Identification of Central Points in Road Networks using Betweenness Centrality Combined with Traffic Demand

Rodrigo de Abreu Batista and Ana Lucia Cetertich Bazzan

Abstract—This paper aims to identify central points in road networks considering traffic demand. This is made with a variation of betweenness centrality. In this variation, the graph that corresponds to the road network is weighted according to the number of routes generated by the traffic demand. To test the proposed approach three networks have been created, which are Porto Alegre and Sioux Falls cities and a regular 10×10 grid. Then, trips were microscopically simulated and the results were compared with the proposed method.

Index Terms—Traffic assignment, complex networks, centrality, betweenness centrality.

I. INTRODUCTION

METROPOLITAN regions are currently facing major problems regarding urban mobility. For example, in 2014 Brazil had 78.1 million private vehicles, which represents an increase of 229.3% of fleet of vehicles. This means one private vehicle per 2.6 inhabitants in comparison to 2004.

Dealing with such growth in the fleet requires duly planning the highway system in order to reduce the effects of traffic on population and environment. Planning transportation systems involves, among other factors, analyzing the distribution of traffic flow throughout road networks.

There is evidence that the measure of betweenness centrality, as proposed by Freeman [1], is capable of predicting traffic flow in road networks. However, this measure ought to be adapted because it does not suitably represent the distribution of demand [2], [3], [4]. The objective of this paper is to analyze how betweenness centrality may be adapted and used in identifying central points in a road network. In this paper the term central points refers to those points that are most often traversed by road users when traveling along their routes.

II. DEFINITIONS

A. Transport Networks

Transport networks can be defined as networks that are composed of roadways and junctions between roadways (*e.g.*

intersections). Such networks are typically represented as weighted directed graphs $G = \{V, E\}$, composed of a set of vertices V (junctions) and edges E (roadways) and a cost function C(e) which associates weight to each edge. In the context of transport networks, the length, the travel time or the capacity are commonly used as the weight of the edges.

Demand for traffic represents the behavior of users in using the network infrastructure. By behavior, it is understood the decisions made by the users that are relevant to the problem which is being modeled (*e.g.* choice of routes).

The locations of origin and destination of demand are grouped in districts. Districts can be defined based on information obtained through sociodemographic studies, data of georeferencing and urban statistics, so that these variations may be the least possible within a given district [5]. In general, and also in this work, each district is associated to a location within the network and it is composed of a set of vertices and edges without there being overlapping with other districts.

The demand of a network is commonly represented by a matrix that relates districts of origin to destinations, associating each possible combination, to a figure that corresponds to the intensity in which these trips occur. As it relates origins to destinations, it is called an origin-destination matrix, or OD matrix.

A driver who wishes to travel from district s to district t, represented by (s,t), may use more than one series of edges that lead from s to t. Each these possible paths is called a route. Since each of these edges has an associated cost, there is particular interest in the path with the lowest cost, which in this article is called the path with the least length, or the shortest path. Therefore, creating the routes consists of associating series of edges to trips that are specified by the OD matrix.

B. Betweenness Centrality

Betweenness centrality is based on the idea that a vertex is more central as more low cost paths pass through it. The shortest paths between all pairs of vertices in the network are considered in this calculation. The traditional method for calculating betweenness, as well as other centrality measures, were originally developed in the scope of studying social networks, and they have recently been highlighted in the literature [6].

Manuscript received on June 30, 2015, accepted for publication on September 23, 2015, published on October 15, 2015.

The authors are with the Institute of Informatics, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil (e-mail: {bazzan, rabatista}@inf.ufrgs.br).

Betweenness for a vertex u is defined according to Equation 1, in which σ_{ij} is the number of shortest paths between i and j, and $\sigma_{ij,u}$ is the number of these paths through which u passes. The shortest paths are calculated on the basis of cost of the edges, thus centrality is sensitive to the function of cost that has been chosen.

$$B_u = \sum_{i \in V} \sum_{j \in V \setminus \{i\}} \frac{\sigma_{ij,u}}{\sigma_{ij}}.$$
 (1)

III. METHODS

This section presents a method for calculating betweenness by taking into consideration the demand that is going to use the network. The method consists of constructing a graph that represents the network and weighting it by using the demand. In order to calculate betweenness, the occupation of the roadways is taking into consideration and it results from the demand that is represented by the OD matrix.

A. Assigning Weights to the Edges

Before considering betweenness by considering the demand, it is necessary to determine the routes that correspond to the demand. As an OD matrix has only origins and destinations, calculating the routes is necessary for obtaining a table that adds up the amount of routes that pass through each edge. The routes were obtained by calculating the path of lowest cost between the origin and the destination of each trip that is presented in the OD matrix. That is why sequences of edges that form the paths with the lowest cost were found for each OD pair, by using the Dijkstra algorithm [7].

Once the routes that correspond to the demand have been calculated, it is possible to construct a table of the occurrence of the amount of routes which pass through each edge. This procedure is a basic stage for defining the weights of the edges in the network graph, since the amount of routes will serve as input for the functions used to calculate the cost of the edges. Weight to the edges can be carried out in distinct manners. Several studies use the length of the roadway as the cost of the edges [3], [8], [2], [9]. In this paper we have used the occupation rate of the roadways as the weight of the edges, as it is understood that it reflects the use of the network related to the demand. Furthermore, the unitary function, in which the lowest cost routes refer to the number of hops necessary to go from the origin to the destination, was also considered.

In this study, we decided to use decreasing cost functions to attribute costs to the edges. Figure 1 illustrates the situation that brought about this decision. A network in which 10 routes pass along each of the edges is shown in Figure 1a; the values of betweenness for each of the vertices are shown in the same figure. One may notice that the betweenness values are evenly distributed since this is a regular network.

Supposing that the number of trips between vertices A and B is increased by 5 trips, the natural logic is to increase the number of trips on the edge to 15. Figure 1b illustrates whats happens to the values of betweenness when the cost of edge

AB is increased. In this case, as the betweenness algorithm considers the paths of lowest cost, paths that previously passed through AB have ceased to do so. Thus, vertices C and D have come to receive the greatest betweenness values, while vertices A and B were those that received a real increase in demand. This would require the reader to use and inverted interpretation of the measure, so that the vertices with the lowest betweenness values are central in relation to the demand.

In order to solve this problem and have the greatest values of betweenness be attributed to the vertices with the greatest volume of demand, we decided to use decreasing cost functions. In this case, an increase of demand between vertices A and B causes a decrease in the weight of the edge, as this was attributed by a decreasing function. Figure 1c shows that in this case the vertices with the highest values of betweenness coincide with the vertices that have the highest demand.

The experiments were guided by taking the following cost functions into consideration:

 F_1 : Decreasing Exponential It is possible to use an exponential function for modeling the cost of an edge according to its occupation. Likewise, assuming that the cost of an edge also decreases exponentially in relation to its rate of occupation, the cost attributed to the edges is defined as function *C*, as defined in range (0; 1]. This weight is calculated in accordance with the decreasing exponential function that is shown in Equation 2, a particular case of the family of equations $y = a(1-b)^x$, and it only considers the amount of trips *n* that pass through a given edge.

$$C(n) = (1 - 0.001)^n.$$
 (2)

 F_2 : Rational Function The rational cost function shown in Equation 3 was also considered in the experiments. This function was chosen to explore the behavior of betweenness distribution when cost decreases faster than the exponential function had previously explained.

$$C(n) = 1/n. \tag{3}$$

 F_3 : Decreasing Linear Function The linearly decreasing function exhibited in Equation 4, in which k > n, was also considered in the experiments. The main objective of using a linear cost function is to study the behavior of the proposed method when edge costs are diminishing more smoothly than the decreasing exponential function.

$$C(n) = k - n. \tag{4}$$

- F_4 : Number of Hops Considers the number of hops that were performed. This function represents the number of edges there are in the path with the lowest cost calculated for an OD pair.
- F_5 : Length of an Edge Considering that several studies have used the length of a roadway as weight of the edges, this was used in the cost function F_5 as a way to compare this study to previous studies.



Fig. 1. Example of influence of the cost of the edges in betweenness.

B. Calculation of Betweenness Centrality Considering Demand

After the graph that represents roadways has been constructed and its edges have been properly weighted by using the routes generated from the OD matrix and cost function, it is possible to calculate the betweenness. As the interest here lies in identifying the central points, the results of the betweenness algorithm shows high values for vertices that have greater demand.

Algorithm 1 lists the steps involved in calculating betweenness centrality considering the demand.

Algo	Algorithm 1 Betweenness considering demand						
Requ	Require: Network R, OD Matrix M, Cost Function C						
Ensu	re: Betweenness from all vertices						
1: p	rocedure DemandBetweennessCentrality (R, M, C)						
2:	Construct graph G with the same topology as R						
3:	Calculate the routes from M over G						
4:	Create a table T of routes passing through each edge						
5:	Attribute costs to the edges of G by using T and C						
6:	Calculate betweenness over G						
7:	return betweenness from the vertices						

IV. EXPERIMENTS AND RESULTS

The proposed approach was tested by means of experiments on three networks. Two of these are abstractions of real networks in the cities of Porto Alegre and Sioux Falls, and a third network consists of a regular 10x10 grid. All three networks are shown in Figure 2. Furthermore, with the aim of analyzing the behavior of betweenness at different occupation levels, demands with volumes of 10%, 25%, 50% and 75% of the total capacity of each network are used.

A. Traffic Demand

For the Porto Alegre Arterials network a pattern of demand was specified with the aim of reproducing the flow patterns of drivers that are observed in the city at the beginning of the day, in which they leave the outskirts of the city and go downtown. In this demand, which is called Non-Uniform Outskirts \rightarrow Downtown Demand (NUODD), seven distinct

points in the outskirts of the city and one point in the central region of the city were used as origin and destination respectively. Considering that the capacity of the Porto Alegre Arterials network is 127,320 vehicles, demands of volume of 10% (12,372 trips), 25% (31,830 trips) and $50\%^1$ (63,660 trips) were defined.

ISSN 2395-8618

For the Sioux Falls network, the same model of demand used in the paper by Chakirov and Fourie [10] was used. In their study the authors based their work on census data to create a de-aggregate demand and a microscopic model of the Sioux Falls network, based on the network that was originally used in the study by LeBlanc *et al.* [11]. In this study only the volume of demand corresponding to the morning rush hour was used. Thus, the demand used in this network has a volume of 44,652 trips and it was generated by an iterative model in order to achieve the stochastic user equilibrium. See [10] for details.

For the regular 10x10 grid two regions were determined edge and center—on which three types of demand were defined. The first of these, uniform demand (UD), shows uniform distribution of the origins and destinations of the trips that were generated, and its aim is to create random trips within the 10x10 grid. The second, Non-Uniform Edge—Center Demand (NUECD), is composed of trips that go from the border toward the center of the grid, and which have the aim of creating congestion in the central region. The third type of demand which is called Non-Uniform Center—Edge Demand (NUCED), is composed of trips that leave the center and go toward the edge of the network. Considering that the grid has a capacity of 4,890 vehicles, demands of volume equivalent to 10% (489 trips), 25% (1,223 trips) and 50% (2,445 trips) are used.

B. Comparing the Proposed Method to a Microscopic Simulation

Since access to the real measurements that were carried out on roadways of the cities of Porto Alegre and Sioux Falls were not available, it was decided to test the proposed

¹The total volume is equal to 50% of the maximum capacity of the network. Demands with volumes over 50% higher were not simulated.



Fig. 2. Networks used in the experiments.

technique by means of comparing it to a simulation. In this case, a microscopic simulation performed in the SUMO [12] simulator was used, and it applies the routes calculated from an OD matrix to a given network. Figure 3 shows the steps involved in the microscopic simulation process and the steps of the proposed method. Both methods receive the OD matrix and the road network files as input and calculate the betweenness of the vertices at the end. For the microscopic simulation, additional steps for trips and routes generation are required to produce the SUMO related files.

By using SUMO it was possible to obtain information about occupation of the edges at the peak of the occupation of the network. Figure 4 shows the mean occupation curve of the Sioux Falls network along time, highlighting the time-step of the peak of mean occupation. The rates of occupation obtained were used to weight a graph that represents the network, on which betweenness which serves as a basis for comparing it to the model proposed was calculated. Table I shows the vertex-by-vertex details of betweenness that were calculated by the proposed method and the betweenness values calculated at the peak of occupation of the network. The five most significant values of each case study were highlighted to make visualization easier.

C. Results

With the aim of comparing the results of betweenness obtained by the proposed method to the results obtained through simulation, Pearson's correlation coefficient was used. Thus, the correlation between the results generated by the proposed method with the results obtained by the simulation were calculated for each of the experiments. This correlation was calculated between the results of betweenness over each set set of vertices. Table II shows the results that were detailed by the experiments and the cost function.

In the case of the Porto Alegre Arterials network, decreasing linear function and rational function showed the best results.

Considering that hops and edge length functions disregard demand, it is possible to note that even so, the former showed results that were significantly better than the latter. It is also possible to note that, for lower volumes, the correlation values obtained were greater, which may be attributed to the fact that microscopic simulation considers factors which the static model does not consider.

In the experiment on the Sioux Falls network, the decreasing exponential functions and the rational functions were those that showed the best results. This experiment was the one that showed the lowest rates of correlation. This can be attributed to the fact that the routes of this demand were generated by a different process than the others. In this case, the routes were generated by the model developed by Chakirov and Fourie [10], while in the other case studies, the routes were calculated by considering the shortest paths.

In the case of the 10x10 grid, hops and edge length functions were the ones that showed the best results. Specifically for this case study, the fact that shortest paths with same value exist between a given origin and destination showed a deviation that may have distorted the results. Another point to be noticed is the strong correspondence between the hops and edge length columns, due to the regularity of the grid, which makes the edge length function just as accurate to the hops function.

We observed that the decreasing exponential function and the decreasing linear function showed the best results when the instances of Sioux Falls and Porto Alegre Arterials were considered. As the occupation peak of the network is being considered, many edges have occupation rates that are near 1. The hops function also showed significant results, exceeding the others in some cases.

V. RELATED WORK

In Holme's paper [6], he investigates the relation between traffic flows in communication networks and centrality measures. In this model, particles are moved along the



Fig. 3. Steps involved in the betweenness centrality calculation using the proposed method and the microscopic simulation.

TABLE I

Betweenness calculated by using different functions for attributing cost and the simulation over the Sioux Falls network, detailed by vertices. The top five betweenness values are shown in red for each case.

Vertex		Simulation					
VEILEX	F_1	F_2 F_3		F_4	F_5	Simulation	
4	19,294	17,562	17,038	19,117	17,006	12,907	
5	6,554	10,559	15,610	18,557	24,134	7,572	
6	15,091	18,700	15,519	15,242	16,639	14,279	
8	18790	18808	12784	12961	13774	16894	
9	13,674	13,623	19,573	20,462	27,655	14,953	
10	42,980	37,882	36,784	30,539	33,664	31,824	
11	35,269	33,075	31,314	25,068	23,418	42,001	
14	4,657	6,185	9,400	12623	15,148	29,303	
15	16,300	15,283	17,327	18,509	17,621	29,817	
16	25,777	21,679	12,148	12,148	13,699	29,798	

TABLE II Correlation between betweenness centrality calculated by the proposed method and simulation, detailed by experiments and cost functions.

Instance	Cost Function						
Instance	F_1	F_2	F_3	F_4	F_5		
Sioux Falls	0.61	0.61	0.59	0.54	0.45		
POA Arterials NUODD Vol. 10%	0.96	0.92	0.99	0.98	0.69		
POA Arterials NUODD Vol. 25%	0.80	0.84	0.84	0.81	0.58		
POA Arterials NUODD Vol. 50%	0.76	0.78	0.77	0.74	0.69		
Grade 10x10 NUECD Vol. 10%	0.84	0.79	0.88	0.89	0.89		
Grade 10x10 NUECD Vol. 25%	0.90	0.83	0.91	0.86	0.86		
Grade 10x10 NUECD Vol. 50%	0.91	0.85	0.93	0.80	0.80		
Grade 10x10 UD Vol. 10%	0.67	0.61	0.67	0.89	0.89		
Grade 10x10 UD Vol. 25%	0.63	0.59	0.63	0.91	0.91		
Grade 10x10 UD Vol. 50%	0.73	0.73	0.73	0.75	0.75		
Grade 10x10 NUCED Vol. 10%	0.90	0.73	0.90	0.88	0.88		
Grade 10x10 NUCED Vol. 25%	0.84	0.82	0.85	0.87	0.87		
Grade 10x10 NUCED Vol. 50%	0.78	0.67	0.78	0.83	0.83		

edges of a graph, constrained by the restriction that two particles may not occupy the same vertex at the same time. The particles move along between their randomly defined origins and destinations, and three different updating policies are considered, which are: random walk, in which particles randomly choose a position; detour-at-obstacle, in which a particle randomly chooses a position among their neighbors that are nearest the destination; and wait-at-obstacle, in which if no vertices are free near the destination, the particle does not move.

In order to monitor the traffic density regarding betweenness, the author chose the scale-free network model of Barabási-Albert, because it shows a wide distribution of betweenness values. Regarding betweenness, the author noted that the vertices with low or average betweenness rates showed steady occupation rates, and concluded that betweenness itself cannot estimate the capacity of a vertex. At this point, our work differs from Holme's work since we use a microscopic simulation to compare to the betweenness. We also consider nonuniform demands that differ from randomly defined OD pairs used in Holme's work. Beyond that, the subject of study in our work was road networks, while in Holme's work the author focused on communication networks. It influences basically the network types studied: communication networks can be explained by scale-free models, while road networks can be better explained by random graphs.

In the study of Kazerani and Winter [2], the issue related to the capacity of betweenness for explaining traffic flows was analyzed. In their study, they came to the conclusion that the traditional betweenness measure [1] is unable to explain traffic flows significantly because it does not consider the traffic demand that flows in a network, nor its dynamics. ISSN 2395-8618



Fig. 4. Mean occupation of the Sioux Falls network related to time, with the peak of occupation occurring at the time-step 32,671.

Still regarding the study of Kazerani and Winter, the authors suggest that an adaptation of the traditional centrality measure is necessary, in which the physical and temporal aspects would be considered, so that a significant correlation with the traffic observed in the network may be attained. In this work, we proposed a variation of the betweenness that, by considering traffic demand, achieved higher correlation with the observed traffic in microscopic simulation.

In the study of Gao *et al.* [3], the authors investigate the capacity of betweenness to preview traffic flows by analyzing the correlation between the centrality measure and real traces collected from GPS. For that, data collected from GPS installed on 149 taxis of the city of Qingdao (China) were used.From this analysis, the authors concluded that the betweenness measure does not explains the traffic flows well, and they attribute it two main reasons: first, when calculating betweenness, origins and destinations of trips are vertices of the graph, while in real OD pairs origins and destinations are associated with edges; second, the OD pairs distribution is not uniform, being associated to the distribution of human activity, which, is influenced by factors like area occupation.

In a tentative to explain the differences between betweenness and traffic flows, the authors established a comparison between the centrality measure and a model developed in three steps: in the first step it is assumed that the demand occurred uniformly over geographic space; in the second, the model was extended to consider the distribution of human activity using data collected from the use of cellphones; in the third step, the demand model developed in the previous step was extended to support the distance decay factor, that models the behavior of people of seeking resources in nearest places.

When compared with betweenness values, the model developed in the second step has shown the greatest correlation. The study suggests that the betweenness presents low correlation with the model developed in the third step because it does not consider the distance decay factor. However the authors do not suggest modifications to the betweenness, which is known to consider uniform demand and disregard distance decay factors.

In this sense, we have also addressed the problem of the distance decay factor found in the model presented by Gao *et al.* by using an OD matrix. Since the OD matrix represents in fact the traffic demand, only the OD pairs related to the desired trips will be found in this matrix.

In the study of Galafassi and Bazzan [4] a betweenness variation that considers traffic demand is suggested. Differently from the metrics proposed by Freeman [1], in their study only the routes that belongs to the specified traffic demand were considered. The authors compare the correlation between the modified metrics of betweenness with the amount of waiting vehicles on the edges, and show that the proposed method explains traffic flows better than the original metrics. The experiments were executed over a 6x6 regular grid and Porto Alegre network, both considering different demand volumes and types (uniform vs nonuniform).

The method proposed here is an extension of the study developed by Galafassi and Bazzan. In this work we kept the betweenness calculation module unchanged, modifying the way of how weights are attributed to the edges to consider the traffic demand. We also compare the proposed method with the occupation of edges during simulation, instead of the waiting vehicles' queue and extended the experiments to consider Sioux Falls network.

VI. CONCLUSIONS AND FUTURE WORK

The problem addressed along this work is central points identification in road networks using betweenness centrality. We have noticed that some authors tried to explain traffic flows using betweenness and failed because this metric itself assumes a uniform distribution of demand. Our method, on the other hand, consisted in combining the betweenness algorithm with the traffic demand so that higher values of betweenness were attributed to the vertices with higher demand.

The proposed method was tested in three networks and a microscopic simulation was performed for each one of them. The occupation of the edges was extracted from the simulations and the results were correlated with the betweenness values calculated by the proposed method. In general, the exponential decay and linear cost functions showed the best results among the studied functions.

The improvement in the proposed method was basically caused by two factors. First, the shortest paths calculated by the betweenness algorithm were influenced by the demand that uses a route. The second point is credited to the use of decreasing cost functions, which caused the weight of an edge to decrease as a function of the number of routes that pass through it.

Despite the fact that the technique proposed in this study is able to help identify central points in transport networks, it is only a small step towards a larger goal, which is to improve the road users' travel times. Hence, a possible extension of this study would be to assess whether traffic light operations at points with high values of betweenness would improve the average travel time for drivers, and whether these points are, in fact, the most critical ones.

Another aspect that could be investigated is the use a weighted correlation coefficient that is calculated considering the capacity of each vertex. Vertex capacity could be estimated by the capacities of its incident edges. Thus, vertices that have large capacity would receive greater weight in the calculation of the correlation.

Another possible extension of this work is to analyze the occupation of links individually during the microscopic simulation, and to approximate a function that models its behavior. This function could be used in the algorithm proposed in this work so that its performance could be compared with other functions considered here.

ACKNOWLEDGMENT

Ana Bazzan is partially supported by CNPq.

REFERENCES

- L. C. Freeman, "A set of measures of centrality based on betweenness," Sociometry, pp. 35–41, 1977.
- [2] A. Kazerani and S. Winter, "Can betweenness centrality explain traffic flow," in *Proceedings of the 12th AGILE International Conference on GIS*, 2009.
- [3] S. Gao, Y. Wang, Y. Gao, and Y. Liu, "Understanding urban traffic-flow characteristics: A rethinking of betweenness centrality," *Environment* and Planning B: Planning and Design, vol. 40, no. 1, pp. 135–153, 2013.
- [4] C. Galafassi and A. L. C. Bazzan, "Analysis of traffic behavior in regular grid and real world networks," in *Proceedings of the Fifth International Workshop on Emergent Intelligence on Networked Agents (WEIN)*, 2013. [Online]. Available: http://www.inf.ufrgs.br/ maslab/pergamus/pubs/GalafassiBazzan2013-wein.pdf

- [5] A. L. Bazzan and F. Klügl, "Introduction to intelligent systems in traffic and transportation," *Synthesis Lectures on Artificial Intelligence and Machine Learning*, vol. 7, no. 3, pp. 1–137, 2013.
- [6] P. Holme, "Congestion and centrality in traffic flow on complex networks," *Advances in Complex Systems*, vol. 6, no. 2, pp. 163–176, 2003.
- [7] E. W. Dijkstra, "A note on two problems in connexion with graphs," *Numerische mathematik*, vol. 1, no. 1, pp. 269–271, 1959.
- [8] P. Crucitti, V. Latora, and S. Porta, "Centrality in networks of urban streets," *Chaos: an interdisciplinary journal of nonlinear science*, vol. 16, no. 1, p. 015113, 2006.
- [9] S. Derrible and C. Kennedy, "Applications of graph theory and network science to transit network design," *Transport reviews*, vol. 31, no. 4, pp. 495–519, 2011.
- [10] A. Chakirov and P. Fourie, "Enriched sioux falls scenario with dynamic and disaggregate demand. working paper," Future Cities Laboratory, Singapore-ETH Centre (SEC), Singapore, Tech. Rep., 2014.
- [11] L. J. LeBlanc, E. K. Morlok, and W. P. Pierskalla, "An efficient approach to solving the road network equilibrium traffic assignment problem," *Transportation Research*, vol. 9, no. 5, pp. 309–318, 1975.
- [12] M. Behrisch, L. Bieker, J. Erdmann, and D. Krajzewicz, "Sumosimulation of urban mobility-an overview," in *SIMUL 2011, The Third International Conference on Advances in System Simulation*, 2011, pp. 55–60.