# Bayesian Network Data Fusion Visualization

**Charles Hodanics, BS**
*Johns Hopkins University Applied Physics Laboratory, Laurel, MD*

## OBJECTIVE

This paper describes the use and visualization of Bayesian Networks (BN) to better assists public health users. The Data Fusion Visualization (DFV) provides an intuitive graphical interface that supports users in three ways. The first is by providing a seamless drill down interpretation of a dataset. The second is by providing an intuitive interpretation of BN. Finally, by abstracting the visualization from the underlying model, the DFV is capable of masking inter-operating BNs into a single visualization. The DFV provides a graphical representation of BN Network Data Fusion.

## BACKGROUND

A BN (Bayesian Network) is a probabilistic graphical model representing dependencies and relationships. The structure of the network and conditional probabilities capture an expert's view of a system. BN have been applied to the public health domain for research purposes [1,2], but have not been used directly by the end users of public health systems. As BN technology becomes more and more accepted in the public health domain, the data fusion visualization becomes a critical component of the overall system design. The tools developed utilize computer assisted analysis on BN in the public health domain, provide a concise view of the data for better decision support, and shorten the decision making phase allowing rapid dissemination of information to public health.

## METHOD

The DFV provides screens with several charts and various tools for user interaction. The first screen includes a time series graph that shows the probability of a public health event in a selected jurisdiction for each day. The time series graph is annotated with a color scheme highlighting potential events. A potential event is the BN detection of epidemiological significance. The DFV allows the user to further investigate a detected event. The resulting second layer of screens presents a node-based graph visualization that shows hierarchical drill down. For example, the National Capital Region network can be presented as the parent node NCR with child nodes DC, Maryland and Virginia. Likewise, a parent node for the state Virginia can be represented as the parent node VA with child nodes for the different counties in Virginia. On each of these node graphs, the user is able to investigate further into the data visually by opening a leaf node. The leaf nodes in the graph are entry nodes into deeper data visualizations. For example, by clicking on the Montgomery County node, the DFV will open a new screen with Montgomery County as the parent node and intermediate nodes that represent the decision making process implemented in the BN. The county level screens visualize decision support logic implemented in the BN. The behavior of the system is transparent and intuitive to the decision makers and allows them to accept or decline the system's recommendation in a timely manner. The county level screen provides data grouped by different criteria. As an example, influenza detection BNs show that both age group and Military/Civilian ill population distributions as important criterion. This provides visual summaries of more specific details about data distribution in case of public health event. From county level screens the user can drill down to actual data records that are contributing to the event detection. The set of data can now be researched and further narrowed based on specific dates of inquiry.

## CONCLUSION

From the various graphs in this visualization, a user is capable of walking through an event of interest from a higher level perspective down to actual data occurrences that may produce this event. The concept of taking a high-level alert and finding data that caused this alert will be very useful in research. The user is also able to do this investigation intuitively without BN experience. Data fusion techniques are transparent to the user. Future scenarios will give advanced users the capability to change underlying Bayesian networks and apply different data fusion techniques. They may also change algorithms for event notification on the fly. The user will also have the capability to provide simulated data and observe the effects of the new data in the higher level views of the system.

## REFERENCES

1. Z. Mnatsakanyan, S. Murphy, R. Ashar, H. Burkom: Hybrid Probabilistic Modeling and Automated Data Fusion for Biosurveillance Applications. PHIN 2006 Atlanta, GA.
2. Weng Keen Wong, G. Cooper, D. Dash, J. Levander, J. Dowling, W. Hogan, M. Wagner: Use of Multiple Data Streams to Conduct Bayesian Biologic Surveillance. MMWR 2005, Vol 54(suppl); 63-69.

Further Information:
charles.hodanics@jhuapl.edu