Forecasting diseases using Wikipedia

November 13 2014

Analyzing page views of Wikipedia articles could make it possible to monitor and forecast diseases around the globe, according to research publishing this week in *PLOS Computational Biology*.

Dr Sara Del Valle and her team from Los Alamos National Laboratory successfully monitored influenza outbreaks in the United States, Poland, Japan and Thailand, dengue fever in Brazil and Thailand, and tuberculosis in China and Thailand.

The team was also able to forecast all but one of these outbreaks (tuberculosis in China) at least 28 days in advance. The results suggest that people start searching for disease-related information on Wikipedia before they seek medical attention.

The paper shows the potential to transfer models across different regions; that is, one can "train" a computer model using public health data in one location and implement the model in another region. For example, researchers could create models using data from Japan to track and forecast disease in Thailand. This is particularly important for countries that do not offer reliable disease data.

Sara Del Valle says: "A global disease-forecasting system will change the way we respond to epidemics. In the same way we check the weather each morning, individuals and public health officials can monitor disease incidence and plan for the future based on today's forecast. The goal of this research is to build an operational disease monitoring and forecasting system with open data and open source code. This paper

shows we can achieve that goal."

More information: Generous N, Fairchild G, Deshpande A, Del Valle SY, Priedhorsky R (2014) Global Disease Monitoring and Forecasting with Wikipedia. *PLoS Comput Biol* 10(11): e1003892. DOI: 10.1371/journal.pcbi.1003892

Provided by Public Library of Science