# STRUCTURAL HOLE ANALYSIS FOR STRUCTURING HIERARCHICAL ROAD NETWORKS

Hong Zhang
oceanzhhd@gmail.com

Zhilin Li*
lszlli@polyu.edu.hk

Department of Land Surveying and Geo-Informatics
The Hong Kong Polytechnic University
Hung Hom, Kowloon, Hong Kong

## Abstract

Studies on the structural properties of the road network have received intensive interdisciplinary attentions, and its close relationship with traffic flow has been widespread received. However, most of these attempts were theoretical and further validation is needed. It is also a challenge to understand how the structure and morphology of a road affect its accommodation of traffic flow. In this study, the structural hole theory rooted in social science was introduced to rank roads in a road network. A new measurement named centrality rank ($CR$) is built to assign an order value to each node in a complex network based on possible structural holes in the ego network. A set of real-life data is used for evaluation and result is in accordance with those obtained by other researchers from practical data. In other words, this method works well both in theory and practice.

**Keywords**: Structural Hole; Road network; Stroke; Hierarchy

## 1. Introduction

Network is ubiquitous in nature and society. Among them, road network has been a central subject and has received intensive interdisciplinary attentions. For instance, mathematicians and physicists focused on the statistical regularities that most road networks seem to share (Kalapala, et al., 2006; Volchenkov & Blanchard 2007; Wagner, 2008; Ferber, et al., 2009), while researchers from computer and information sciences tried to explore the spatial patterns and spatial knowledge of road networks (Rosvall et al., 2005; Thomson, 2006; Tomko, et al., 2008; Barthélemy & Flammini 2008). In geographic science, a number of studies on road network have been carried out, including graph representation, pattern detection, property exploration, map generalization, route optimization and behavior analysis (Mackaness, 1995; Li & Choi 2002; Salingaros, 2003; Hillier & Iida 2005; Marshall, 2005; Xu & Sui 2007). Amongst all the above attempts, the detection of the hierarchical structure of a road network has nowadays attracted more and more attentions. Fruitful results on this topic can be found in many fields. For instance, some measures have been developed to define the hierarchies of roads in a road network including entropy, local integration, degree (or connectivity) and flow dimension and flow capacity (Sukhov, 1970; Hillier & Hanson 1984; Bjørke, 2003; Jiang, 2008a). However, most of these attempts were theoretical and further validations are still needed. In addition, how to define the orders of roads in a road network from the aspect of transfer of flow is still a challenge.

This study aims to deal with the above challenge. In this study, the structural hole analysis rooted in social science is firstly introduced to rank roads in a road network. It describes the status of each node from the aspect of transfer of flow in its ego network. The reminder of this paper is organized as follows, section 2 introduces the principles and mathematics of structural hole theory in social science; section 3 demonstrates how to apply the structural hole analysis to a road network; section 4 uses an actual road network to evaluate this approach; some discussions and conclusions are given in section 5 and 6 respectively.

## 2. Structural hole: principles and mathematical formulation

Social sciences focus on structure and conceptualize social structure as a network of social ties (Nooy, et al., 2005). Sociologists either examine the structure of the entire social group, or turn to the position of each individual in the local network. Structural hole analysis belongs to the latter.

### *2.1 Structural hole and ego network*

The structural hole theory developed by Burt (1995), is one of the most efficient ways to define the positional status of each node in its ego network (a ego network is defined as a road network consisting of a single actor (ego) together with the actors they are directly connected to (or alters) and all the links among them). The structural hole theory believes that in a social network, the individual's advantage or power is based on his or her control over the spread of information, goods or services between his or her immediate neighbors, and the absence of a tie between either ego or alter and other alters would induce a structural hole. Here we use three simple ego networks to demonstrate the principles of structural hole analysis (see Figure 1).

It is obvious that Figure1 (a) is a complete ego-network where ego, alter1 and alter2 can directly communicate with each other, while both Figure 1(b) and Figure 1(c) have structural holes corresponding to different conditions. Or in other words, in Figure 1(b), the structural hole is caused by the lack of linkage between the immediate neighbors of ego, i.e., alter1 and alter2, while in figure 1(c), the structural hole lies in the lack of linkage between ego and other alters (alter2 in this case).

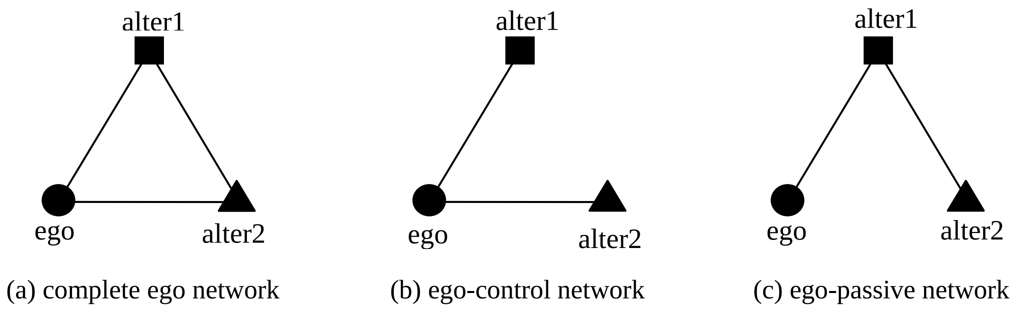



Figure1. Three forms of ego networks

***2.2 Constraint and aggregate constraint***

According to the structural hole theory, the more structural holes in the ego network of a node, the less constraint the node obtains from its immediate neighbors, and the high opportunity for the node to spread of information. In social science, two measures, i.e. constraint and aggregate constraint have been developed to describe structural holes.

To better understand the mathematics of constraint and aggregate constraint, some basic definitions and notions of graph theory worth to be described. Mathematically, a graph is formed by a set of vertices (or nodes) and a set of edges (or links) that connected pairs of vertices. A graph can be directed or undirected, weighted or unweighted, complete or incomplete, and connected or unconnected. In order to quantitatively describe the structural holes of any given node, the links between pairs of actors need to be directed and weighted (see figure 2).

The weight of each link ( $p_{ij}$ ) (also called link strength or proportional strength) from node $i$ to all its immediate neighboring nodes can be defined as the reciprocal of the degree (or connectivity) ( $k$ ) of node $i$ . Mathematically

$$p_{ij} = \frac{1}{k_i} \qquad \left( j \in i_{ne} \right) \tag{1}$$

Where, $ne$ is the immediate neighbors of the node $i$. For instance, in Figure 2(a), the ego is connected to both alter1 and alter2, so its degree of connectivity is 2. The strengths (weights) of links from this ego to alter1 and to alter2 are both 1/2=0.5. In Figure 2(b), however, the degree of alter1 is only one, thus the weight of the link from alter1 to ego is 1 (=1/1) although the weight of the link from ego to alter1 is 0.5.

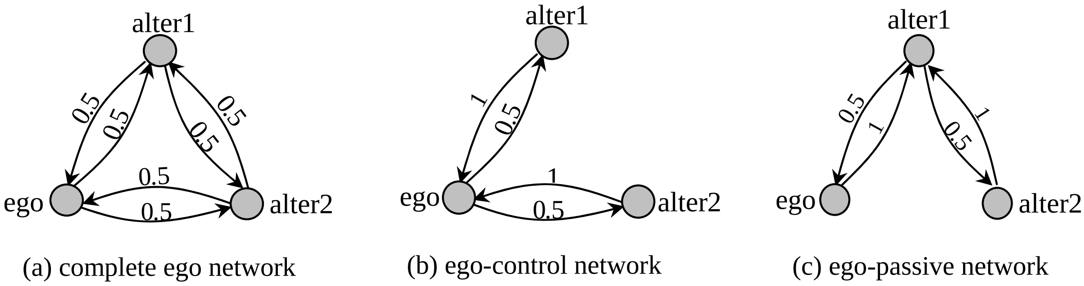



Figure 2. Graphs and proportional strengths of three kinds of ego networks

If any two nodes (node $j$ and node $q$), which are neighbors ($ne$) of node $i$, are directly connected, then the indirect link strength ($p'_{ij}$) from node $i$ to node $j$ is defined as:

$$p'_{ij} = p_{iq} p_{qj} \qquad \left( j \in i_{ne}, q \in i_{ne}, q \neq j, q \neq i \right) \tag{2}$$

This is like the relationship between the ego and the alter2 via the alter1 as shown in Figure 2(a). In other words, in this case, the strength of the indirect link from ego to alter2 is 1×0.5=0.5. In this case, nodes $i$, $j$ and $q$ form a triangle.

The constraint ($C_{ij}$) of node $i$ by node $j$ is computed by the squares of the sum of the direct link strength and the indirect link strength from node $i$ to node $j$ (Burt, 1995):

$$C_{ij} = \left( p_{ij} + \sum p'_{ij} \right)^2 = \left( p_{ij} + \sum_q p_{iq} p_{qj} \right)^2 \qquad \left( j \in i_{ne}, q \in i_{ne}, q \neq i, q \neq j \right) \tag{3}$$

The $C_{ij}$ value reveals the constraint of node $i$ by node $j$. The larger the $C$ value, the larger the constraint over node $i$, the smaller the opportunity for node $i$ for spread of information. In the case of Figure 2(a), the constraint from ego to alter1 is $(0.5+0.5*0.5)^2=0.5625$. On the other hand, the $C$ values for the links from ego to alter1 for Figures 2(b) and 2(c) are respectively 0.25(=$(0.5)^2$) and 1(=$(1)^2$). This shows that in these three Figures, the ego has the largest control over alter1 in Figure 2(b) (i.e. the ego-control network) and most constrained by the alter1 in Figure 2(c), (i.e. the ego-passive network).

The constraint is for analysis at the link level (e.g. ego to alter). This is not enough because there might be more than one link connected to a node (e.g. ego). Therefore, a new measurement, called aggregate constraint ($AC$), needs to be introduced to reflect the status of each node $i$. $AC_i$ is defined as the sum of the constraint of node $i$ by all its immediate neighbors.

$$AC_i=\sum_s C_{is} \qquad \left(s\in i_{ne}, s\neq i\right) \tag{4}$$

The $AC_i$ value reveals the constraint of node $i$ by all its neighboring nodes. The larger the $AC_i$ value, the larger the constraint over node $i$, the smaller the opportunity for node $i$. The $AC$ values for the ego in the Figure 1 are respectively 1.125 (=0.5625+0.5625), 0.5 (=$(0.5)^2+(0.5)^2$) and 1(=$(1)^2$). The results reveal that the ego in Figure 2(b) is least constrained by its neighbors and has most opportunity to spread of information.

***2.3 Centrality Rank***

It can be noted here that a large aggregate constraint ($AC$) means a small opportunity for the given ego. This is inconvenient to use. There a new measure named Centrality Rank ($CR$) is introduced to make the relationship more intuitive. The centrality rank is defined as follows:

$$CR_i=\frac{1}{AC_i}=\frac{1}{\sum_s C_{is}} \qquad \left(s\in i_{ne}, s\neq i\right) \tag{5}$$

According to the above equations, it is easy to obtain that the higher the $CR$, the more important the node in transfer of flow in the entire network.

Since $CR$ focuses on structural holes instead of centre and periphery, it provides a different perspective for structuring network and has some advantages over traditional network measures. Taking the degree (or connectivity) as an example, the degree of ego in Figure 1(a) and Figure 1(b) is both 2 although the structures are quite different. This is also the case if the clustering coefficient is considered. For example, the clustering coefficient of the ego in Figure 1(b) and Figure 1(c) is both 0, while centrality rank is able to differentiate these different structures and thus will be applied to the structure of hierarchy for a complex network, which is to be discussed in section 3.

It should be noted that figure 1 is very simple, real life networks are much complex than it. Therefore, it is necessary to decompose a complex network to a number of ego networks by treating each node in the network in turn as an ego and all its immediate neighbors as alters.

## 3. Structural hole for structuring road network

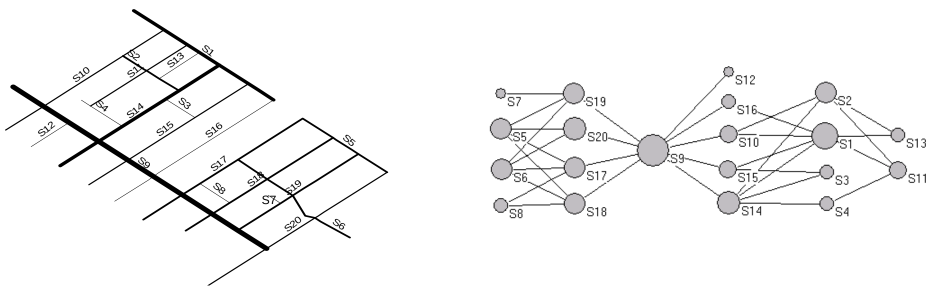
The previous section introduces the definition of centrality rank, this section applies it to a real life road networks–the sampled San Francisco road network, which is very regular and grid-like.

It seems clear that the most important step is to form a natural road from road segments because the topologic structure of segment-based road network is quite tedious. As mentioned in section 1, various approaches have been developed for representing a road network. In this study, natural road (or stroke in terms of Thomson (2003)) is used to represent a road network.

As centrality rank is for node (ego) in an ego network but one wants to use centrality rank to weigh the importance of strokes, it is necessary to construct a connectivity graph of the stroke-based road network after deriving strokes in a road network. In such a graph, strokes need to be represented by nodes and thus the intersections represented by links, In this way, a centrality rank for each stroke in the network (i.e. each node of the graph) can be computed. The stroke representations and their corresponding connectivity graphs are illustrated in Figure 3.

From these connectivity graphs, centrality rank of all strokes (i.e. nodes in Figures 3(b) is computed and visualized by size. In Figures 3(a), the thicker the line, the higher the centrality rank of the stroke. In Figure 3(b), the larger the node, the higher the centrality rank of the corresponding stroke.

It is illustrated that in Figure 3(a), stroke $S_9$ has the highest centrality rank ($CR$) value, followed by stroke $S_1$ and stroke $S_{14}$. According to the definition of equation (4) and equation (5), the high $CR$ values for these three strokes imply that: (1) there are many structural holes in the immediate neighbors of the three strokes; (2) the three strokes play important roles in transfer of flow in the entire network. These inferences are intuitively represented in Figure 3(b), where none of the immediate neighbors of stroke $S_9$, or $S_1$ or $S_{14}$ is directly connected with each other. The three strokes are the backbones to hold the integrity of the entire network. Indeed, other strokes all rely on these three strokes to communicate with each other. For instance, the left part of the network relies on stroke $S_9$ to reach the right part of the network, and vice versa. If we only look at the left part, strokes $S_1$ and $S_{14}$ serve as transferring ports between its neighbors. By contrast, all the ´dangling´ strokes (i.e. a stroke that is only directly connected to one stroke), such as stroke $S_7$ and stroke $S_{12}$ have the lowest $CR$ values, which mean there are few structural holes in their ego networks and they tend to be strongly constrained by their neighbors.




(a) Sampled road network Of San Francisco

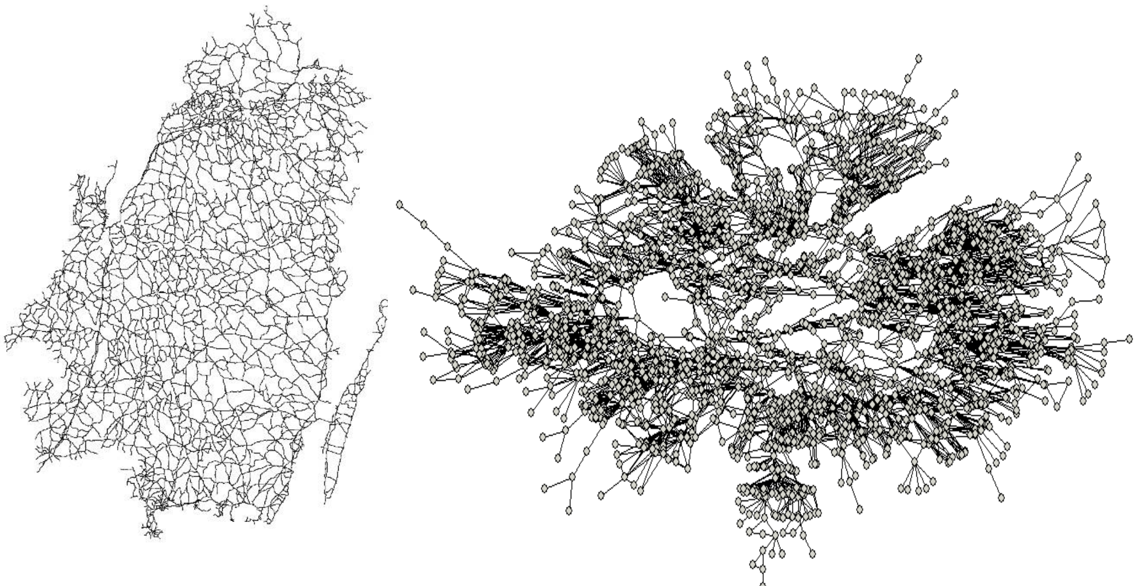
(b) Connectivity graph of (a)

Figure 3. Two real life road networks and their connectivity graphs

This section demonstrates how structural hole analysis can help define the orders of roads in a road network. It should be noted that some physical properties of the road itself including road width, number of lanes, built hierarchy (e.g. national highway, provincial highway and road etc.), and location of traffic lights, are not considered in this study.

**4. Evaluation of structural hole for structuring hierarchical road networks**

To evaluate the applicability of structural hole analysis in structuring hierarchical road network, some real-life data should be used for experimental testing. The major dataset involved in this study is the Sydost highway network (see Figure 4(a)) and its traffic flow data. The road segments are first connected into strokes and the traffic flow of roads are used as a benchmark. Such a use is based on the assumption that roads with higher orders in terms of centrality rank tend to be more important for the entire network and thus will accommodate higher traffic flow.  The connectivity graph of stroke is shown in Figure 4(b).



(a) Sydost road network (b) Connectivity graph of Sydost road network

Figure 4. The Sydost highway network and its connectivity graph

After ranking strokes in terms of centrality rank, we respectively selected the top 1%, top 5%, top 10%, top 15% …… top 100% strokes in terms of centrality rank values, and computed the corresponding percentages of the traffic flow accommodated by the selected strokes to the total traffic flow of the entire road network. It is found out that the top 1%, top 5%, top 10% and top 20% strokes respectively accommodate 30%, 50%, 65% and 80% traffic flow. This finding is consistent with Jiang (2008b), i.e., in a road network, there are top 1% strokes that accommodate nearly 20% traffic flow, and top 20% strokes that accommodate 80% traffic flow.

## 5. Conclusion

In this article, we have introduced the concept of structural hole analysis for structuring hierarchical road network and built a new measurement, named centrality rank, for assigning an order value to each road in a network. The effectiveness ands stability of centrality rank had been evaluated by a real-life data -- the Sydost highway network. It is found out that the results are quite consistent with practical data.

It is hoped that the ideas and methods presented here will prove useful in the analysis of many other types of urban networks. Possible applications include map generalization, park location, population migration, and daily communication.

## Acknowledgement

This research is supported by the Hong Kong Polytechnic University and RGC of HK (PolyU5221/07E). This research is supported by the Hong Kong Polytechnic University. The data about Sydost highway network is provided by Jiang Bin, and the San Francisco sampled road network is obtained from TIGER data of U.S.Census Bureau (http://www.census.gov/geo/www/tiger/).

## References

Barthélemy, M. & Flammini, A., 2008. Modeling urban street patterns. *Physical Review Letters*, 100(13), pp. 138702.01–138702.04.

Bjørke, J.T., 2003. Generalization of road networks for mobile map services: an information theoretic approach. In *Proceedings of the 21st International Cartographic Conference (ICC)*. Durban, South Africa 10-16 August 2003.

Burt, R.S., 1995. *Structual Holes, The social structure of competition*. 1st ed. Cambridge, Massachusetts, and London, England: Harvard University Press.

Ferber, C.V., Holovatch,T., Holovatch, Y. & Palchykov,V., 2009. Public transport networks: empirical analysis and modeling. *The European Physical Journal B*, 68, pp. 261–275.

Hillier, B. & Hanson, J., 1984. *The social logic of space*. Cambridge: Cambridge University Press.

Hillier, B.& Iida, S., 2005. Network and psychological effects in urban movement, In

COSIT (*Proceedings of the International Conference on Spatial Information Theory*). Elliotville, USA 14-18 September 2005.

Jiang, B., 2008a. Flow dimension and capacity for structuring urban street networks. *Physica A:Statistical Mechanics and its Applications*, 387, pp. 4440–4452.

Jiang, B., 2008b. Street hierarchies: a minority of streets account for a majority of traffic flow. *International Journal of Geographical Information Science,* [Online], 16 September, Available at: http://www.arxiv.org/abs/0802.1284. [Accessed 10th July 2009].

Kalapala, V., Sanwalani, V., Clauset, A. & Moore, C., 2006. Scale invariance in road networks. *Physical Review E*, 73, pp.026130-1–026130-6.

Li, Z.L. & Choi, Y. H., 2002. Topographic map generalization: Association of road eliminationwith thematic attributes. *The Cartographic Journal*, 39, pp. 153–166.

Mackaness, W.A. 1995 Analysis of Urban Road Networks to Support Cartographic Generalization, *Cartography and Geographic Information Systems*, 22(4): 306-316.

Marshall, S., 2005. *Streets & patterns*. London and New York: Spon Press.

Nooy, W. D., Mrvar, A. & Batagelj, V., 2005. *Exploratory social network aAnalysis with Pajek*. New York: Cambridge University Press.

Rosvall, M., Trusina, A., Minnhagen, P. & Sneppen, K., 2005. Neworks and cities: an information perspective. *Physical Review Letters*, 94(2), pp.028701-1–028701-4.

Salingaros, N., 2003. Connecting the fractal city, Keynote speech, In *5th Biennal of towns and town planners in Europe*. Barcelona, Spain 10-12 April 2003.

Sukhov, V.I., 1970. Application of information theory in generalization of map contents. *International Yearbook of Cartography*, X, pp.41–47.

Thomson, R.C., 2003. Bending the axial line: smoothly continuous road centre-line segments as a basis for road network analysis. In *Proceedings of the 4th Space Syntax International Symposium*. London, United Kingdom 17-19 June 2003.

Thomson, R.C., 2006. The ´stroke´ concept in geographic network: generalization and analysis. In *12th International Symposium on Spatial Data Handling*. Vienna, Australia 10–12 July 2006.

Tomko, M., Winter, S. & Claramunt, C., 2008. Experiential hierarchies of streets. *Computers, Environment and Urban Systems*, 32, pp.41–52.

Volchenkov, D. & Blanchard, PH., 2007. Random walks along the streets and canals in compact cities: spectral analysis, dynamical modularity, information, and statistical mechanics. *Physical Review E*, 75, pp.026104-1–026104-14.

Wagner, R., 2008. On the metric, topological and functional structures of urban networks. *Physica A*, 287, pp. 2120–2132.

Xu, Z.W. & Sui, D.Z., 2007. Small-world characteristics on transportation networks: a perspective from network autocorrelation. *Journal of Geographical Systems*, 9, pp.189–205.