



Understanding Trajectory Data Based on Heterogeneous Information Network Using Visual Analytics

Rui Zhang^{1,2,3}, Wenjie Ma³, Luo Zhong^{3(✉)}, Peng Xie³, and Hongbo Jiang⁴

¹ Hubei Key Laboratory of Transportation Internet of Things, Wuhan University of Technology, Wuhan, Hubei, China

zhangrui@whut.edu.cn

² Hubei Key Laboratory of Inland Shipping Technology, Wuhan University of Technology, Wuhan, Hubei, China

³ School of Computer Science and Technology, Wuhan University of Technology, Wuhan 430072, Hubei, China

15827182791@163.com, zhongluo@whut.edu.cn, xiepeng0715@gmail.com

⁴ School of Electronic Information and Communications,

Huazhong University of Science and Technology, Wuhan 430074, Hubei, China

hongbojiang2004@gmail.com

Abstract. With its continuous development, location information acquisition technology is able to collect more and more trajectory data, and the rich information contained therein is gradually attracting attention from researchers. Trajectory data involves complex relationships among moving objects, time, space, which are hard to understand and be used directly. Nowadays, visual analysis of trajectory data is mainly focus on its representation and interaction, but fails to address the complex correlation contained in trajectory data. Hence, we propose TrajHIN, a heterogeneous information network model built on trajectory data, measure the meta path-based similarity and centrality, and use a visual analytics method to deeply understand trajectory data. The example of visual analysis of real trajectory data has been interpreted and given feedback from domain experts, which proves effectiveness of TrajHIN and feasibility of mining implicit semantic information from trajectory data.

Keywords: Trajectory · Heterogeneous information network · Visual analysis

1 Introduction

Nowadays, with the continuous progress of location information acquisition technology, trajectory data has gradually received public attention and concern. Trajectory data plays an important role in behavioral patterns mining, traffic flow prediction, and POI recommendation, etc. However, trajectory data involves complex relationships between moving objects, time and space, making it difficult to be understood intuitively. Most existing research regards trajectory data as homogeneous information networks.

However, moving objects are related to locations, environment and other things in real-life scenarios, so homogeneous information networks are not suitable for analyzing trajectory data.

Han et al. proposed Heterogeneous Information Networks [1, 2], which are the logical networks involving multiple typed objects or multiple typed links denoting different relations, such as bibliographic networks, social media networks. Heterogeneous information networks can be used to model complex interaction data.

By analyzing trajectory data based on heterogeneous information network, we can get the semantics and information that cannot be mined by many homogeneous information networks. For instance, a meta path of region \rightarrow car \rightarrow region suggests the most frequently used region of taxi, and that the region may be the traffic center during this period. In order to further analyze the underlying relevance in trajectory data, we measure the meta path-based similarity and centrality.

Visualization is desired since it allows the domain users to incorporate their domain knowledge and human intelligence in the exploratory analysis process. However, the scale and complexity of the trajectory data make interactive visualization a challenging task. Some researchers also introduce graph into visual analysis of trajectory data [8], but they fail to pay sufficient attention to the various types of objects and relationships involved in trajectory data, and visualize the high dimensional features of trajectory data. We hope that the implicit correlation information in trajectory data can be displayed to users more clearly. We integrate visualization methods with trajectory data analysis based on heterogeneous information networks so that the information obtained from analysis can be fully utilized.

The main contributions of this paper are as follows:

- We build TrajHIN, a heterogeneous information network model based on trajectory data, it is constructed to model complex correlation of trajectory data and express trajectory data more clearly.
- With TrajHIN, we measure the meta path-based similarity and centrality.
- We integrate the heterogeneity information network model TrajHIN with visual analysis so that users can easily understand and analyze the relationship between corresponding objects and mine correlation information in trajectory data.

The respect of this paper is organized as follows. Section 2 describes the related work. The third section gives the definition and description of our model. The fourth section presents the visualization and some experiments about the method. Section 5 concludes the paper.

2 Related Work

In this section, we explain some other work related to our research, including others' work in trajectory data, brief introduction of heterogeneous information networks and some work in visual analysis.

2.1 Trajectory Data

Many scholars have mined behavior patterns through analyzing and understanding trajectory data. For example, Hirokazu Madokoro modeled trajectory data by using hidden markov model and used behavior patterns in an interest-recommending website [3]. Also, Mahdim Kalayeh proposed a dynamic model of mining behavior patterns from trajectory data [4]. Trajectory data was studied in behavior and path planning that Bucher et al. proposed a path planning algorithm which required less computation based on individual user trajectory log [5]. There are also some research on trajectory data for predicting location, such as taxi status inquiry and waiting time forecasting based on taxi trajectory data [6, 7]. According to current position and historical trajectories of a moving object, predicting location is able to forecast the location of this object [8]. On the other hand, there are some other research on semantic information mining of trajectory data as well. For instance, Liu et al. analyzed the best location for setting up billboard from urban taxi trajectory data [9]. By regarding trajectory data as link relation, Huang et al. constructed an urban road network relationship and analyzed the traffic condition of roads in urban center by the link relations between road sections [10].

2.2 Heterogeneous Information Networks

There are some research on the similarity measurement of heterogeneous information networks. After Han et al. proposed the meta paths for DBLP, the concept of meta paths was widely introduced into similarity measurement on heterogeneous information networks. Subsequently, Han et al. proposed Pathsim, a novel similarity measurement method based on meta path which is able to find peer objects in the network, making it possible to accurately distinguish different latent semantics in heterogeneous information networks [11, 12]. Also, there are some other research on clustering analysis of heterogeneous information networks. For example, Aggarwal et al. used local optimal features to balance heterogeneous information networks which can achieve clustering [13]. In link prediction of Heterogeneous Information Networks, some studies predicted possible relationship between two nodes by using observed links and node attributes [14–16].

2.3 Visual Analysis

Visual analysis related to our work is often focused on two primary aspects. The first aspect is analysis related to graph. For example, Pienta et al. designed a locally adaptive exploration model for it, which is of data graph [17]. Chau et al. developed an interactive visualization system and iteratively improved it to interpret large-scale deep learning models and results [18]. They even presented a novel interactive visual analytics system to explore and comprehend them completely [19]. Another aspect in visual analysis is related to trajectory. One of the most classic applications is proposed by Huang et al. [10], in which they used taxi trajectory data and graph-based visual analysis to study urban network centers. Al-Dohuki et al. put forward SemanticTraj as well, which can be used to link the map and users' semantic information, make users querying much more efficient than

before [20]. However, visual analysis neither related to graph nor trajectory has considered the various types of objects and relationships involved in trajectory data, which may make it not much suitable when dealing with complex relationships. Therefore, our method should take good care of this.

3 TrajHIN Model

In this section, we first constructed a heterogeneous information networks model based on trajectory data. We then described the Pathsim algorithm and measured the meta path-based similarity in Sect. 3.2. In Sect. 3.3, we designed a new degree centrality measure of trajectory data and evaluated meta path-based degree centrality.

3.1 TrajHIN Construction

Trajectory data is data information formed by sampling the movements of a moving object. A trajectory can be seen as a sequence of time-stamped positions. In this paper, the trajectories of ships are used as an example of visual analysis. Specifically, ship trajectories are taken from AIS equipment and include information such as unique identification, position, course, and speed, name of ship, type of ship, destination and timestamp.

Heterogeneous information networks can be denoted by $G = (V, E)$, while V, E are object and link respectively [1]. Each V has a function: $\Psi:V \rightarrow T$, for T is a set of a kind of objects; Each E has a function: $\Phi:E \rightarrow R$, for R is a set of a kind of links. In heterogeneous information networks, $|T| > 1$ or $|R| > 1$. TrajHIN is constructed by extracting the moving objects in trajectory data and related concepts such as time, space and inter-relationship. In this paper, the set of object types includes region, ship and destination while the adjacent, contained and included form the set of relationship types. Region is obtained from geographical coordinates converted by anti-geocoding after trajectory data is de-noised and compressed. TrajHIN model construction is shown in Fig. 1.

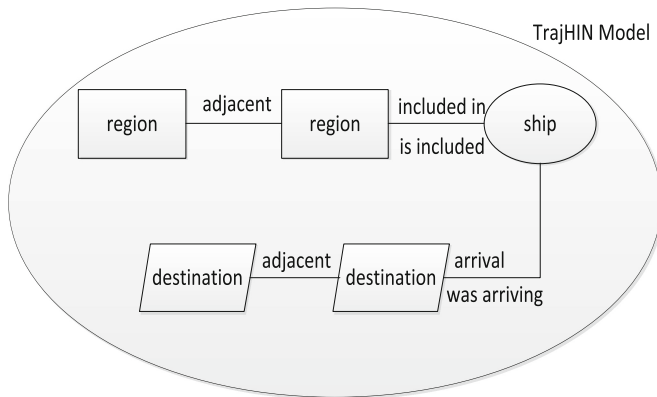


Fig. 1. TrajHIN model.

TrajHIN treats region, ship and destination as different types of objects respectively. In this paper, we mainly examine the following meta paths where a meta path is a path consisting of a sequence of relations defined between different object types:

- *ASDSA* (region *A*, ship *S*, destination *D*)
- *DSASD* (destination *D*, ship *S*, region *A*)
- *SAS* (ship *S*, region *A*, ship *S*)
- *SDS* (ship *S*, destination *D*, ship *S*)

Both similarity and centrality measures use the above meta paths as one of their factors.

3.2 Measuring Similarity in TrajHIN

The settings of heterogeneous information network model TrajHIN and meta paths generate semantic meaning of similarity between objects in trajectory data. For example, similarity of two trajectories is no longer limited to the shape and so on, and we can also mine semantic information through meta path *ASDSA* and measure similarity by analyzing meta path between two objects. Pathsim proposed by Sun [15] can well measure the similarity between nodes in heterogeneous information networks. For example, given a symmetric meta path $P = ASDSA$, Pathsim measures in areas a and b as below:

$$S(a, b) = \frac{2 \times \left| \{P_{a \rightarrow b}: P_{a \rightarrow b} \in P\} \right|}{\left| \{P_{a \rightarrow a}: P_{a \rightarrow a} \in P\} \right| + \left| \{P_{b \rightarrow b}: P_{b \rightarrow b} \in P\} \right|} \quad (1)$$

$Pa \rightarrow b$ refers to a path instance between a and b , $Pa \rightarrow a$, $Pb \rightarrow b$ also represent the paths from a to a and b to b , respectively.

3.3 Measuring Centrality in TrajHIN

Centrality demonstrates a degree that whether a node is in the center of the information network. If a node has directly link with many other nodes, it is more like a center than those nodes which don't have so many links. We studied trajectory data according to measuring meta path-based centrality and designed a new centrality measure of trajectory data in the basis of the heterogeneous network. Given a meta path $P(ASA)$, degree centrality of a node v is the number of entries back to this node along path P . Then, when comparing different graphs, we need to normalize degree centrality. From meta path P , we can see that if the first and the last nodes of the path are in same type, it can be divided by maximum number of possible connections $Num(A) - 1$, where A is the set of points of the same type as point v and $Num(A)$ described those points generated by path P .

4 Visualization

Through the model TrajHIN we constructed and measured the meta path-based similarity and centrality above. Next we conduct visualization of our method. We first designed a interface by integrating TrajHIN with visual analysis in section A and B. In section C and D, we used real trajectory data to explore similarity and centrality in TrajHIN. Then, we interpreted visual analysis of real trajectory data and compared it with feedback from domain experts.

4.1 Interface

We integrate heterogeneous information network based on trajectory data with visual analysis to analyze trajectory data. Functions include: map matching, region selection, graph visualization, similarity query and centrality query, the interface is shown in Fig. 2. Module (1) shows the map; Module (2) displays the trajectory data graph; Module (3) shows the results of measuring meta paths-based similarity and centrality. (4) represents the trajectory data information search module.

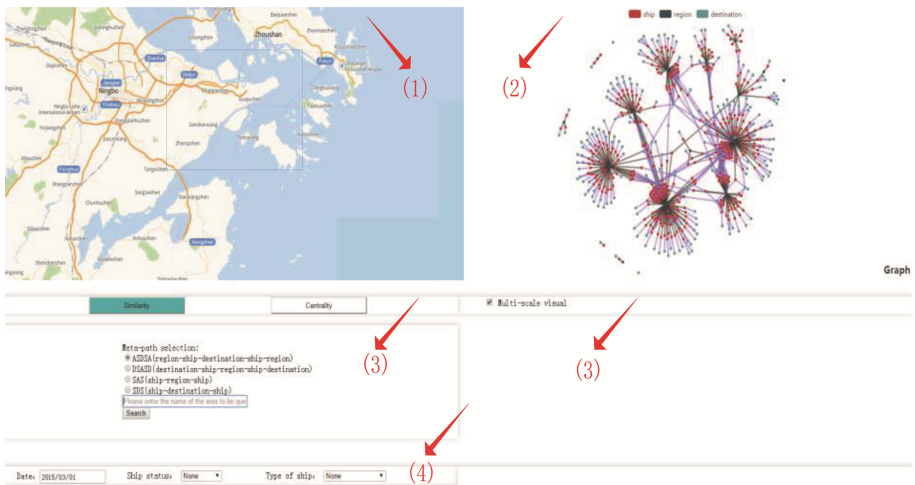


Fig. 2. Visual analysis interface.

4.2 Visualizing TrajHIN

The module shows a graph of heterogeneous information network model constructed on trajectory data, which contains three types of objects: region, ship, destination and different types of edges. In Fig. 3, ship GANGFENG8 is connected to region Ningbo-Ninghai. The graph in this module also has the function of dragging zooming, where different colors of nodes are used to distinguish different types of objects and links.

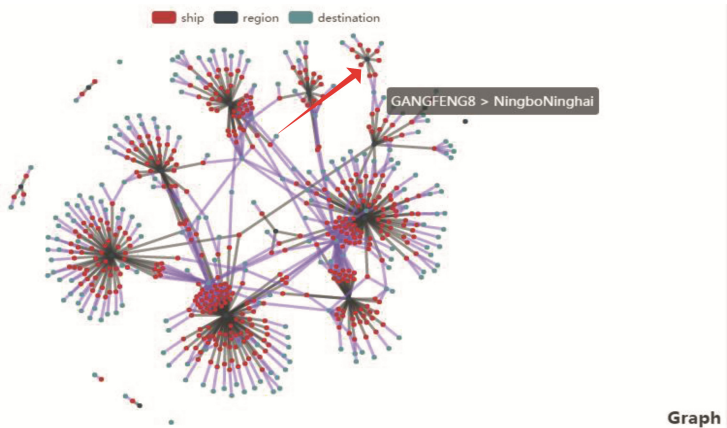


Fig. 3. Graph (Color figure online).

4.3 Exploring Similarity in TrajHIN

By selecting the meta path and inputting the object to be studied, we can display the top-4 objects in the form of histogram. The names of similar objects are shown in the abscissa and similarity measurement scores are shown in the ordinate. The histogram is shown in Fig. 4. The local heterogeneous information network formed by top4 and the object to be studied are shown in Fig. 5.



Fig. 4. Histogram of similarity analysis.

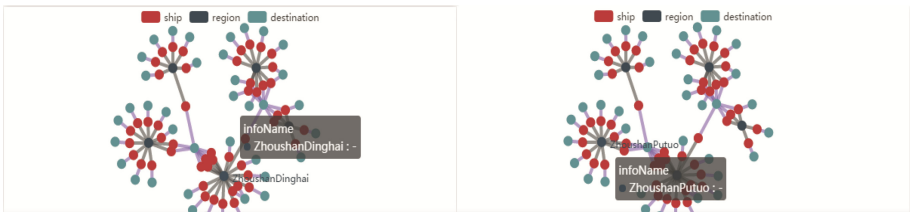


Fig. 5. Similarity analysis graph.

In this example, area is Dinghai, and the research meta path is *ASDSA*. The histogram shows that the similarity between Dinghai and Putuo is the highest by measuring similarity through *ASDSA*. It can also be seen from graph that the number of meta paths of Dinghai → Putuo will be greater than other areas. Through the meta path analysis like *ASDSA*, it can be inferred that the reachability of Dinghai and Putuo is the most similar for some destinations. As confirmed by domain experts, many ships sail through common channels in Dinghai and Putuo, so the two areas are similar.

4.4 Exploring Centrality in TrajHIN

By setting the region and time threshold, the user can draw a line chart reflect the degree centrality with the change of time. Figure 6 shows degree centrality of the area Dinghai on March 1, 2015. By analyzing the meta path *ASA*, we can understand that degree centrality actually refers to the navigation of ships in the area within time threshold. We can draw the semantic result that centrality of the area is the highest in early morning when the time threshold for Dinghai area is set to one hour. This means that Dinghai is the area where fishing vessels work. After doing a field investigation, we find that Dinghai is indeed the scope of fishing vessels activities on that day. It is shown that the improved centrality method can be applied to heterogeneous information networks and obtain semantic information.



Fig. 6. Centrality Results.

4.5 Case Study

We choose Dinghai as the area and observe trajectory data of fishing vessels. Through similarity analysis of Dinghai based on meta path *ASDSA*, it can be inferred that Dinghai and Putuo have the highest reachability on some destinations. By setting time threshold and meta path *ASA*, the degree centrality of Dinghai within a specific day was analyzed. We found that degree centrality of Dinghai was the highest and the area was the most active from 0:00 to 2:00 and 21:00 to 24:00. Therefore, the conclusion is that this area is the scope of fishing vessels activities, which was confirmed by a field investigation. From the results above, we can conclude that the integration of TrajHIN and visual analysis makes it easy for users to understand and analyze relationship between corresponding objects in trajectory data, where semantic information can be mined from trajectory data at the same time.

5 Conclusion

The rapid development of location logging has led to the explosive growth of trajectory data. Meanwhile, the abundant information hidden in trajectory data has drawn a lot more attention. Based on AIS navigation trajectory data, a heterogeneous information network model TrajHIN is constructed and combined with visual analysis. Experimental results have validated the effectiveness of TrajHIN and visual analysis. In the future, we will expand the scale of trajectory data and incorporate the idea of parallel computing into the model to iterate the visual analysis model so that it can be used for visual analysis of large-scale trajectory data.

Acknowledgements. This work was supported in part by the National Natural Science Foundation of China under Grants 61572219, 61502192, 61671216, 61471408 and 51479157; by the China Postdoctoral Science Foundation under Grants 2017T100556; and by the Fundamental Research Funds for the Central Universities under Grant 2015QN073, 2016YXMS297, 2016JCTD118 and WUT:2016III028; by fund of Hubei Key Laboratory of Inland Shipping Technology under Grant NHHY2015005.

References

1. Sun, Y., Han, J.: Mining heterogeneous information networks: a structural analysis approach. *ACM SIGKDD Explor. Newsl.* **14**(2), 20–28 (2013)
2. Deng, H., Han, J., Lyu, M. R., King, I.: Modeling and exploiting heterogeneous bibliographic networks for expertise ranking. In: *Proceedings of the 12th ACM/IEEE-CS joint conference on Digital Libraries*, pp. 71–80. ACM, June 2012
3. Madokoro, H., Honma, K., Sato, K.: Classification of behavior patterns with trajectory analysis used for event site. In: *International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8. IEEE, June 2012
4. Kalayeh, M.M., Mussmann, S., Petrakova, A., Lobo, N.D.V., Shah, M.: *Understanding Trajectory Behavior: A Motion Pattern Approach* (2015). arXiv preprint [arXiv:1501.00614](https://arxiv.org/abs/1501.00614)
5. Bucher, D., Jonietz, D., Raubal, M.: A heuristic for multi-modal route planning. In: Gartner, G., Huang, H. (eds.) *Progress in Location-Based Services 2016*. LNGC, pp. 211–229. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-47289-8_11
6. Luo, W., Tan, H., Chen, L., Ni, L.M.: Finding time period-based most frequent path in big trajectory data. In: *Proceedings of the 2013 ACM SIGMOD international conference on management of data*, pp. 713–724. ACM, June 2013
7. Su, H., Zheng, K., Huang, J., Jeung, H., Chen, L., Zhou, X.: Crowdplanner: a crowd-based route recommendation system. In: *IEEE 30th International Conference on Data Engineering (ICDE)*, pp. 1144–1155. IEEE, March 2014
8. Xue, A.Y., Zhang, R., Zheng, Y., Xie, X., Huang, J., Xu, Z.: Destination prediction by sub-trajectory synthesis and privacy protection against such prediction. In: *IEEE 29th International Conference on Data Engineering (ICDE 2013)*, pp. 254–265. IEEE, April 2013
9. Liu, D., Weng, D., Li, Y., Bao, J., Zheng, Y., Qu, H., Wu, Y.: SmartAdP: visual analytics of large-scale taxi trajectories for selecting billboard locations. *IEEE Trans. Visual Comput. Graphics* **23**(1), 1–10 (2017)

10. Huang, X., Zhao, Y., Ma, C., Yang, J., Ye, X., Zhang, C.: TrajGraph: a graph-based visual analytics approach to studying urban network centralities using taxi trajectory data. *IEEE Trans. Visual Comput. Graphics* **22**(1), 160–169 (2016)
11. Shang, J., Qu, M., Liu, J., Kaplan, L.M., Han, J., Peng, J.: Meta-Path Guided Embedding for Similarity Search in Large-Scale Heterogeneous Information Networks (2016). arXiv preprint [arXiv:1610.09769](https://arxiv.org/abs/1610.09769)
12. Yu, X., Sun, Y., Norick, B., Mao, T., Han, J.: User guided entity similarity search using meta-path selection in heterogeneous information networks. In: *Proceedings of the 21st ACM International Conference on Information and Knowledge Management*, pp. 2025–2029. ACM, October 2012
13. Gupta, M., Aggarwal, C., Han, J., Sun, Y.: *Evolutionary Clustering and Analysis of Heterogeneous Information Networks*. IBM Research Report, 1006-064 (2010)
14. Zhang, J., Yu, P.S., Zhou, Z.H.: Meta-path based multi-network collective link prediction. In: *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 1286–1295. ACM, August 2014
15. Kong, X., Zhang, J., Yu, P.S.: Inferring anchor links across multiple heterogeneous social networks. In: *Proceedings of the 22nd ACM international conference on Information & Knowledge Management*, pp. 179–188. ACM, October 2013
16. Zhang, J., Kong, X., Yu, P.S.: Transferring heterogeneous links across location-based social networks. In: *Proceedings of the 7th ACM international conference on Web search and data mining*, pp. 303–312. ACM, February 2014
17. Pienta, R., Lin, Z., Kahng, M., Vreeken, J., Talukdar, P.P., Abello, J., Chau, D.H.: Seeing the Forest through the Trees: Adaptive Local Exploration of Large Graphs (2015). arXiv preprint [arXiv:1505.06792](https://arxiv.org/abs/1505.06792)
18. Kahng, M., Andrews, P.Y., Kalro, A., Chau, D.H.P.: ActiVis: visual exploration of industry-scale deep neural network models. *IEEE Trans. Visual Comput. Graphics* **24**(1), 88–97 (2018)
19. Pienta, R., Hohman, F., Endert, A., Tamersoy, A., Roundy, K., Gates, C., Chau, D.H.: VIGOR: Interactive Visual Exploration of Graph Query Results. *IEEE Trans. Visual Comput. Graphics* **24**(1), 215–225 (2018)
20. Al-Dohuki, S., Wu, Y., Kamw, F., Yang, J., Li, X., Zhao, Y., Wang, F.: SemanticTraj: A new approach to interacting with massive taxi trajectories. *IEEE Trans. Visual. Comput. Graphics* **23**(1), 11–20 (2017)