Better predicting flu outbreaks with Wikipedia

May 15 2015

From Left: Kyle Hickmann, Nick Generous, Geoffrey Fairchild, James Hyman (back), Alina Deshpande and Sara Del Valle (front).

Scientists at Los Alamos National Laboratory have the ability to forecast the upcoming flu season and other infectious diseases by analyzing views

of Wikipedia articles. "The ability to more accurately forecast the flu season and other infectious diseases will transform the way health departments prepare for and respond to epidemics, ultimately saving lives," scientist Sara Del Valle said.

Del Valle and her team recently published "Forecasting the 2013-2014 Influenza Season using Wikipedia," in the *Public Library of Science*.

"Infectious diseases are one of the leading causes of morbidity and mortality around the world; because of this, forecasting their impact is crucial for planning an effective response strategy," Del Valle said.

According to the Centers for Disease Control and Prevention, seasonal influenza effects up to 20 percent of people in the United States and causes major economic impacts resulting from hospitalization and absenteeism.

Understanding influenza or other infectious disease dynamics and forecasting their impact is fundamental for developing prevention and mitigation strategies. To do this, researchers at Los Alamos combined modern data assimilation methods with Wikipedia access logs and CDC influenza-like illness (ILI) reports to create a weekly forecast for seasonal influenza.

"We used techniques often seen in weather forecasting to iteratively tune a model of influenza dynamics based on Wikipedia observations so that our forecast agrees with the most current ILI data," said researcher Kyle Hickmann, who is the lead author of the paper and a member of Del Valle's team.

These methods were applied to the 2013-2014 influenza season but are standard enough to forecast any disease outbreak, given incidence or case count data, the team believes. Del Valle and her team adjusted the

initialization and parameterization of a disease model, which allows them to determine systematic model bias. They also provided a way to determine where the model diverges from observation, and evaluate forecast accuracy.

Wikipedia article access logs are shown to be highly correlated with historical ILI records and allow for accurate prediction of ILI data several weeks before it becomes available. The results showed that prior to the peak of the flu season, their forecasting method projected the actual outcome with a high probability.

"Disease forecasting is still in its infancy and there is much more to learn in this field," Del Valle said. "We are continuing to refine our approach so our forecasts can be used for actionable decision-making."

More information: "Forecasting the 2013–2014 Influenza Season Using Wikipedia." *PLoS Comput Biol* 11(5): e1004239. DOI: 10.1371/journal.pcbi.1004239

Provided by Los Alamos National Laboratory