Similarity of paths in spatial networks: The case of long-distance athletic events

Dimitris Kavroudakis	Panagiotis Agourogiannis	Marios Batsaris	Michalis Vaitis	Efstratios Kavroudakis	Panayiotis Kouloumentas
Assistant Professor University of the Aegean Department of Geography Mytilini, Lesvos, Greece dimitrisk@aegean.gr	PhD Candidate University of the Aegean Department of Geography Mytilini, Lesvos, Greece geod17010@geo.aegean.gr	PhD Candidate University of the Aegean Department of Geography Mytilini, Lesvos, Greece m.batsaris@aegean.gr	Associate Professor University of the Aegean Department of Geography Mytilini, Lesvos, Greece vaitis@aegean.gr	MD, MSc, IOC DipSpPhy Orthopaedic Surgery, Sports Medicine, Traumatology Athens, Greece ekavroudakis@hotmail.com	MD, Msc Orthopaedic Surgery, Sports Medicine, Traumatology info@kouloumentas.gr

Abstract

The introduction of graph theory approaches in spatial sciences has open new horizons regarding spatial network analysis. Mathematical analysis of large spatial networks is now possible mainly due to increase of computational capabilities and the development of network algorithms based on modern database structures. This work illustrates a new algorithm which deals with paths along spatial networks and quantifies similarity between paths. After discussing the algorithm we illustrate a case study related to athletic events where we use the algorithm for the identification of similar paths in two different spatial networks. The identification of similar paths is useful for training reasons and has some potentials regarding searching between large scale spatial networks.

Keywords: Spatial Networks, Similarity of paths, Graph Theory, Athletic Events

1 Introduction

Spatial analysis approaches have been used in a number of contexts over the last couple of decades. With the evolution of computational approaches and the availability of hardware, a new era of large scale spatial analysis has been introduced. The availability of spatial data and the evolution of novel computation algorithms, has also boosted the development of new approaches for dealing with spatial problems and analyzing even more complex and interconnected spatial data-sets.

Spatial analysis along networks has been a relatively new field which is related to graph theory approaches and spatial sciences. There is currently a limited number of software capable of analyzing spatial networks.

The aim of this work is to investigate the process of identifying similar routes (paths) on large spatial networks based on multivariate similarity measures. We propose a similarity algorithm which compares spatial network paths based on a number of topographic criteria. Finally, we use the proposed algorithm in a case study related to long-distance athletic events such as the Marathon race.

2 Background

Affordable high performance computation systems, high speed internet connection and evolved Geo-computation approaches have enabled the ability of analyzing and solving large scale spatial problems in reasonable time. Analyzing large scale spatial networks, has been a very computationally intensive task, especially due to complexity and topology problems. Over the last decade, a growing number of spatial networks related applications has been approached regarding navigation systems and transportation optimization. theory approaches have been introduced in spatial sciences especially for the analysis and quantification of spatial networks. General network analysis approaches have been used in a number of contexts in spatial sciences' literature such as location analysis for optimal siting of fuel stations on spatial networks (Kuby et al., 2007), earthquake evacuation planning (Yamada, 1996). Demšar et. al. (2008) illustrate the identification of critical location in a spatial network with the use of classical Graph Theory approaches. On a more theoretical basis, the work of Reggiani et al. (2009) is a very good example of quantifying the notion of complexity in spatial networks as well as the work of Rozenblat et al. (2013) which focused on a number of methods for multilevel analysis and visualization of geographical networks. Regarding road networks, the work of Okabe et al. (2012) is one of the first works which formulates a number of methodological approaches of spatial analysis along networks, transforming traditional spatial measures to this new form of discrete geographical spatial space.

Regarding spatial network paths, the extended use of interconnected portable multi-purpose devices offer a huge amount of almost real time trajectory information available for analysis. Feng & Zhu (2016) review methods and applications on trajectory data mining such as path discovery, location destination prediction, object and movement behavior analysis and several urban services. Chen et al. (2011) attempt to identify the most popular routes based on moving object trajectories while Liu & Schneider (2013) study geographic/semantic similarity of moving object trajectories.

Spatial paths analysis is important especially in the fields of transportation and navigation. Regarding paths similarity over spatial networks, there is a lack of literature.

Path similarity may be useful in a number of contexts such as finding similar routes between two networks for sports training. For example, athletes preparing for participating in athletic marathon events (such as Olympic games) may need to train and study the race path without visiting the actual location where the event will take place. The very same need has been also identified for bicycle athletes who ask for similar paths for training reasons. It is therefore important for athletes of such sports to identify similar routes close to their home residence for training and studding reasons.

3 Spatial Network

Spatial networks are conceptualized in the basis of the scientific field of Graph Theory as sets of Nodes and Vertices. Nodes which represent junctions of streets with single x,y coordinates forming a point in space and connecting one or more vertices. Vertices are connections between two nodes and do not have coordinates. Both nodes and vertices have attributes which describe the characteristics of this specific part of the spatial network.

A sequence of nodes and vertices on a spatial network forms a “walk”. This can be conceptualized as a subgraph of the network which is used for the movement from a source node to a destination node. A walk includes the aggregate characteristics of its individual parts. For example that the total length of a walk is the sum of all the vertices it includes and the total altitude range is the range of altitude information in all nodes included in the walk of the network. The number of possible walks in a network depends on the number of nodes and the amount of connections it includes. The number of walks can be potentially infinite as there are many walks that connect node A with node B. It heavily depends on the available connections and the number of steps required from node A to node B.

Let A denote the adjacency matrix of a spatial network (graph), then, the i,j entry of the matrix. A_n will denote the number of walks of length n from vertex i to vertex j. So for example if we have a 3x3 adjacency matrix, with ones in every position, setting A_4 will result to 27 possible walks of length 4. It is clear that for a fairly complex spatial network of thousands vertices and edges, there may be a very big number of possible walks of n steps.

Similarity between walks between spatial networks can be conceptualized after setting necessary similarity thresholds. These thresholds will form ranges of defining two walks as “similar” based on their geometrical or numeric attributes. The more tight the thresholds of similarity, the more difficult to identify two walks as “similar”. Identifying similar walks on a network can be a time-consuming and computationally expensive task as it requires numerous comparisons of geometrical and numerical attributes between all pairs of possible walks in the network.

In the context of this research there is a need for identifying similar walks between two spatial networks. Athletic teams before joining an athletic event, conduct training in order to acquire better results for the coming athletic event. This training is conducted mainly close to their residence as there are necessary resources (temporal/financial) for relocation to the are where the athletic event will take place. As a result, usually the teams do not train on the very same terrain where the event will take place. This lead to unfamiliarity with the geometrical features of the actual walk of the event which may have consequences for the overall performance of the athletes. There is a need for familiarization with the route of the athletic event without spending extra time and funds. Ideally this would also include a training field close to home-town that will be similar to the actual route of the athletic event in order to use it frequently as a training field.

4 Path similarity algorithm

In order to satisfy the need described above, we developed a spatial algorithm that identifies similar paths between two different spatial networks. The algorithm is based on geographical constrains that form a serial sequence of constrains (filters). The number and type of filters is related to the complexity of the target path which the algorithm aim to identify. The steps of the algorithm are the following:

Step 1: Preparation of spatial network

The algorithm requires a source path and a target network to conduct the spatial search. It is necessary to transform source path and target spatial network in a mathematical graph in order to facilitate mathematical operations.

Step 2: Identification of filters

Next we construct a number of filters which will be used for the spatial search. These filters describe the topographical characteristics of the source path. These filters could be information such the total length of a path, the minimum and maximum altitude, the starting and ending altitude, the existence of a “hill topography” after a specific part of the path, proximity to populated places, proximity to Points Of Interest etc.

Step 3: Recursive Search

Next the algorithm performs a recursive search based on the sequence of filters, eliminating parts of the target network which do not fulfill the criteria set.

Step 4: Multivariate ranking of solutions

Finally, having indemnified a number of possible similar paths in the target network, the algorithm rank the solutions based on a multivariate approach.

The proposed algorithm has been implemented in R programming language along with PostGIS/ PgRouting database.

5 Case study

In order to illustrate the potential of the algorithm, we contacted the local Lesvos Mountain-Bike team which participates in regional athletic events in Greece, for the development of a case study based on their future training schedule and their future participation in regional sport events. A target sport event has been selected which took place in Chios island, GR. The mountain bike team ask for the identification of a path in Lesvos island, that would have the same topographic characteristics as the one they would face in the sport event in Chios island. The athletes, had the need to find a training field close to their home residence in order to train daily without traveling to the target area. This will help them to be prepared for the athletic event and help them achieve better results.

The total length of the target athletic event is 11 km and has been analyzed in a GIS system in order to obtain its geometrical characteristics. We identified the necessary constrains for the searching algorithm. The constrains we have developed for this athletic event along with the acceptable solution margins for each step of the algorithm. The number of potential solutions are the following:

1. Total length of 11km (with a deviation of +- 500m): found 63 candidate results
2. Start-End altitude diff of 600m: found 55 candidate results (rejected 8 solutions)
3. No hill (100m): found 51 candidate results (rejected 4 solutions)
4. Start close to any city (1000m): found 30 candidate results (rejected 21 solutions)
5. Start altitude 0-40: found 16 candidate results (rejected 14 solutions)
6. Stop altitude 600m (+- 50m): found 5 candidate results (rejected 11 solutions)

Finally, the algorithm identified 5 acceptable candidate walks in the spatial network of Lesvos island that fulfill the geometrical characteristics of the athletic event in Chios island. Figure 1 depicts a combined map the two routes (Left: target athletic events, Right: optimal solution in Lesvos island).

Figure 1: Comparison of initial data (left) and results (right) of searching for a bike event that took place in Chios island

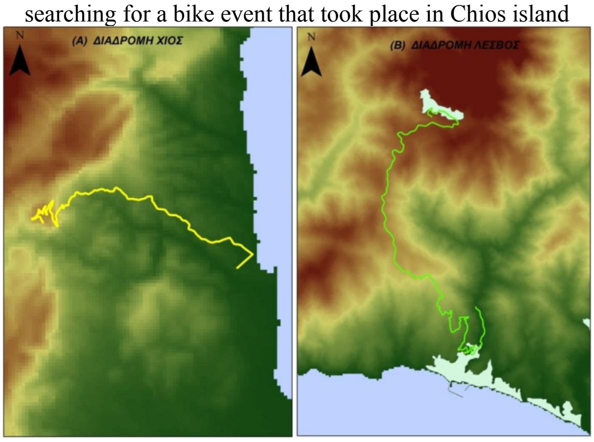

Figure 2: Profile plot of altitude of the target athletic event that took place in Chios island

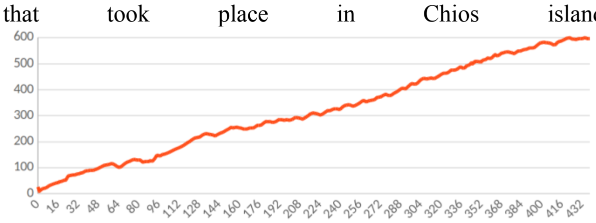

Figure 3: Profile plot of altitude of the result walk found in Lesvos island for training.

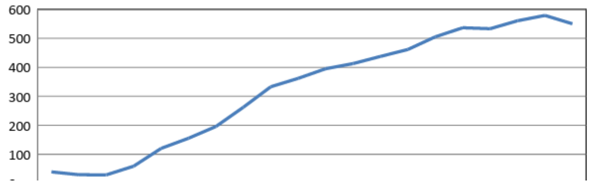

Figure 2 and 3 depict the altitude profile diagram of the target route and the optimal solution respectively. It is important to note that 4th constrain, is based on convenience grounds in order to facilitate the training process and is not related with the targeting event in Chios island. This illustrates the ability to set spatial queries during the search process which lead to acceptable solution, fulfilling not only similarity constrains but also convenience factors. This approach can also be used when planning an athletic event, by setting convenience constrains such as: proximity to touristic activities and preferability of specific areas regarding residence (hotels) or even sightseeing.

6 Results and conclusion

The results of the case study indicate that there are acceptable paths in the source geographic network fulfilling the geometric criteria of the target sport event. The athletes used the results of this work for training and acquired somehow better results. 3 out of 5 participants that used the results of this work acquired better results than previous participation in the same event. Currently we are not able to cross-check if the acquisition of better running times is a result of better training or other physical characteristics of the athletes. We are working along with a specialized team of doctors in order to identify the improvements of each athlete combining athletic sensors' data (hart-rate, effort, etc) along with Geographic Information Systems.

The identification of similar paths in different spatial networks has also some potentials related to touristic development of an area. The identification of known athletic events in Lesvos island, may lead to the development of athletic oriented touristic activity which is a very promising new sector in touristic sciences.

Finally, the very same algorithmic approach can be used in athletic events of running, biking or even car racing as long as the focus of similarity is related to topographic characteristics of the source-target network.

Future steps of this work include the optimization of database searching capabilities by exploring indexing approaches for PostgresDB.

7 References

- Agourogiannis, Panagiotis, Charalampos Lepeniotis, Georgios Tataris, and Dimitris Kavroudakis. 2018. “Geographical Analysis of Road Networks for the Identification of Similar Routes: The Case of Long Race Athletic Events.” In 11th International Conference of the Hellenic Geographical Society, 1–7. Lavrion.
- Chen, Zaiben, Heng Tao Shen, and Xiaofang Zhou. 2011. “Discovering Popular Routes from Trajectories.”

2011 IEEE 27th International Conference on Data Engineering 4 (c): 900–911. https://doi.org/10.1109/ICDE.2011.5767890.

- Demšar, Urška, Olga Špatenková, and Kirsi Virrantaus. 2008. "Identifying Critical Locations in a Spatial Network with Graph Theory." Transactions in GIS 12 (1): 61–82. https://doi.org/10.1111/j.1467-9671.2008.01086.x.
- Feng, Zhenni, and Yanmin Zhu. 2016. "A Survey on Trajectory Data Mining: Techniques and Applications." IEEE Access 4: 2056–67. https://doi.org/10.1109/ACCESS.2016.2553681.
- Kuby, Michael, and Seow Lim. 2007. "Location of Alternative-Fuel Stations Using the Flow-Refueling Location Model and Dispersion of Candidate Sites on Arcs." Networks and Spatial Economics 7 (2): 129–52. https://doi.org/10.1007/s11067-006-9003-6.
- Liu, Hechen, and Markus Schneider. 2013. "Similarity Measurement of Moving Object Trajectories," 19–22. https://doi.org/10.1145/2442968.2442971.
- Okabe, Atsuyuki, and Kokichi Sugihara. 2012. Spatial Analysis Along Networks: Statistical and Computational Methods. 1 edition. Wiley.
- Reggiani, Aura, and Peter Nijkamp. 2009. Complexity and Spatial Networks.
- Rozenblat, Céline, and Guy Melançon, eds. 2013. Methods for Multilevel Analysis and Visualisation of Geographical Networks. Dordrecht: Springer Netherlands. http://link.springer.com/10.1007/978-94-007-6677-8.
- Yamada, Takeo. 1996. "A Network Flow Approach to a City Emergency Evacuation Planning." International Journal of Systems Science 27 (10): 931–36. https://doi.org/10.1080/00207729608929296.