A SOCIAL NETWORK ANALYSIS APPROACH TO ANALYZE ROAD NETWORKS

Kyoungjin Park
Alper Yilmaz
Photogrammetric and Computer Vision Lab
Ohio State University
park.764@osu.edu
yilmaz.15@osu.edu

ABSTRACT

Depending on relations and transient properties, real world events or physical settings can be represented with different types of network structures. To realize this purpose, a social network topology becomes a complex and a scale free network. In this paper, we propose to use this topology to represent road networks. Generally speaking, road networks and their spatial relations reside on a plane, which generates a specific network structure termed as a planar network. This observations result in the degree distributions in road networks which contain a considerably different characteristic compared to other types of networks. In order to analyze these special networks, we introduce a novel approach using the *centrality* and *entropy* of various distributions estimated from the network topology. In context of sociology, the centrality is used for find an important actor in a social setting. The entropy of these centrality distributions provides a way to examine the characteristics of selected road networks at city, county or state levels. We conjecture that the proposed tools will help transportation experts to gather in depth information on existing road networks. Our experiments prove a concept design, which can be used to quantitatively compare residential area and downtown area road networks.

INTRODUCTION

The network representations have been commonly used in different problem domains. An Internet network and a web site link network are among many common examples, which describe how systems are connected to each other. In Biology, and chemistry use network to explain how proteins and atoms are interacts with each other (Wu *et al.*, 2006). In the field of sociology, researchers have long considered social networks to represent relationships between actors in a social setting (Freeman 1979; Knoke *et al.*, 2007; Wasserman *et al.*,1994). These representations have served as a strong tool to conduct research analyzing various technological or social phenomena. Recently, network representations have been used to analyze patterns in a street network in the field of geography, and geosciences. Particularly, they are used to explain the growth of cities or urban area (Masucci *et al.*, 2004; Ji *et al.*, 2008; Cardillo *et al.*, 2006). These studies, unlike other fields, do not adapt social network analysis methods and do not consider importance of intersections, relations between roads for resolving transportation problems.

In this paper, we proposed a novel network analysis approach, which exploits social network analysis for analysis of road networks. The topology of a road network is comparably different from other common types of networks, and provides a planar topology where the links between the nodes do not intersect. In addition, the degree of a node in a road network is less than a typical social or Internet network.

We introduce two main concepts in analysis of road network. The first one is the centrality of nodes in a road network, which provides a distribution. The centrality is widely used to detect important nodes and find nodal characteristics in networks. As it is proposed by Freeman (Freeman 1979), we compute three types of centralities; degree, closeness, and betweenness. Generally, the centralities explain how nodes play an important role in a social network structure. In a road network, degree of a node can be provides connectivity and popularity of an intersection with respect to spatially neighboring intersections. In similar fashion, the closeness and the betweenness of a node (intersection) provide alternative routes with in the network.

The second concept we introduce to the road network analysis is the entropy of distributions computed from the network. Entropy is generally used to explain the uncertainty from a probability distribution and is commonly used in information coding theory. In our context, we compute the graph entropy to measure information encoded in the distributions generated from the road networks. We suggest interested reader to read a through surveyed in (Dehmer 2008; Riis 2007; Simonyi 1995) for a mathematical discussion on graph entropy. It is not uncommon to come across articles that do not clearly discuss how nodal distributions are computed and how entropy is evaluated. In addition, the entropy definitions are commonly varied across different application domains. In this study, we examined the graph entropy for three different centrality distributions.

The remainder of the paper is organized as follows: Section 2 introduces social network analysis methods for node centralities. Also, the concept of entropy applies in information coding theory is also presented. In section 3, experimental results are shown. Finally, we conclude in Section 4.

METHODOLOGY

This section discusses construction of the road network, estimation of centrality measurement and computation of entropy as it is used in information coding theory. For road network analysis, we first construct a road network from maps, and compute all three nodal centralities. Finally, we estimate the entropies from the distributions of centralities computed from the entire network.

Construction of the Road Network

We extracted the road networks selected sites using the Google Map™ by selecting the intersections and junctions between the roads as nodes of the network, and road segments between these intersections as the links in the network. We, particularly, chose four different sites to perform our analysis. Two of these sites are selected from the downtown areas and two sites are selected from residential areas. All the sites have same square foot coverage and the resulting network is an undirected and unweighted graph, which generates a network with a symmetric adjacency matrix. More formally, let G be a graph of a road network. $G = (V, E)$. Where V is a set of nodes of a graph, and E is a set of edges in G. And let **A** be an adjacent matrix of road network G. Then adjacent matrix **A** has the following form

$$A = \begin{cases} a_{ij} = a_{ji} & \text{if node } v_i \text{ and node } v_j \text{ has a link} \\ 0 & \text{otherwise} \end{cases}.$$

Node Centrality

One of the main goals in social network analysis is to find the most important node in the network structure. Many of algorithms for finding important node have been studied (Freeman 1979; Knoke *et al.,* 2007; Wasserman *et al.* 1994). Centrality is one of the measurements to rank nodes. Freeman developed most widely used three centrality concepts: Degree, closeness, and betweenness.

- Degree

  Degree indicates the connectivity of nodes. In particular, it provides information on how many other nodes are connected with a particular node. Since the number of degree of a node in a network counts the number of edges directly linked with that node, the degree centrality is considered as a centrality of local measurement. In road network application, the large number of degree means how many ways are linked at a junction point. It may imply that higher degree of nodes could have crowded traffic at those points than lower degree nodes.

$$C_{Degree}(N_i) = \sum_{j=1}^{n} a_{ij}$$

- Closeness

  In social network, closeness indicates how a node is close is to the other node. Thus, the closeness is computed as what is the shortest geodetic path between two nodes. Since closeness finds the shortest path in the whole network structure, it considers the global connectivity of network structure.

$$C_{Closeness}(N_i) = \frac{1}{\sum_{j=1}^{n} d(N_i, N_j)},$$

  where $d(N_i, N_j)$ is a geodetic distance between N_i, and N_j. Since closeness is global centrality, the closeness is affected the number of the nodes in a whole network. Thus, we used the normalized closeness for the experiments.

- Betweenness
  Betweenness explains how a node can control the other nodes which have no direct connectivity between them. It counts how many times a node intervene to connect the other nodes. It is also one of global measurement of centrality, because betweenness investigate the whole network to find the connecting path.

$$C_{Betweennes\ s}(N_i) = \sum_{j<k} \frac{g_{jk}(N_i)}{g_{jk}}$$

where, g_{jk} is the number of geodesic paths between two nodes N_j and N_k, and $g_{jk}(N_i)$ is the number of geodesic between the N_j and N_k that contain node N_i. Likewise closeness, betweenness is also global centrality, so normalized betweenness is used for the experiments. In a road network application, betweenness tells how the intersection points are important to reach the destination from the start points.

Entropy

Entropy is a quantitative measurement used to explain the probability distributions. That is an important concept in the information coding theory to be used for measurement the probability distribution. If the probability has uniform distribution, it is called the uncertainty of the distribution is uniform, such that, the states of system are highly disordered. In such cases the entropy of the probability distribution increases. On the contrary, when the probability distribution is not uniform distribution, so some of state could be predictable. It means the probability of predictable states has higher probability than the others. In this case, the state of order is high, and the entropy goes down to the lower. Therefore entropy is helpful tool to describe the state of order in a system. Let *P* be the probability distribution on the node set of *V(G)*, and $p_i \in [0,1]$. The entropy of the graph G is

$$H(G,P) = \sum_{i=1}^{N} p_i \log_2(p_i)$$

In the previous phase, we computed three types of centralities, and we earned the histogram of those centralities. The histogram is directly converted into the probability distribution, and entropy of the each centrality was computed.

EXPERIMENT

For the road network structure comparison, we selected four sites from two types of settings. One study area is Columbus Ohio and the other study area is Washington D.C area. From those areas, we selected two sites, one is in a downtown area and the other one is in a residential site. The graphs for the selected sites are generated from the Google Map™. Figure 1 shows the network topology of selected sites. As can be observed the downtown area has a nearly grid structure, while the residential area is more circular.

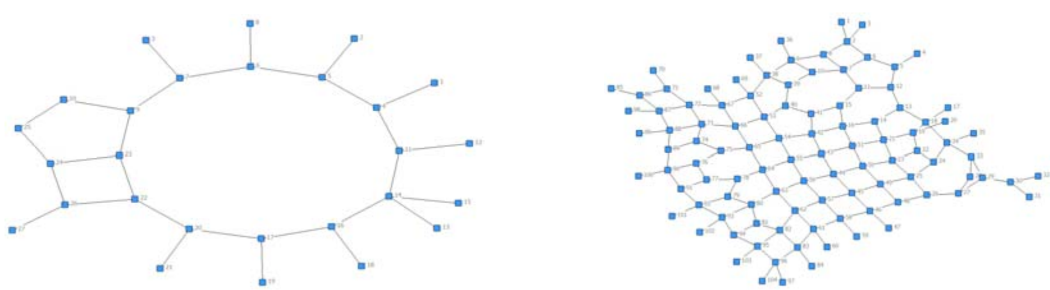

Figure 1. (a) Network of residential area of Columbus Ohio, (b) Network of downtown area of Columbus Ohio.

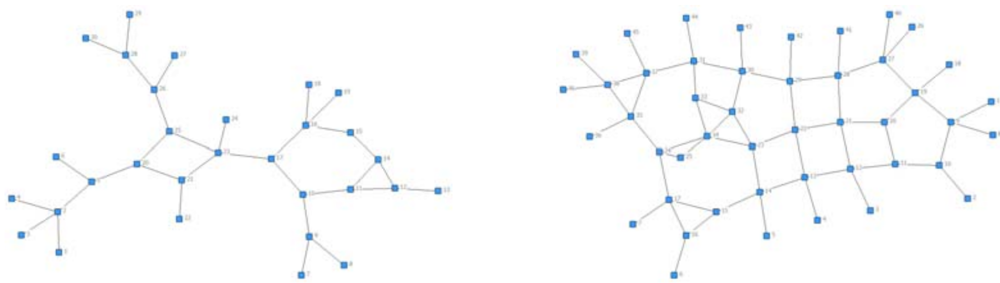

Figure 2. (a) Network of residential area of Washington D.C. (b) Network of downtown area of Washington D.C.

Table 1 shows constructed network properties, and it tells the number of nodes and edges in a network is different even selected site has same coverage of area. The ratio of vertices to edges in residential area is larger than that of in downtown area. Because, the downtown has nearly gird, so many of the nodal degree is four.

Table 1. Graph properties of selected study sites

	Number of Nodes	Number of Links	Ratio of nodes to Links
Columbus Downtown Area	104	152	0.68
Columbus Residential Area	27	29	0.93
Washington D.C Downtown Area	46	60	0.76
Washington D.C Resident Area	30	32	0.94

Centrality of the Networks

We found that the degree range in the above network is one to five. The degree of the nodes could not be expanded like as other types of complex network. Because that road network is embedded into 2D plane, it has planar network property. The histogram of degree in figure 3 shows that the downtown area tends to have more nodal degree of four than the residential area. However in a view of probability distribution, degree of one and degree of three in Columbus residential area is similar. And in case of Washington D.C. downtown area, the probability distribution is close in case of degree.

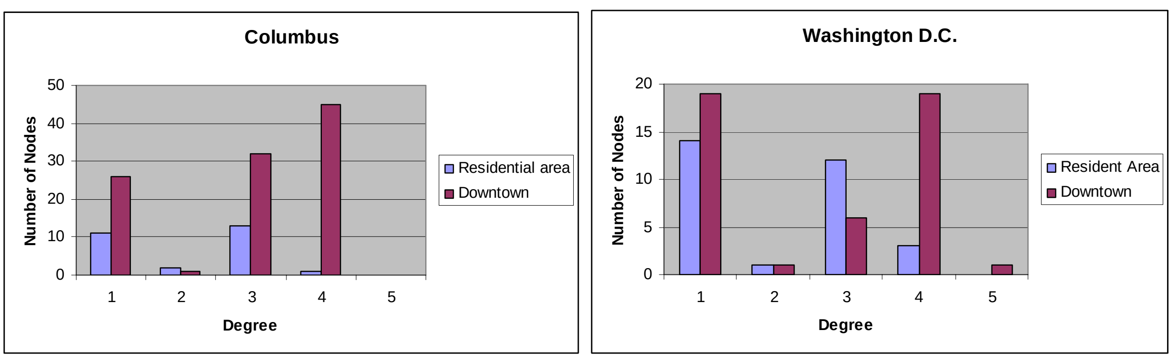

Figure 3. Histogram of Degree.

From the above distribution, degree entropy is computed and shown in figure 4. This figure describes that the entropy of degree distribution dose not discriminate different types of road network topologies. This is cased by the range of the degree is too narrow. So we experiment the entropy with the closeness centrality distribution and betweenness centrality distribution. Figure 5 and figure 6 show the examination results.

Figure 5 shows how the closeness is distributed. The closeness in a road network indicates how far from one node to the other nodes. So the larger value of closeness means that a node dose not have a short cut path to arrive to the other nodes. The study sample in Columbus area explains well that the downtown area has more alternative shortest routes to than the residential area. That implies grid-like network topology generate circulated routes, and it has shorter geodesic paths than residential area.

Figure 6 shows the probability of distribution of betweenness. Betweenness indicates the controlling power between other nodes which are not connected. Thus the higher value of betweenness means that the node plays an important role to connect the other two nodes. On the contrary, a node having smaller value of betweenness implies it is less important to connect other two nodes. So we assume that those nodes could be a leaf node a network. Graph of figure 6 of Columbus downtown area shows that downtown area has grid type of road networks, so there are more alternative ways to reach to the other node than residential area.

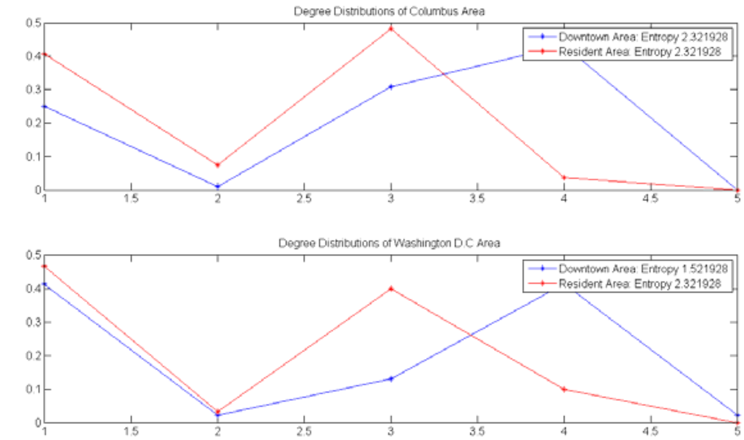

Figure 4. Probability distribution of degree in study area and its entropy.

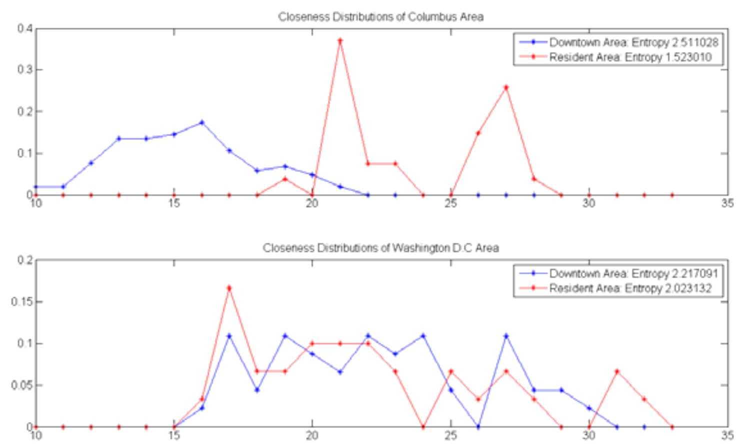

Figure 6. Probability distribution of closeness and its entropy.

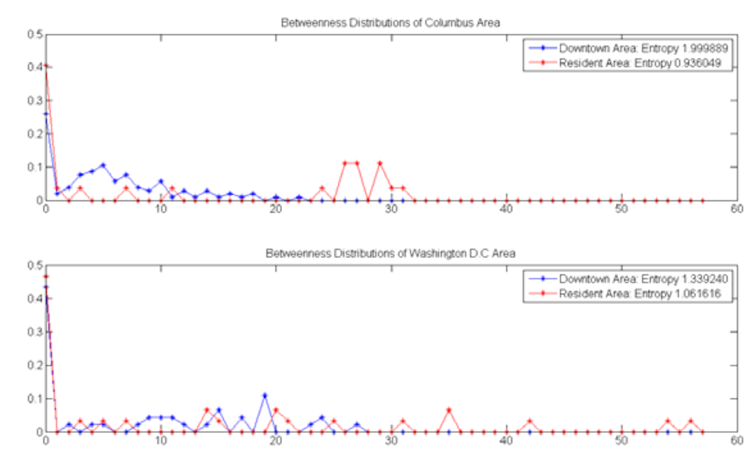

Figure 5. Probability distribution of betweenness and its entropy.

CONCLUSION

We propose road network analysis applying social network analysis methods which are widely used in the other areas. First we compute the three types of centralities, degree centrality, closeness centrality, and betweenness centralities. Those node centralities describe how a node important in a network, and we extract the histogram of those centrality values respectively. By comparing the distribution of node centrality in downtown area and residential area, this work showed that the degree distribution dose not have distinction between downtown and residential area. Because road network is planar network so the range of nodal degree is narrow, so the entropy of the degree distribution dose not have difference between downtown area and residential area. However the distribution of closeness and betweenness show difference of downtown area and residential area. In a downtown area which having grid-like network topology has higher entropy than the residential area having radiant network topology. In this work we examine the distribution properties in unweighted and undirected network structure. We conjecture that the proposed analysis tool can be extended to include other types of information related to transportation such as traffic flows, accidents, and number of cars on a certain road. We will represent this information as weighted graph, and provide detail report for future road intersection planning.

REFERENCE

Cardillo, A., S. Scellato, V. Latora, and S. Porta, 2006. Structural properties of planar graphs of urban street patterns. *Physical Review E 73* 066107.

Dehmer, M., 2008. Information processing in complex networks: Graph entropy and information functionals. *Applied Mathematics and Computation 201*, 1-2, pp82 – 94.

Freeman, L. 1979. Centrality in social networks: Conceptual clarification. *Social Networks 1*, 3 pp215–239.

Ji, L., W. Bing-hong, W. Wen-xu, and Z. Tao, 2008. Network entropy based on topology configuration and its computation to random networks. *Chinese Physics Letters 25*, 11, 4177.

Knoke, D., and S. Yang, 2007. *Social Network Analysis (Quantitative Applications in the Social Sciences)*, 2nd ed. Sage Publications, Inc, November.

Masucci, P., D. Smith, A. Crooks, and M. Batty, 2004. Random planar graphs and the London street network.

Riis, S., 2007. Graph entropy, network coding and guessing games.

Shetty, J., and J. Adibi, J., 2005. Discovering important nodes through graph entropy the case of enron email database. In *LinkKDD '05: Proceedings of the 3rd international workshop on Link discovery* (New York, NY, USA, 2005), ACM, pp. 74–81.

Simonyi, G., 1995. Graph entropy: a survey. In *Combinatorial Optimization*, DIMACS Series in Discrete Math. and Theoretical Comp. Sci., Vol. 20 (W. Cook, L. Lovász and P. D. Seymour, eds.), pp. 399-441.

Wasserman, S., K. Faust, and D. Iacobucci, 1994. *Social Network Analysis : Methods and Applications(Structural Analysis in the Social Sciences)*. Cambridge University Press, November.

Wu, D., and X. Hu, 2006. Mining and analyzing the topological structure of protein-protein interaction networks. In *SAC '06: Proceedings of the 2006 ACM symposium on Applied computing* (New York, NY, USA, 2006), ACM, pp.185–189.