Demo Abstract: Visual Analytics of Higher-order Dependencies in Sensor Data

Jian Xu
University of Notre Dame
Department of Computer Science & Engineering
Notre Dame, Indiana 46556
jxu5@nd.edu

Jun Tao
University of Notre Dame
Department of Computer Science & Engineering
Notre Dame, Indiana 46556
jtao1@nd.edu

Nitesh V. Chawla
University of Notre Dame
Department of Computer Science & Engineering
Notre Dame, Indiana 46556
nchawla@nd.edu

Chaoli Wang
University of Notre Dame
Department of Computer Science & Engineering
Notre Dame, Indiana 46556
chaoli.wang@nd.edu

ABSTRACT

Existing approach to model sensor movement data as pairwise connections in networks implicitly assumes the Markov property and loses higher-order movement patterns. While the higher-order network (HON) captures higher-order movement patterns, there has not yet been a visualization tool tailored for HON. Based on our prior work, in this demo we present HoNVis, a comprehensive visualization and interactive exploration framework for HON. We demonstrate domain experts' workflow for analyzing global ship movement sensor data with HoNVis.

CCS CONCEPTS

•**Information systems** →**Network data models;** *Spatial-temporal systems;* •**Human-centered computing** →**Visualization systems and tools;** •**Computer systems organization** →*Sensor networks;*

KEYWORDS

higher-order, sensor network, data mining, visualization

1 INTRODUCTION

In the era of Internet of Things, sensors are generating massive amount of data at an unprecedented speed. For example, in the global ocean shipping network, more than 100,000 ships are now equipped with radio- or satellite-based automatic identification
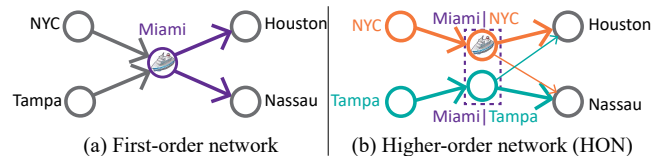
Figure 1: Higher-order networks can more accurately capture higher-order dependencies in sensor movement data.

systems [1], yielding 3-4 millions of movement records per year, according to Lloyd's List Intelligence. However, such big data collected from sensors will only be valuable until properly stored, analyzed, visualized, and interpreted. Recently, networks (or graphs) that are composed of nodes and edges have emerged as a powerful tool in reducing the complexity of the raw data while capturing the main patterns. For example, it is common to represent the global shipping data as a network, with nodes representing ports and edges representing the number of ships sailing between ports [2, 5].

However, the current approach to model sensor movement data as pairwise connections in networks is limited. For example, in the conventional network in Figure 1 (a), it appears that a ship's next movement from Miami does not depend on where the ship came to Miami, implicitly making the Markov assumption to the raw sensor movements data. In reality, a ship is more likely to continue onto Houston if the ship came from NYC to Miami, and more likely to go to Nassau if the ship came from Tampa to Miami. Such *higher-order dependencies* are lost if the raw sensor data are represented, visualized and interpreted using the conventional first-order network. The higher-order network (HON) [4] as illustrated in Figure 1 (b) offers a flexible and powerful representation of the raw sensor data, by breaking down the node *Miami* into higher-order nodes *Miami|NYC* and *Miami|Tampa*.

Despite the importance of HON on sensor network analysis, there has not yet been a visualization tool that is capable of handling the richness of the HON representation. When using the sensor data for decision making, crucial questions such as "where do ports exhibiting higher-order dependencies locate on a map", "how do the connections look like at the country level", and "how do higher-order dependencies influence diffusion dynamics such as invasive species carried through global shipping" cannot be answered with existing tools.
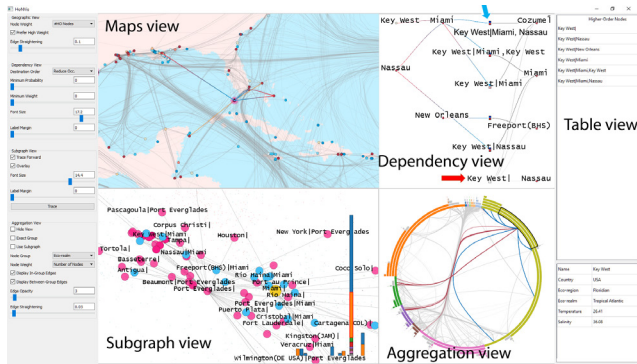
Figure 2: Demonstration of user interface and interactions.

Based on our prior work [3], in this demo we present HoNVis, a comprehensive visualization and interactive exploration framework for HON. It features five coordinated views that enables exploration of HON at different levels of details, and provides a solution to the key challenge of visualizing variable orders of dependencies in networks. The visualization of the global shipping sensor data is used for case study throughout this demo as an important application for devising effective species control strategies. We invited two domain experts from the fields of network science and marine ecology to use HoNVis to explore the shipping sensor data, who performed a case study different from that in [3], and reported critical insights that are only available through HoNVis. Meanwhile, we note that HoNVis is generalizable for other types of sensor data such as taxi trajectories collected by GPS, human movements and interactions collected by cell phones, and so on, facilitating the visualization and interpretation of the increasingly complex IoT data.

2 DESIGN

The design of HoNVis focuses on the unique structural features of HON, i.e., the formation and impact of higher-order nodes, including why a higher-order node exists in a HON, where do they locate, and how do they influence network dynamics? Specific to the shipping network case study, we aim to answer these questions through a three-step exploration process: 1) global identification of ports of interest, 2) detailed observation of the connections of an individual port, and 3) tracing the propagation of invasive species from port to port through shipping. The input trajectories data collected from sensors are converted directly to the conventional first-order network (FON); meanwhile, higher-order dependencies are extracted to construct the HON using the algorithm in [4]. By aligning these two network representations, and mapping them geographically, users can quickly identify structures of interest.

As shown in Figure 2, the visualization includes five coordinated views: *map view* and *table view* show information related to a single node; *dependency view*, *subgraph view*, and *aggregation view* show connections among multiple nodes. Together these five views enables the exploration of higher-order nodes and their dependencies at different levels of detail.

3 DOMAIN EXPERT'S WORKFLOW

Data import. The domain expert first imports diverse types of sensor data for this case study. The global ship movement data are collected by Lloyd's List Intelligence using sensors of the automatic identification system, which contain 3,415,577 individual ship voyages corresponding to 65,591 ships that move among 4,108 ports globally between May 1, 2012 and April 30, 2013. The expert also imports metadata of ships, such as ship type, voyage start and end time, ship size, as well as metadata of ports such as coordinates, country, environmental conditions (temperature and salinity), and eco-region information.

Identifying ports with higher-order dependencies. The expert wanted to investigate species invasion from Florida to Mexico. On the map view, the expert chose to highlight ports that demonstrate higher-order dependencies, and quickly identified the port Key West. Upon clicking, shipping routes associated with Key West are highlighted in the map view, including the route to Cozumel in Mexico.

Exploring higher-order dependencies. The expert was then intrigued by what higher-order dependencies are associated with Key West. In the dependency view, every curve indicates a shipping route, with the nodes on the left indicating ports prior to the arrival at Key West, and the rightmost column of nodes showing the potential destinations. With the edge darkness indicating the transition probabilities, the expert noticed that ships are equally likely to go to Cozumel in Mexico and other ports such as Miami in the U.S. if we do not know where the ships came to Key West. However, as indicated by the blue arrow, the path $Nassau \rightarrow Miami \rightarrow KeyWest$ will lead to Cozumel with high certainty. The expert commented that instead of the ambiguous information like "Key West might bring invasive species to Cozumel, Miami, Freeport and Nassau" made available by conventional network models, information such as "ships from Miami to Key West will only bring invasive species to Cozumel" made available by HON are crucial when devising species control strategies that target on specific routes.

Exploring the influence of higher-order dependencies on diffusion. The expert further wanted to explore how the higher-order shipping route to Key West will influence the propagation of invasive species. The expert selected the path $Nassau \rightarrow Miami \rightarrow KeyWest$ in the dependency view, and the corresponding higher-order node is highlighted on the network view, indicating that the node has been "infected". The expert then clicked "trace" multiple times to observe how the invasive species will propagate given this pathway. It is observed that the Port Everglades and Miami are central to the propagation of invasive species. Through the aggregation view, the expert further understands the connections between ports at a higher level. Therefore, the expert has gained the knowledge to devise port-targeted control policies.

REFERENCES

[1] International Maritime Organization 2012. *International Shipping Facts and Figures - Information Resources on Trade, Safety, Security, Environment.* International Maritime Organization.
[2] P. Kaluza, A. Kölzsch, M. T. Gastner, and B. Blasius. 2010. The Complex Network of Global Cargo Ship Movements. *Journal of the Royal Society Interface* 7, 48 (2010), 1093–1103.
[3] J. Tao, J. Xu, C. Wang, and N. V. Chawla. 2017. HoNVis: Visualizing and Exploring Higher-Order Networks. In *Proceedings of IEEE Pacific Visualization Symposium.*
[4] J. Xu, T. L. Wickramarathne, and N. V. Chawla. 2016. Representing Higher-Order Dependencies in Networks. *Science Advances* 2, 5 (2016), e1600028.
[5] J. Xu, T. L. Wickramarathne, N. V. Chawla, E. K. Grey, K. Steinhaeuser, R. P. Keller, J. M. Drake, and D. M. Lodge. 2014. Improving Management of Aquatic Invasions by Integrating Shipping Network, Ecological, and Environmental Data: Data Mining for Social Good. In *Proceedings of ACM SIGKDD.* 1699–1708.