ABSTRACT

Using cross-correlation networks to identify and visualize patterns in disease transmission

T-C Lu and N Abernethy

Biomedical and Health Informatics, School of Medicine, University of Washington, Seattle, WA, USA
E-mail: neila@uw.edu

Objective

Time series of influenza-like illness (ILI) events are often used to depict case rates in different regions. We explore the suitability of network visualization to highlight geographic patterns in this data on the basis of cross-correlation of the time series data.

Introduction

Syndromic surveillance data such as the incidence of influenza-like illness (ILI) is broadly monitored to provide awareness of respiratory disease epidemiology. Diverse algorithms have been employed to find geospatial trends in surveillance data, however, these methods often do not point to a route of transmission. We seek to use correlations between regions in time series data to identify patterns that point to transmission trends and routes. Toward this aim, we employ network analysis to summarize the correlation structure between regions, whereas also providing an interpretation based on infectious disease transmission.

Cross-correlation has been used to quantify associations between climate variables and disease transmission.[1,2]

The related method of autocorrelation has been widely used to identify patterns in time series surveillance data.[3] This research seeks to improve interpretation of time series data and shed light on the spatial–temporal transmission of respiratory infections based on cross-correlation of ILI case rates.

Methods

For this pilot study, we analyzed patient visits to health care providers for ILI, collected through the US Outpatient Influenza-like Illness Surveillance Network (ILINet). Aggregate data for the 27-week period from the 35th week of 2009 to the 9th week of 2010 were used. The model involves the 10 Human and Health Services (HHS) regions for which ILI data are publicly available through the CDC. The data consist of the weighted percentage of all patient visits to healthcare providers for ILI reported each week. Additional networks were generated using confirmed cases from the 2009 H1N1 pandemic and city-level data from Google Flu Trends.[4] Using the cross-correlation function to measure maximal correlation and the corresponding latency

Figure 1 Cross-correlation network representing fraction of ILI cases in the 10 HHS regions during the 2009–2010 influenza season. The corresponding latency between time series is represented by directed ties in the graph. Gray (bidirectional) edges represent regions correlated with no time lag.

between each pair of regions, the resulting matrix was visualized with the R statistical package.

Results

We demonstrate this technique on several data sources. An example network (Figure 1) shows ILI correlations between HHS regions. For clarity, a minimum threshold was used to highlight regions having the most similar rates. Regions four and nine are disconnected from the core network because of weak correlation with other regions. These outliers reflect an early peak (region four) and weak transmission (region nine) of ILI cases in hospital visit data. Other observations from the network are discussed. We demonstrate how this visualization motif might enhance existing tools such as DISTRIBUTE.

Conclusions

Network analysis has a role in several aspects of infectious-disease modeling, including simulation, contact investigation, and sampling. We have proposed a new technique to interpret the correlation between case rates in geographic regions. This method may help epidemiologists to quickly visualize the similarities in case rates between regions. In future work, we plan to validate the ability of these networks to reveal patterns in disease transmission. Comparison of cross-correlation networks to transmission routes such as airline networks or US Census worker flow data may reflect the route of transmission of new outbreaks or diseases such as SARS. The networks may then serve as a basis to evaluate intervention options during outbreak of pandemic influenza or other emerging infectious diseases.

Acknowledgements

References

1 Naish S, Hu W, Nicholls N, Mackenzie JS, McMichael AJ, Dale P, *et al.* Weather variability, tides, and Barmah Forest virus disease in the Gladstone region, Australia. *Environ Health Perspect* 2006;**114**:678–83.

2 Zhang WY, Guo WD, Fang LQ, Li CP, Bi P, Glass GE. Climate variability and hemorrhagic fever with renal syndrome transmission in Northeastern China. *Environ Health Perspect* 2010;**118**: 915–20.

3 Espino JU, Hogan WR, Wagner MM. Telephone triage: a timely data source for surveillance of influenza-like diseases. *AMIA Annu Symp Proc* 2003; 215–19.

4 Valdivia A, Lopez-Alcalde J, Vicente M, Pichiule M, Ruiz M, Ordobas M. Monitoring influenza activity in Europe with Google Flu Trends: comparison with the findings of sentinel physician networks—results for 2009–10. *Euro Surveillance* 2010;**15**, pii: 19621.