Modeling Ebola Spread Using Big Data Analytics

A model of Ebola spread was developed using innovative big data analytics techniques and tools. Massive amounts of data were used from various sources including Twitter feeds, Facebook and Google. This data was then fed into a decision support system. It models the spread pattern of the Ebola virus and creates dynamic graphs and predictive diffusion models based on the outcome and impact on either a specific person or on entire communities.

This CAKE research created computational spread models for Ebola that will potentially lead to more precise forward predictions of disease propagation. The tool will help identify individuals who are possibly already infected. It is capable of performing trace-back analysis to locate the possible sources of infection for particular social groups. Working with Florida International University researchers and other partner universities, the Florida Atlantic University research team also collaborated with LexisNexis Risk Solutions (LN); a leading big data company and CAKE member. LN provided large amounts of data about relationships among people in the United States. Researchers used data analytics and tools to model disease spread patterns. The project performed modeling, analytics, and development of a Decision Support System (DSS), that calculates probabilistic outcomes of Ebola impact on either specific persons or communities at a specific locations. This information is then fed to FIU's Terafly system for geospatial mapping and other services.

Jointly with the LN research team, people clusters were created based on proximity and a model using weighted scores was built to approximate physical contacts. In creating people clusters, researchers used public record graphs to calculate distances between an affected person and his/her relatives and friends. Based on this model, the disease propagation paths were developed.

As part of this research, the NewsCubeSum was developed. This is a personalized multidimensional news update summarization system. It collects data from news articles and Twitter. The system utilizes OLAP (OnLine Analytical Processing) and supervised sentence selection techniques to generate brief summaries. It delivers news summaries in multiple dimensions, such as time, entity, and topic. This project demonstrated how this system can be used in improving situational awareness during disease outbreaks.

Tracking and containing Ebola requires enormous resources. This system provides a proactive approach to reasonably reduce the risk of exposure of Ebola spread within a community or a geographic location. This work represents an improvement over previous state-of-the-art, because it used innovative data analytics techniques and the latest HPCC Systems technology to developed models of Ebola spread. With information from multiple sources indicating infected individuals and their personal relationships and social groups, dynamic graphs can be created, and predictive diffusion models can be used to study key issues of Ebola epidemics, e.g., location, time and number of expected new cases. The two fundamental diffusion models are Independent Cascade Model (IC) and Linear Threshold Model (LT). Both models follow an iterative diffusion process wherein infected nodes infect their uninfected neighbors with certain predictable probabilities. Based on fundamental models, advanced propagation models were developed to estimate an influence function by examining past and newly infected notes and predict subsequent infections.

The DSS allows the individual to enter their specific information through a Web-application such as travel information. The mobile interface application would automatically extract the geo-coordinates of the individuals. This would allow the system to intelligently query the movement of the person and a possible con-

tact in the areas affected by Ebola or in the areas affected by Ebola. For instance, if found that a person affected with Ebola was in a theater the previous evening, then a monitoring alert could be issued to the other people within the theater to take extract precaution and watch any sign of Ebola impact such as fever or headache.

The mobile interface will further allow the people to enter the signs of Ebola such as fever specifically above 100.4^O, chills, headache and vomiting, myalgia, intense weakness. While these signs are very common with other diseases as well, such as Malaria and Typhoid, laying over these impacts on the geo-coordinates of having the person in a geographic proximity of a person in a community with impact of Ebola, will provide more focused results with higher accuracy.
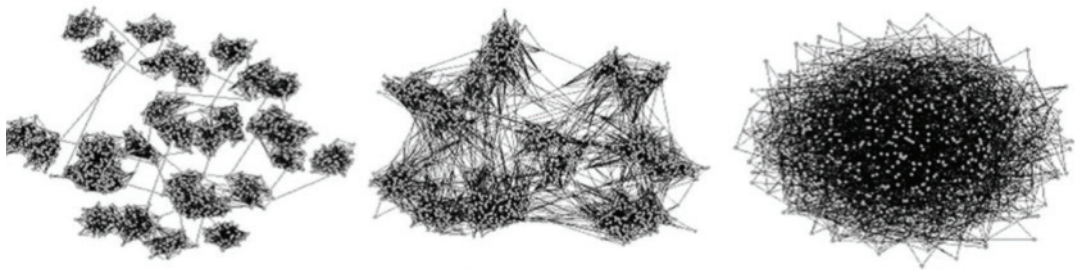
Social network analysis types showing increasing risks of Ebola propagation. Image provided by CAKE.

The DSS will also interface with the social groups of the person. This allows the extraction of relationship maps to predict the higher probability of Ebola spread within specific communities. The social group integration such as a LinkedIn or a Facebook application programmable interface (API) can extract such data very accurately thereby providing high precision and accuracy in risk prediction.

> **Economic impact**: The proposed methodology, including coalition-building efforts, supports solutions to a wide range of other public health issues. The economic impacts can be tremendous by predicting outbreaks of the deadly Ebola virus (or any other epidemics) and directing potential victims to the nearest suitable medical facility. This research can also help in indicating the areas in which new facilities should be opened, where disease outbreaks are beginning to occur, and how they are likely to expend.

For more information, contact Borko Furht at Florida Atlantic University, bfurht@fau.edu, Bio http://www.cse.fau.edu/~borko/, phone 561.297.3180.