Detecting Anomalies in Sequential Data with Higher-Order Networks

Mandana Saebi
University of Notre Dame
msaebi@nd.edu

Jian Xu
Citadel LLC
i@jianxu.net

Lance M Kaplan
U.S. Army Research Lab
lance.m.kaplan.civ@mail.mil

Bruno Ribero
University of South Florida
ribeiro@cs.purdue.edu

Nitesh V Chawla
University of Notre Dame
nchawla@nd.edu

Complex systems, represented as dynamic networks, comprise of components that influence each other via direct and indirect interactions. A trending line of research has shown the importance of using Higher-Order Network (HON) models for analyzing such complex systems, as the typical Markovian assumption in developing the First Order Network (FON) is highly limiting. This higher-order network representation not only creates a more accurate representation of the underlying complex system, but also leads to more accurate network analysis. In our prior work [1] we developed (`BuildHON`) for extracting higher-order dependencies from sequential data to build the Higher-Order Network (HON) representation. `BuildHON`, although accurate, faced the challenges of computational complexity as well as parameter dependency. In this work, we address the limitations of our prior work by proposing a scalable and parameter-free algorithm, `BuildHON+` for timely and accurate extraction of higher-order dependencies from sequential data. Given `BuildHON+`, we ask the following research question: *Does incorporating higher-order dependencies improve the performance of existing network-base methods for detection anomalous signals in the sequential data?*

To answer the above question, we evaluate the performance of `BuildHON+` vs `BuildHON` for detecting anomalies in a real-world taxi trajectory data. We use the ECML/PKDD 2015 challenge data, which contains one year of all the 442 taxi GPS trajectories in Porto, Portugal. We map all coordinates to the nearest 41 police stations. We compare the 52 networks for `BuildHON`, with *MaxOrder* of 2 (indicated as HON-2) and `BuildHON+` with no *MaxOrder*, (indicated as HON+) in Fig. 1. We compute the graph distances (using weight distance) for neighboring time windows. Fig. 1 shows the comparison of HON+ and HON-2. While the trend of HON+ resembles that of HON-2, the graph distances in Week 44 are particularly more significant in HON+ than HON-2 . We notice that Porto's second most important festival, "Burning of the Ribbons", lasts from May 4 to May 11 in 2014 and falls within Week 44 of our study. The festival involves parades, road closures, and is popular among tourists, which could be the underlying reason to the changes in taxis' movement patterns. We notice that multiple higher-order nodes and edges emerge in Week 44, indicating the emergence of higher-order traffic patterns. Note that, FON's performance in capturing this higher-order anomaly is worse than both HON-2 and HON+. Therefore, it is not included in the final results.

Finally, we compare the running time and memory consumption of `BuildHON+` and `BuildHON` on the global ship movement data which is known to exhibit dependencies beyond fifth order. It consists of up to three years of shipping data, aggregated over

Anomaly Detection in Finance (ADF'19), August 4-8, 2019, Anchorage, Alaska, USA

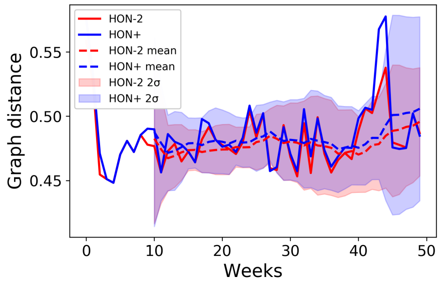

Figure 1: The anomalous traffic pattern is noticeable in HON+ (c), while it is not clear from HON-2.

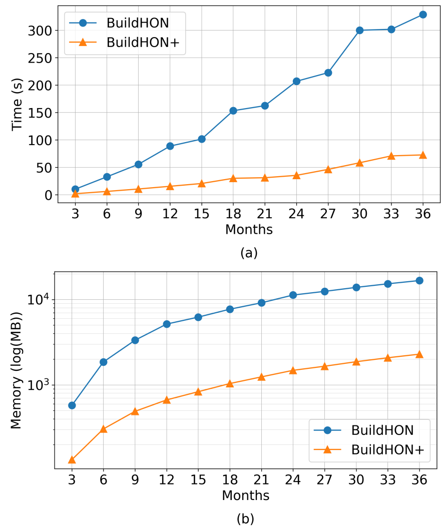

Figure 2: For the maximum data size, `BuildHON` takes 4.5 times longer than `BuildHON+` to run (a), and requires approximately 7.2 times more memory than `BuildHON+` (b)

3-months intervals. In this experiment, `BuildHON+` is parameter-free, and for `BuildHON` we set *MaxOrder* = 15. Fig. 2 illustrate the time and memory usage of both algorithm as the size of the data increases. We observe that `BuildHON` is highly sensitive to the size of the data. For the maximum data size, `BuildHON` requires approximately 7.2 times more memory than `BuildHON+` and takes 4.5 times longer to run.

In conclusion, we presented a scalable and parameter-free algorithm for extracting higher-order dependencies from sequential data, and demonstrated the success of higher-order network modeling for anomaly detection in dynamic networks.

REFERENCES

[1] Jian Xu, Thanuka L Wickramarathne, and Nitesh V Chawla. 2016. Representing higher-order dependencies in networks. *Science advances* 2, 5 (2016), e1600028.