

infectious disease specialists, CDC investigators, and public health officials. Since its inception, the EIN listserv has served over 2800 discussions on the identification of new infectious diseases, treatments, and policy implications. Identifying important topics of discussion on the EIN is currently ad hoc, done a list administrator reading all discussions. There is significant interest in improving the accuracy and timeliness with which this important information is identified so that it can be distributed to the CDC and other healthcare organizations.

There are a number of approaches that have been taken previously to improve disease surveillance methods. Polgreen et al. [95] considered the problem of finding optimal placement to increase coverage of an influence surveillance network. They show that maximum coverage models can greatly increase the coverage level for the state of Iowa. Polgreen et al. [97] have considered the use of healthcare prediction markets, emerging from economics [42], to give timely predictions based on healthcare related forecasting. Other recent work has focused on using the collective wisdom of crowds to track disease outbreaks using search engine queries [45, 93]. During the H1N1 outbreak in 2009, a number of projects considered the use of Twitter posts to track the spread of the infection [101].

Our solution to improve disease surveillance is to develop a simple procedure for identifying discussions on the EIN that have the potential to become “important,” and ignore threads that will remain “unimportant.” To solve this problem we leverage the social network of individuals and their participation in threads, based on historical EIN data, and identify a set of “bellwether” users who typically participate in the