Big Data Opportunities: Use of Social Media Analytics for Forecasting Infectious Diseases.

L. Leticia Ramirez-Ramirez*
Instituto Tecnologico Autonomo de Mexico (ITAM), Mexico City, Mexico –
lilialeticia.ramirez@itam.mx

Yulia Gel
University of Dallas at Texas, Dallas, USA – ygl@utdallas.edu

J. Octavio Gutierrez-Garcia
Instituto Tecnologico Autonomo de Mexico (ITAM), Mexico City, Mexico –
octavio.gutierrez@itam.mx

Mark Dredze
Johns Hopkins University, USA – mdredze@cs.jhu.edu

Patrick Thompson
University of Dallas at Texas, Dallas, USA – patrickthompson03@gmail.com

Many currently employed models for prediction dynamics of infectious diseases suffer from inclusion of data elements that are not universally available in real time and, most importantly, from the lack of observations for near real-time forecasting. This in turn substantially reduces practical utility of these prediction models. (For instance, CDC data for well monitored areas in USA are two weeks behind, as it takes time to confirm ILI as flu). In this talk we propose an innovative individual-specific early warning and awareness platform for real-time tracking and forecasting of infectious diseases that coherently harnesses informational wealth from various traditional and from novel data sources such as social media (e.g. Twitter, Weibo, etc) and individual socio-demographics and social interaction patterns.

Our proposed procedure to do real time forecasting is based on information on confirmation cases (available until two weeks prior to the present date) and social media data, such as twitter (available until present date). We use the information on twitter to describe the confirmed cases that are unavailable. This proxy information is used to initialize Monte Carlo simulations that evolve in a social network previously constructed to represent the most important contact characteristics in the population.

We illustrate our approach in forecasting influenza 2012-2013 in the Region of Peel (Ontario, Canada). We discuss the potential extension of the proposed methodology to forecasting infectious diseases such as Chikungunya.

This is where the abstract is placed. It should include a statement about the problem being addressed in the presentation (and paper, if submitted). Continue with a discussion of why it is important to address this problem. This may be followed by some summary information about the models and methods developed and/or used to address the problem. Conclude with a description of the key results and contributions that will be covered in the presentation (and paper).