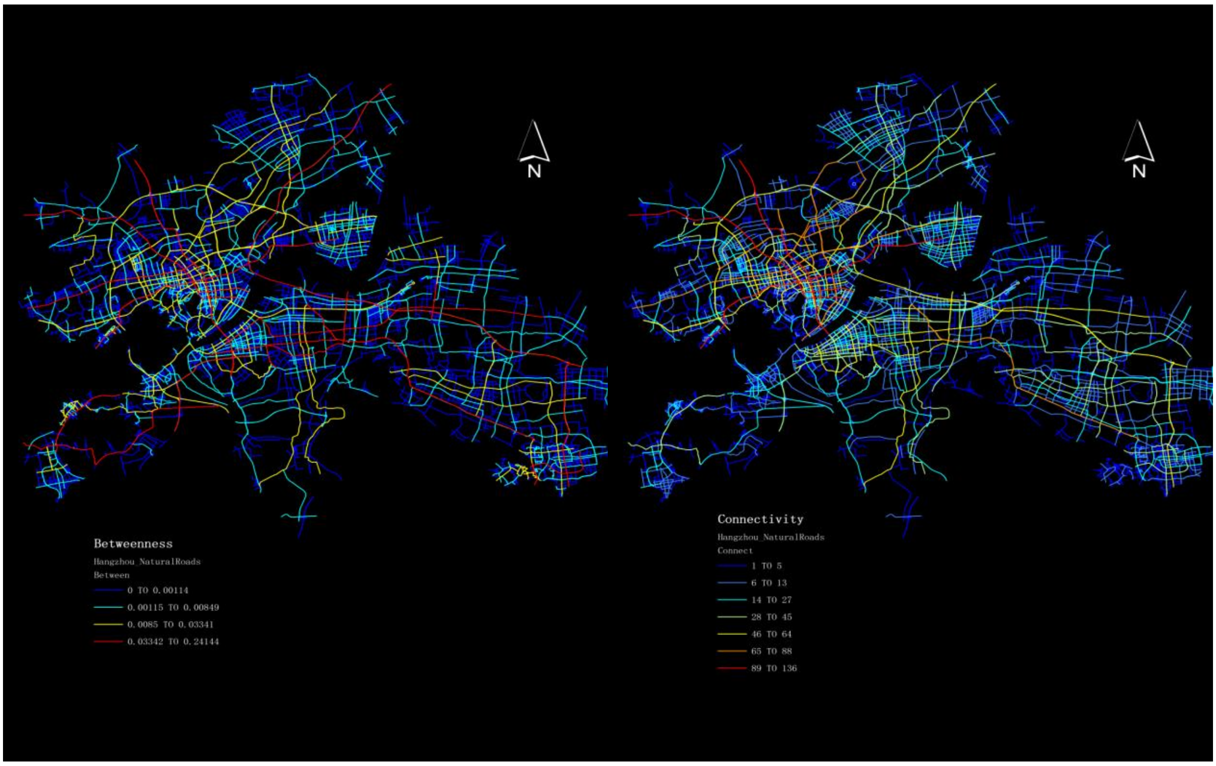


Figure B 4. Hierarchical pattern of Hangzhou named streets based on betweenness and connectivity

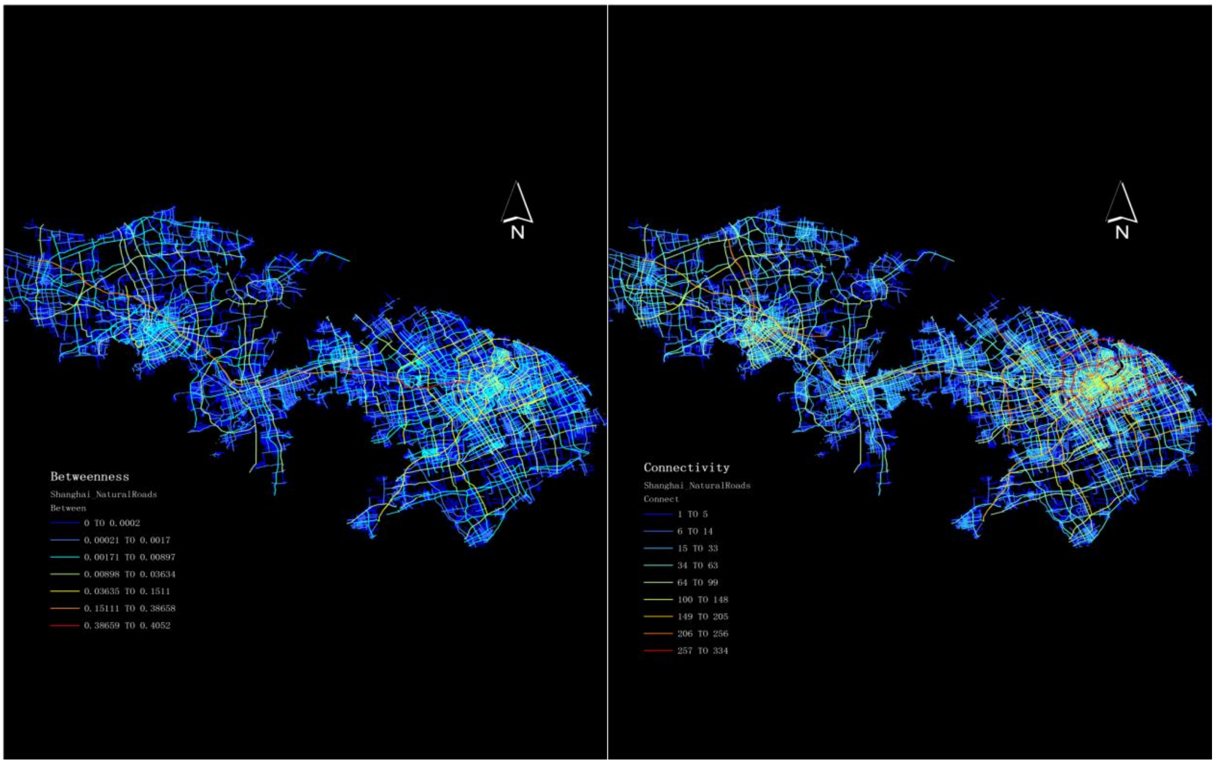


Figure B 5. Hierarchical pattern of Shanghai named streets based on betweenness and connectivity

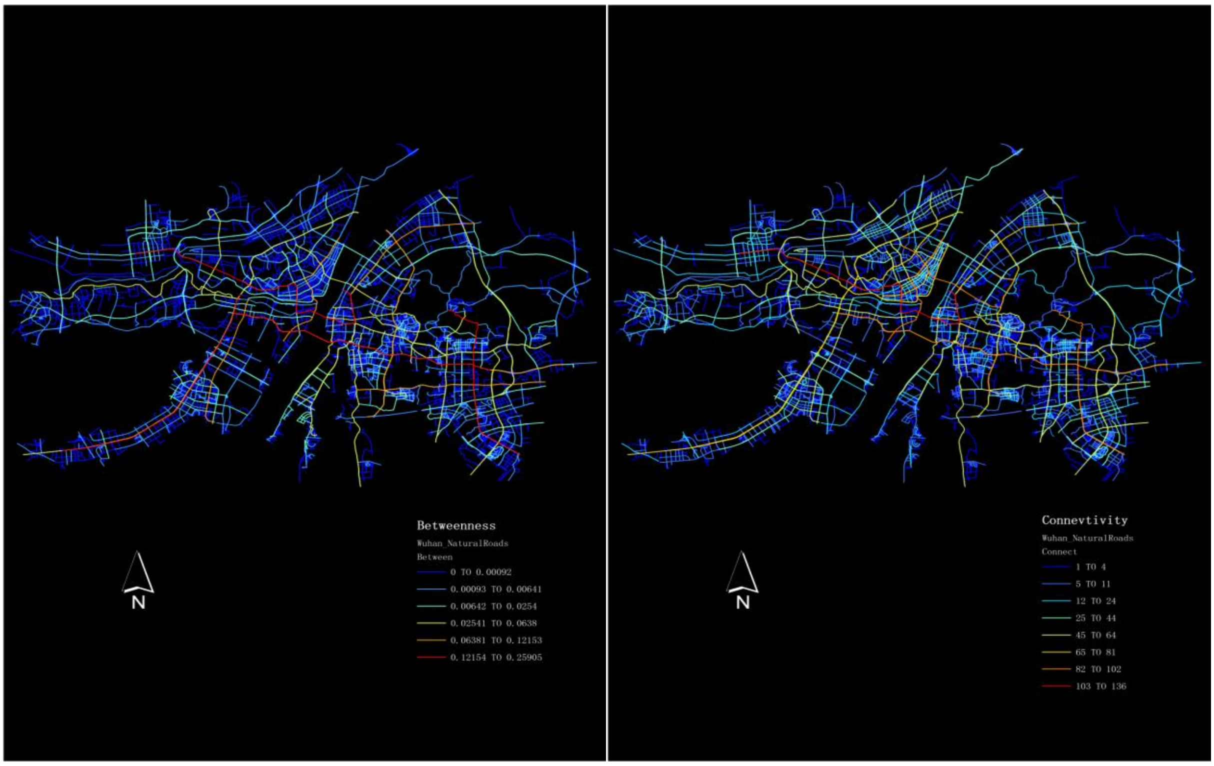


Figure B 6. Hierarchical pattern of Wuhan named streets based on betweenness and connectivity