Uncovering Correlations Between Urban Road Network Centrality and Human Mobility

Yury Halavachou

Department of Railway Transportation Control
Belarusian State University of Transport
34, Kirova Street, Gomel, 246653, Republic of
Belarus
E-mail: yurafromgomel@gmail.com

Li Jinwei
Daihatsu Motor Co., Ltd
E-mail: vjo16252@gmail.com

Abstract—**Urban planners have been long interested in understanding how urban structure and activities are mutually influenced. What are the forces that lead to the current patens of road network and what is the influence of the latter's structure on urban activities? Network centrality measures have been traditionally utilized to uncover the structural properties of urban street networks. In this paper we are interested in examining the correlation between the centrality of street network and the intensity of human movement in urban areas. We focus on two cities and we utilize a dataset of geo-tagged tweets that can form a proxy to urban mobility and the corresponding street networks as obtained from OpenStreeMap. Our results indicate that different centrality metrics have different levels of correlations with the intensity of human movement. Furthermore, the strength of the correlation varies in the two cities examined.**

Keywords-Human Mobility; Urban Street Network; Network Centrality; OpenStreetMap

I. INTRODUCTION

Urban spaces are typically highly localized but they are globally connected [1]. In particular, the urban space consists of local patchwork, which serve some specific functionality. Nevertheless, these patchwork are linked by the urban street network into a whole at a global scale. While the structure of urban space is greatly influenced by the history of each city [2], researchers have long been analyzing its properties in order to facilitate planning functionalities, such as resource allocation and transportation planning. Human activities in urban environments, such as business and travel, are often shaped and constrained by the geographical distance to and accessibility of the resources.

The urban street network functioning as the backbone of urban space. plays a vital role in connecting urban neighborhoods together and supporting the local/global movement in/between urban areas. Its structural properties, such as centrality and accessibility, can reveal many implications on human activities. Centrality [3], which is a network-based metric measuring the structural. importance of nodes in complex networks is often utilized to capture the importance of different parts of road networks. Former studies [4, 5] indicate that the structural properties of urban road networks as captured by the betweenness centrality can explain the observed traffic flow. Another form of centrality,

closeness centrality is shown to be highly correlated with the intensity of economic activities [6] and land use [7]. Furthermore, the aggregated human travel flow on streets is shown through simulations to be mainly shaped by the underlying street structure [8].

In this work, we conduct a study on the correlation between the centrality of the urban street network and the intensity of human movement over it using data from Pittsburgh and NYC. Our results imply that different centrality metrics correlate with the intensity of human movement at different levels. The correlation strength further differs in the two cities examined.

Roadmap: Section 2 describes our analysis set up and the dataset. Section 3 presents and analyzes our experimental results, while Section 4 concludes by also discussing the future work.

II. EXPERIMENTAL SETUP

In this section we will introduce the network structures that capture the intensity of human movements and the urban road network as well as the data that drive their realizations in Pittsburgh and NYC.

A. Human Transition Network

In the human transition network $G_T = (U, E)$, the set of nodes U is a collection of non-overlapping areas/neighborhoods in the city under examination. Further, a directed edge e_{ij} between two areas $u1, u2 \in U$ exists if there has been observed a transition by a city-dweller from $u1$ to $u2$. The definition of uis can be arbitrary (e.g., municipal neighborhood borders, grid etc.). In our analysis, we divide the whole city (10^2 miles rectangle area considered around the center of each city) into 400 neighborhood areas, each one of 0.5 miles^2

In order to obtain the structure of G_T for both cites we use geo-tagged social-media

user-generated content. In particular, we collect Tweets using Twitter's streaming API from Jul 15 to Nov 15 2013. Each tweet has a tuple format<*user Id, place Id, time, latitude, longitude*>. In total, we have 526,799 geo-tagged tweets in Pittsburgh and 3,715,016 in NYC. Figure 1 presents the distribution of tweets in two cities examined. Using these data, we generate edge (transition) $e_{ij} \in E$ if the same Twitter user has generated two consecutive tweets in locations $li \in ui$ and $lj \in uj$ within a predefined time interval Δt and the distance between these two locations is greater than a threshold Δd. In our experiment, we set $\Delta t = 4$ hours and $\Delta d = 10m$. Finally, we have 172,887 such pairs in Pittsburgh and 961,671 in NYC. Note that the above definition allows for self-edges in *GT*. We can also annotate every edge e_{ij} with a weight, which captures the number of transitions between the two urban areas i and j.

Centrality in GT: To capture the centrality of human movement in different neighborhoods, we calculate the PageRank [9] for each node in *GT*. In particular, we calculate a weighted PageRank score Pi of area i as:

$$P_i = \alpha \sum_j A_{ij} \frac{P_j}{k_j^{out}} + \beta_i, \qquad (1)$$

Where a=0.85 and k_j^{out} is the weighted out-degree of node j which counts self and outgoing edges β_i is a personalized (external) priority importance for area i, which is defined as the fraction of tweets taking place in area i.

This work will also use a second simple centrality metric for G_T, which is the number of tweets $n_{t,i}$ generated in area i. The latter does not incorporate mobility information, but rather captures the intensity of activity in each area.
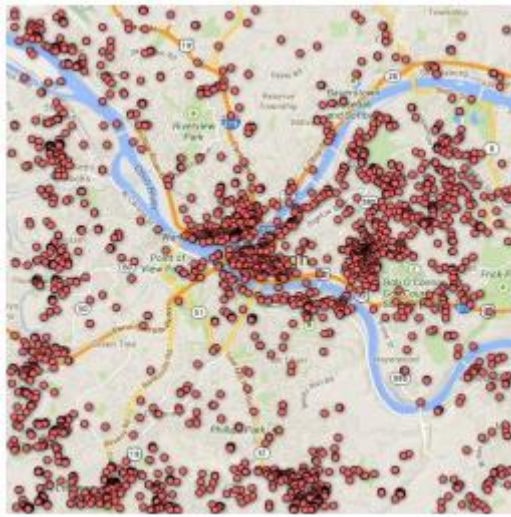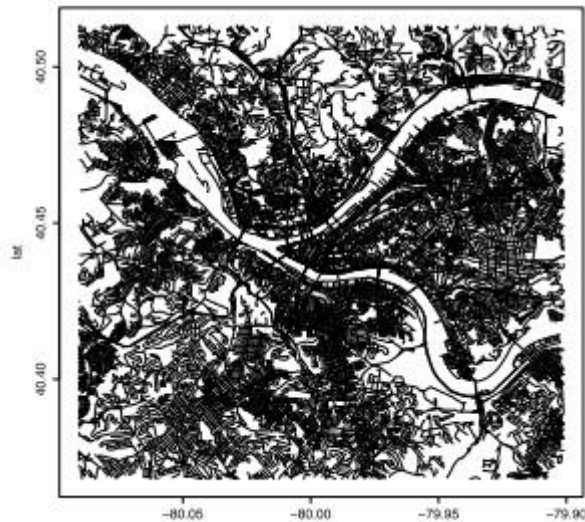
(a) Pittsburgh   (b) New York City

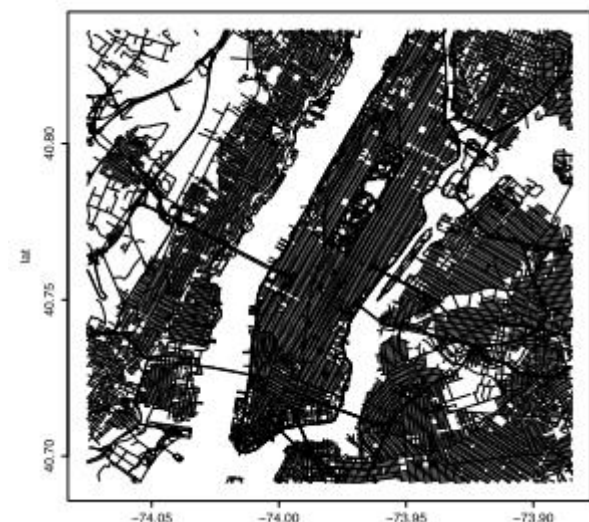Figure 1.   Street network in selected urban areas of two cities

B. Street Network

This paper will model the street network through a graph *Gs =(V,S)*, where the set of nodes represents the intersections in the street spatial structure and an edge $s_{ij} \in S$ represents the street segment that connects intersections *i* and *j*. We fetch the street networks from OpenStreetMap and process them using osm4arouting [1] into the G_s

network format. osm4routing provides additional metadata such as the coordinates of each intersection, the length of each street segment and accessibility flags for each street segment in two directions(e.g accessibility by car, foot, bike etc.). Figure 2 further gives a visualization of the street networks in both cites.

(a) Pittsburgh   (b) New York City

Figure 2.   Street network in selected urban areas of two cities

Centrality of Street Network: For a road network with *n* nodes and *m* edges, we calculate three well-established measures of node centrality: closeness centrality betweenness centrality C^b and straightness centrality C^s, C_i^c caps the accessibly of node and is defined as [3]:

$$C_i^c = \frac{n-1}{\sum_{j=1, j\neq i}^n d_{ij}} \qquad (2)$$

Where d_{ij} is the shortest path length between nodes *i* and *j*. C_i^b quantifes to what extent node I serves as a "broker" betteen nodes, is formally defined as [3]:

$$C_i^b = \frac{1}{(n-1)(n-2)} \sum_{s=1; t=1; s\neq t \neq i}^n \frac{n_{st}^i}{n_{st}} \qquad (3)$$

Where, n_{st} is the number of shortest paths been nodes *s* and *t* while n_{st}^i is the number of such shortest paths that traverse node *i*. C_s measures the extent to which node *i* can be reached directly, on a straight line, from all other nodes, which is defined as [6]:

$$C_i^s = \frac{1}{n-1} \sum_{j=1; j\neq i}^n \frac{d_{ij}^{Eucl}}{d_{ij}} \qquad (4)$$

Where, d^{Eucl} is the Euclidean distance between nodes *i* and *j*.

Finlly, we calculate three global and nine local indices of street centralities. The global in indices, $C^c glob$, $C^b glob$ and $C^s glob$, are calculated using the whole road network. We also consider the local version of centralities $C^c local,d$, $C^b local,d$ and $C^s local,d$, where we compute the centrality of node *i* considering only the nodes that are within a radius *d*.

C. Analysis setup

[1] https://github.com/Tristramg/osm4routing

Our goal is to examine the relation between the central areas in a city as captured through the mobility of people, and the central areas of the city as captured through the street network. For that, we will utilize the Spearman's rank correlation coefficient *ρ*. In particular, the first variable for this correlation will be the PageRank centrality *Pi* of nodes *i* $\in U$ (as well as *nt,i*). However, the centrality values that we got from the street networks are defined on a different set of nodes(set *V*). Thus, we will use a spatial mapping $\Phi : V \rightarrow U$ utilizing the lat/lon coordinates we have for every *v* $\in V$. With Φ in place, the second variable for calculating *ρ* will be the average road network centrality. $\overline{C_v^*}$ of all nodes *v* $\in V$ that map to *i* $\in U$, that is, $\Phi(v)=i$.

III. RESULTS AND ANALYSIS

We take the urban street network as an directed network without consideration of the traffic accessibility in two directions. Table 1 presents the correlation results for Pittsburgh and NYC. We can see that the global closeness centrality $C^c glob$ and betweenness centrality $C^b glob$ highly correlate with the intensity of human movement in both environments. This suggests that center areas in urban cities tend naturally to be more accessible from/to other places (higher $C^c glob$) and thus function as city "hubs"(higher $C^b glob$). In contrast the global straightness centrality $C^s glob$, local closeness centrality $C^c local,d$ and local betweenness centrality $C^b local,d$ present no significant positive correlations. However, the straightness centrality $C^s local,d$ shows an interesting urban difference with a significant level of correlation in Pittsburgh but not in NYC. This is more likely due to the difference of urban space structures or travel patterns between the two cities.

Further analysis is needed to sort out the exact   source of this difference

TABLE I.   CORRELATION P(*INDICATES A P-VALUE<0.05; ** INDICATES P-VALUE<0.01)BETWEEN THE STREET CENTRALITY AND THE INTENSITY OF HUMAN MOVEMENT.

G_T C	Pittsburgh		NYC	
	$n_{t,i}$	P_i	$n_{t,i}$	P_i
C^c_{glob}	0.610**	0.604**	0.509**	0.505**
C^b_{glob}	0.501**	0.497**	0.459**	0.466**
C^s_{glob}	0.021	0.020	0.078	0.074
$C^c_{local,d=800m}$	-0.223**	-0.228**	-0.085	-0.093
$C^c_{local,d=1600m}$	-0.043	-0.046	0.012	0.004
$C^c_{local,d=2400m}$	0.024	0.0189	-0.044	-0.047
$C^b_{local,d=800m}$	-0.001	-0.128*	0.009	-0.127*
$C^b_{local,d=1600m}$	0.017	0.026	-0.072	-0.070
$C^b_{local,d=2400m}$	0.106*	0.112*	-0.014	-0.014
$C^s_{local,d=800m}$	0.348**	0.351**	0.105*	0.104*
$C^s_{local,d=1600m}$	0.410**	0.408**	0.028	0.026
$C^s_{local,d=2400m}$	0.442**	0.438**	-0.031	-0.031

This research further consider the urban street network as a directed graph based on the direction accessibility for three types of movements including driving biking and walking. In this case, there are two different calculation or closeness and straightness centrality based on two types of shortest paths between nodes. The first one is outgoing shortest path d^{out}_{ij}, with the direction starting from node i to node j. The second is incoming shortest path d^{in}_{ij} with direction into node from node j, capturing how easily a traveler can access node i from other locations in the city. Therefore, we have *in* and *out* closeness and straightness centrality based on these two types of shortest path calculations. Table 2 presents the correlation between the centrality of directed street network and the PageRank score of neighborhood areas (results for $n_{i,t}$ are omitted due to space limitations). Compared to Table 1, we do not observe significant differences when considering the directed networks. This might be due to the fact that the transition network G_T essentially captures the starting and ending point of a movement, ignoring the actual path followed and/or due to the high similarity of the different directed network structures. Nevertheless, there is still some significant change for global straightness centrality when considering directed street network-especially for biking and walking-which might be attributed to the fact that for these "slow modes" of transportation short geometric distance is

TABLE II.   CORRELATION RESULTS BY CONSIDERING THE ROAD NETWORK AS A DIRECTED NETWORK BASED ON THE ACCESSIBILITY OF DRIVING, BIKING AND WALKING IN EITHER DIRECTIONS.

PageRank / C	driving		biking		walking	
	Pittsburgh	NYC	Pittsburgh	NYC	Pittsburgh	NYC
C^c_{glob} (in)	**0.597****	**0.473****	**0.622****	**0.397****	**0.616****	**0.393****
C^c_{glob} (out)	**0.594****	**0.481****	**0.623****	**0.391****		
C^b_{glob}	**0.481****	**0.431****	**0.520****	**0.452****	**0.514****	**0.444****
C^s_{glob} (in)	-0.053	0.061	0.200**	0.301**	0.212**	0.313**
C^s_{glob} (out)	-0.002	0.083	0.231**	0.303**		
$C^c_{local,d=800m}$ (in)	-0.253**	-0.143**	-0.250**	-0.087	-0.241**	0.042
$C^c_{local,d=800m}$ (out)	-0.282**	-0.142**	-0.253**	-0.069		
$C^c_{local,d=1600m}$ (in)	-0.133**	-0.170*	-0.123**	-0.117*	0.103*	-0.012
$C^c_{local,d=1600m}$ (out)	-0.067	0.003	-0.103*	-0.012		
$C^c_{local,d=2400m}$ (in)	-0.053	-0.215**	-0.024	-0.178**	0.011	-0.078
$C^c_{local,d=2400m}$ (out)	-0.039	-0.204**	-0.077	-0.171**		
$C^b_{local,d=800m}$	0.042	0.100*	0.044	-0.066	0.041	-0.081
$C^b_{local,d=1600m}$	0.061	0.125*	0.072	-0.035	0.062	-0.044
$C^b_{local,d=2400m}$	0.140**	0.100*	0.161**	0.009	0.143**	0.002
$C^s_{local,d=800m}$ (in)	0.248**	0.053	0.324**	0.002	0.362**	0.094
$C^s_{local,d=800m}$ (out)	0.248**	0.053	0.324**	0.002		
$C^s_{local,d=1600m}$ (in)	0.306**	0.046	0.363**	-0.020	0.396**	0.032
$C^s_{local,d=1600m}$ (out)	0.306**	0.046	0.363**	-0.020		
$C^s_{local,d=2400m}$ (in)	0.349**	0.003	0.386**	-0.051	0.423**	-0.023
$C^s_{local,d=2400m}$ (out)	0.349**	0.003	0.386**	-0.051		

IV. DISCUSSION AND FUTURE WORK

In this paper we examined the correlations between the centrality of street networks with the intensity of human movement in urban areas and we found that the correlation level differs with different centrality metrics, of which some further depend on different cities. Our work provides an illuminating way to study the relationship between urban structure and human movement in a large-scale way.

We would like to emphasize that our analysis methods may suffer from a variety of biases. For example, we examine the correlation by aggregating the road network centrality and human movement in each neighborhood area, while a microscopic study might give a different view. Also, the rectangle urban area we pick may introduce edge effects on the correlation results. Furthermore, the large-scale available dataset used here may have some noises and biases. For instance, the street networks in OpenStreetMap might not that accurate especially for cities that are not that popular, since all the information is crowdsourced y the public. Also, the nature of voluntarily sharing may only give a partial information of human movement captured by geo-tagged tweets, of which the quality depends on many other factors, such as demographic biases, spam tweets and fake location information.

In the future, we plan to examine the levels of correlation by considering the temporal and contextual information of human movement such as the time and type. Furthermore, we aim to examine the centralities of a directed road network by considering the accessibility of different transportation modes (e.g., driving, biking and walking) in two directions on the street. For network centralities, we want to further investigate other practical factors, such as the max flow on a street (number of available lanes), the fastest path

and the density/type of resources surrounding a street intersection.

REFERENCES

[1] Bill Hiller. Alasdair Tumer, Tao Yang, and H-T Park Metric and topo-geometric properties of urban street networks: some convergences, divergences and new results. Journal of Space Syntax Studies, 2009.

[2] J.P. Rodrigue. Transportation and the urban form. Chapter 6, The Geography of Transport Systems, 3rd Ed., 2013.

[3] Mark Newman. Networks: an introduction. Oxford University Press, 2009.

[4] Aisan Kazerani and Stephan Winter. Can betweenness centrality explain traffic flow. In Proceedings of the 12th AGILE International Conference on GIS, 2009.

[5] lan XY Leung, Shu-Yan Chan, Pan Hui, and Pietro Lio. Intra-city urban network and traffic flow analysis from gps mobility trace. arXiv preprint arXiv. 1105.5939, 2011.

[6] Sergio Porta, Vito Latora, Fahui Wang, Salvador Rueda, Emanuele Strano, Salvatore Scellato, Alessio Cardillo, Eugenio Belli, Francisco C'ardenas, Berta Cormmenzana, et al. Street centrality and the location of economic activities in Barcelona. Urban Studies, 49(7):1471-1488, 2012.

[7] Francesco Battaglia, Giuseppe Borruso, and Andrea Porceddu. Real estates ban centra, econome act. a s analysis on e ity of snon() Computational Science and Its Applicanons-ICCSA 2010, pages 1-16. Springer, 2010.

[8] Bin Jing and Tao Jia. Agent-based simulation of human movement shaped by the underlying street structure. International Journal of Geographical Information Science, 25(1):51-64, 2011.

[9] Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. The pager-ank citation ranking Bringing order to the web. Technical report, Stanford Digital Library Technologies Project, 1998.