Computer-assisted city touring for explorers

Gabriel Spadon and Jose F. Rodrigues-Jr

University of Sao Paulo, Sao Carlos – SP, Brazil
spadon@usp.br, junio@icmc.usp.br

Abstract. The basic purpose of a map is to trace shortest paths between two locations in a city. However, this is not always what a user needs. Consider a tourist in an unknown city, he/she might want to trace routes to visit multiple landmarks while passing through the main streets of the city, possibly more than once through the same street. Such functionality is not yet available in online map services, which are prone to provide shortest paths that connect all landmarks. Rather, is of common interest a tour that puts together the most central streets (topologically speaking), minimizes the trajectory and, at the same time, passes through such landmarks. To cope with this problem, it is possible to investigate techniques of Center-Piece Subgraph, Absorbing Random Walk Centrality and Spanning Edge-Betweenness; such techniques can be used to find induced subgraphs that optimize centrality measures for a set of referential nodes or edges, *i.e.* landmarks or streets. The results shall be in the form of optimized algorithms, and how to integrate them into online systems. Studies in this line can succeed if they can guarantee timely scalability at the same time that they provide algorithms that produce tours (1) considering all the known destinations; (2) including the main streets of a city; and, (3) ensuring the shortest routes.

Keywords: Complex Network; Urban Structure; City Tour.

1 Problem Statement

Digital maps are becoming available to everyone, anywhere, through computers, smartphones, car assistants, and electronic devices in general. The most common use of such maps is to trace routes, or shortest paths, among different locations. However, this basic use is not always what one needs. In the case of a tourist visiting an unknown city, for instance, the user is not looking for the fastest and/or shortest path. Rather, he/she is looking for a tour that visits multiple landmarks (*i.e.* points of interest) while passing, possibly more than once, through the main streets of the city. That is, the user not only has several destinations, he/she desires a route that favors hotspots of the city in what is usually called city-tour. For instance, imagine a tourist is on the east side of the Central Park in New York and wants to go all the way down to the World Trade Center; the fastest and shortest path is to go through the monotonous 7th avenue. For a tourist, differently, the best pick might be to go through Broadway, despite its traffic and higher distance. The idea then is to find the tour that puts together

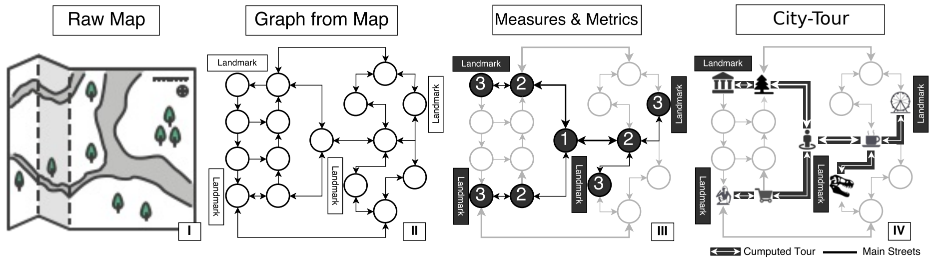

Fig. 1: An illustration of the methodology and desired results of the city-tour. It starts by **(I)** creating a graph from a map, **(II)** then the user informs some points of interest in the city; **(III)** these places are analyzed by means of their placement using metrics of graphs; **(IV)** then the algorithm returns a city-tour that passes through all of them and also through the main streets of the city.

the most central streets (topologically speaking), minimizes the trajectory and, at the same time, passes through points of interest. Differently, current systems are more likely to provide the user with shortest paths that connect all the points of interest. In other cases, the user does not even have destinations to visit; he/she wants to cruise through the city, discovering it iteratively. Such functionalities are not yet available in online map services and can be explored through complex-network tools by using a methodological process similar to the one presented in Figure 1.

2 Research Goals

The goal is to use graph-processing techniques to compute touristic routes with, or without, multiple destinations satisfying performance constraints. Given a city, or a region of a city, such routes might come as the answer to two queries: **(1)** to compute a multiple-destinations tour; or, **(2)** to compute a tour, even if the set of destinations is empty. The second query might sound odd, but it is not an unusual case, especially for explorers in unknown cities. Notice that, such queries depend on the investigation of techniques based on random-walk with restart [12], usually used to find hub nodes of a network, but that, by means of edge processing, might be used to find the most significant edges related to a set of nodes [7]. The queries also depend on techniques that identify the edges with the highest centrality, which, in the context of maps, must be concomitantly related to shortest paths so to reflect good routes — Hannah *et al.* [1] performed a study that explores these properties, but not considering the city-tour constraints.

3 Related Works

Besides the aforementioned work, others touched the issue of processing digital maps. Scellato *et al.* [8] investigated how it is possible to extract the backbone of a city using Spanning Trees based on Edge-Betweenness and Information Centrality; they proposed a method that allows extending the comprehension of the most important routes that affect the city flow, retails, land-use separation, and that impact upon collective behavior. However, the authors used a greedy algorithm, which fails in circumstances when the topology is not adequate to the greedy strategy, for instance, when the lengths of the streets strongly diverge in different regions. Additionally, Delling *et al.* [2] have focused on algorithms to solve problems from the commercial domain; this is the case when one needs to choose the best placement for a new store. Dibbelt, Pajor, and Wagner [3] explored a multi-modal common-route planning problem. These studies consider the possibility of using heterogeneous transportation to go from one point to the other (common-route problem), favoring only the criterion of shortest paths.

4 Suggested Methodology

Nowadays, digital maps can be extracted from many different sources, but some of them are not freely available. OpenStreetMap[1] is a collaborative street-mapping community and an open-source option for digital maps. Over OpenStreetMap, it is possible to investigate techniques related to the questions placed in Section 2. Such research questions can be explored by means of three techniques, as follows:

Center-Piece Subgraph (CEPS). The Center-Piece Subgraph technique summarizes a graph, producing an induced topology that connects a subset of referential nodes provided as input. It follows by using random-walks with restart and cost functions to measure the adequacy of the edges that will be part of the resulting topology [11]. The summarized graph is validated by analyzing the goodness of the graph nodes [6], such that, heuristically, the best simplification is the one that contains the essential elements according to a goodness criterion (as defined by Equation 1), where $r(q, j)$ is the goodness score for a given node j considering a query set $\mathbf{Q}$.

$$g(CP) = \sum_{j \,\in\, nodes(CP)} r(\mathbf{Q}, j) \tag{1}$$

The goodness criterion minimizes the sum of weights to connect all the edges in the subset while inducing a subgraph; however, it does not consider edge properties, leading to a subgraph with semantics related to a route in a map. The metric was not designed to provide a minimum set of edges, as necessary in multiple-destination routes, but only with as many edges as desired by the user.

Example. Let us consider an unknown city represented as a graph $\mathbf{G}(V, E)$ which has a couple of touristic attractions $\mathbf{T}$ and let T' be the subset of those that he/she wants to *visit* or, alternatively, *avoid*. The idea then is to extract a subgraph G' that confers the user a reduced network in which he/she can travel

[1] www.openstreetmap.org

back and forth to visit the desired destinations. The output of CEPS, in this particular case, shall be a subgraph indicating the easiest and most interesting way to travel from a source place t_i to a target one t_{i+1}, where $\{(t_i, t_{i+1}) \in T'\}$.

Absorbing Random Walk Centrality (ARwC). The Absorbing Random Walk Centrality technique is capable of evaluating the centrality of a subgraph considering a set of query nodes Q [5]; ARwC considers the number of steps needed to absorb a walker that starts from a query node $q \in Q$ and walks to all the other nodes C in the graph, denoted as $ac_Q^q(C)$. This metric is based on the *k-Arw-Centrality* optimization, which is an NP-hard problem; to solve that problem, ARwC uses a greedy approach that provides solutions with good approximation guarantees. Equation 2 presents the metric, whose result is the centrality of a set of nodes C with regard to a set of query nodes Q.

$$ac_Q(C) = \sum_{q \in Q} \frac{1}{|Q|} \times ac_Q^q(C) \tag{2}$$

Example. Consider an unknown city with some touristic attractions represented as a graph. It is possible to assess the centrality of these attraction points inside the subset T' using ARwC, which will provide a set of values $\{C_t^{rw} \ \forall \ t \in T'\}$ that represent the importance of an attraction with respect to the others. Consequently, the result of this process is a hierarchical city view, which can: **(1)** represent how critical a node is among the others; **(2)** quantify how drastically he/she needs to avoid critical nodes to improve his/her mobility in a city; and, **(3)** classify the touristic attraction which will receive more or fewer visits considering its positioning.

Spanning Edge-Betweenness (SEB). A Spanning Tree (ST) is a graph structure that contains all the nodes connected by a subset of non-cyclic edges. It can be produced with weighted or non-weighted edges. In the context of a map, an accurately computed ST (not necessarily the minimum) generates a *Backbone* representation of a city, which conforms to the problem of computing a tour without a set of destinations. To this end, SEB is a centrality metric computed by considering all the STs of a graph; it works by quantifying the number of times that each edge pertains to an ST. The result of this metric is a hierarchical formation of streets according to their importance.

SEB was firstly defined over undirected and weighted graphs to improve the analysis of phylogenetic trees [9]; it has been shown to be a powerful tool able to evaluate the most relevant edges of a graph that, if removed, might disrupt the network structure [10]; notice that, the metric is not only to be used in the analysis of phylogenetic trees but also over massive graphs with millions of nodes [4]. Equation 3 formally defines SEB, where τ_G is the number of STs for a graph G, and $\tau_G(e)$ is the number of different STs where edge e occurs.

$$\delta_G(e) = \frac{\tau_G(e)}{\tau_G} \tag{3}$$

Example. Suppose that it is desired to detect the most common and interesting route that connects the nodes of an unknown city. That is, one wants not only the streets with the minimum length, but also wants those that, given the network topology, render better routes more frequently. This process can be achieved by using SEB whenever it is possible to infer weights to the set of network edges.

5 Expected Results

The results shall be in the form of optimized algorithms, and how to integrate them into online systems, including how to extract and prepare the network (maps are not necessarily represented as graphs), how to provide input to these algorithms, and how to interpret the outputs. The results shall be validated through extensive testing over a significant number of representative cities. Studies in this line can succeed if they can process queries over a map within seconds, and if the algorithms produce tours that **(1)** consider all the destinations (if known); **(2)** include the main streets of a city (higher centralities); and, **(3)** minimize the length of the paths (good routes). It is possible to verify those conditions in large scale by using brute-force algorithms or multi-objective optimization techniques; and, in small scales, by considering case studies of known cities and their tours. Research following these lines shall pave the way for future map processing, turning tours into a novel concept on electronic-trajectory computing. The results have the potential to be reported in international conferences and journals of Data Mining, Transportation Systems, and Physics.

6 Further Research Lines

The possibilities of research and investigation are not limited to the calculation of city-tours. It is possible, for instance, to analyze different sets of cities by extracting feature vectors that describe the complex-networks of their tours. This is because feature vectors can describe cities by detailing their topology and structural peculiarities, which would enable further analysis based on clustering detection, similarity search, multidimensional projection, and fractal analysis. All of which have the potential to enhance the understanding both about cities and tourist behaviors in the face of different types of landmarks that can be found in a given city. More specifically, by investigating feature vectors through these tools, it is possible: **(i)** to analyze cities according to their similarities and differences; **(ii)** to determine characteristics shared between cities or groups of cities; and, **(iii)** to disclose routing problems that are exclusive to peculiar cities. In fact, all these activities help in the designing of the urban space, they also help in its improvement and comprehension because similar cities might share the valuable characteristics; meanwhile, exceptional cities might indicate routing problems that can be further studied through a different set of tools.

Acknowledgments

We would like to thank the Brazilian agencies CNPq, FAPESP, and CAPES that fully supported this research.

Bibliography

[1] Bast, H., Delling, D., Goldberg, A., Müller-Hannemann, M., Pajor, T., Sanders, P., Wagner, D., Werneck, R.F.: Route planning in transportation networks. In: Kliemann, L., Sanders, P. (eds.) Algorithm Engineering, pp. 19–80. Springer International Publishing (2016)

[2] Delling, D., Goldberg, A.V., Pajor, T., Werneck, R.F.: Customizable Route Planning in Road Networks. Transportation Science (may 2015)

[3] Dibbelt, J., Pajor, T., Wagner, D.: User-Constrained Multi-Modal Route Planning. In: 2012 Proceedings of the Fourteenth Workshop on Algorithm Engineering and Experiments (ALENEX), pp. 118–129. Society for Industrial & Applied Mathematics (SIAM) (jan 2012)

[4] Mavroforakis, C., Garcia-Lebron, R., Koutis, I., Terzi, E.: Spanning edge centrality. In: Proceedings of the 24th International Conference on World Wide Web - WWW'15. pp. 732–742. ACM Press (2015)

[5] Mavroforakis, C., Mathioudakis, M., Gionis, A.: Absorbing random-walk centrality: Theory and algorithms. In: 2015 IEEE International Conference on Data Mining. pp. 901–906. IEEE (nov 2015)

[6] Rodrigues-Jr, J.F., Tong, H., Pan, J.Y., Traina, A.J.M., Traina, C., Faloutsos, C.: Large Graph Analysis in the GMine System. IEEE Transactions on Knowledge and Data Engineering **25**(1), 106–118 (jan 2013)

[7] Rodrigues-Jr, J.F., Tong, H., Traina, A., Faloutsos, C., Leskovec, J.: GMine: A System for Scalable, Interactive Graph Visualization and Mining. In: VLDB. ACM, Seul, South Corea (jun 2006)

[8] Scellato, S., Cardillo, A., Latora, V., Porta, S.: The backbone of a city. The European Physical Journal B - Condensed Matter and Complex Systems **50**(1-2), 221–225 (feb 2006)

[9] Teixeira, A.S., Monteiro, P.T., Carriço, J.A., Ramirez, M., Francisco, A.P.: Spanning edge betweenness. In: Workshop on Mining and Learning with Graphs. vol. 24, pp. 27–31 (2013)

[10] Teixeira, A.S., Santos, F.C., Francisco, A.P.: Spanning Edge Betweenness in Practice. In: Studies in Computational Intelligence, pp. 3–10. Springer Nature (2016)

[11] Tong, H., Faloutsos, C.: Center-piece subgraphs. In: Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD'06. pp. 404–413. KDD '06, ACM Press (2006)

[12] Tong, H., Faloutsos, C., yu Pan, J.: Fast random walk with restart and its applications. In: Sixth International Conference on Data Mining (ICDM'06). pp. 613–622. IEEE (dec 2006)