Matt Hemmings, V00494368
Ryan Schafer, V00637085
Adrian Kilian, V00762884

Term Project 1 - Epidemiological Spread Modeling

History

Human history is filled with death due to spread of disease. One of history's most notorious killers is Smallpox, which killed an estimated 300 to 500 million people in the past 12,000 years and spread through inhalation of the virus. In the 6th century, a plague spread throughout Asia and Europe. After approximately 200 years, had killed an estimated 25 million people, which was half of Europe's population. Other diseases may have different mechanisms for dispersal, such as through blood or contact, but the spread can be as effective. As humans cluster in larger communities, the imperative to track and contain the spread of disease becomes paramount. With the globalization of society, an unrestrained disease could spread through multiple continents and decimate populations.

Widespread use of antibiotics has led to a culling of the weakest members of the bacterial communities and explosion of antibiotic resistant bacteria known colloquially as "Super Bugs." The spread of antibiotic resistant bacteria is a threat recognized by the Centers for Disease Control (CDC) in the United States. They note that 23,000 people die annually from super bugs and suggest that doctors and hospitals work to prevent the spread of infection and be more conservative in their use of antibiotics.

Project

With the idea of preventing the spread of infection in mind, our team project leverages the ubiquity and power of internet searches to determine the locations of those who may be experiencing symptoms. The information resulting from our project could be used to determine likely epicenters for disease and enable those interested in the welfare of the populace, including family members, the governmental health organizations, and pharmaceutical companies to actively work to limit the risk of a disease spreading. Currently, the project focuses on a made-up disease known as "Raptor flu," which we opined may have be the ancestor for avian flu. Raptor flu, however, turns the sufferer into a Velociraptor including growing claws and a tail. While this disease may be fictional, we have modularized the code to a point that we can change the terms we look for and expand upon our concept.

For our project, we designed a system as an expandable proof of concept of an application that imports internet searches and location data and outputs a visual display (a map) of relevant searches. We implemented techniques covered in our Data Dilemmas in Cloud Computing class, such as Mapreduce and sentiment analysis. Currently, the project focuses on British Columbia, however, it could be expanded to Canada, North America, or even the world.

The project produces a result by the following procedure:

1. Searches are currently generated from a set of predetermined URL strings and randomized locations before being saved into a JSON file
2. The JSON file is parsed

3. For each entry the IP address is sent to be queried and the URL is sent for sentiment analysis

4. The returned location and sentiment are mapped and sent to the reducer for output

5. The output emitted into gephi and a graph is generated

Future Directions

While the concept is currently centered around the spread of a specific disease, the model could be expanded to include many cases, such as product virality or resource priority.

Our project could be used to model product spread in a similar way to the spread of disease. Through the analysis of searches relating to a topic, whether it be specific (e.g. "What's being said about my product?") or general (e.g. "What products are in vogue right now?"), a company could make improvements to a product or tailor advertising. As an example, let's say a hypothetical company named Macrohard releases a software called Doors Century Edition which is known to crash often, especially in specific localities like the North American market, and is widely considered to be the worst version of Doors released. Using the terms being searched for, Macrohard could find the problems experienced most often by users and work to correct those issues. They could also find what people are talking about positively to associate their products through the use of strategic advertising. As an example, if they find that people in a city (or province/state) are looking at pictures of cats and are associating it with happy

emotions (e.g. "I love cats," "Funny cat pictures," or "lolcats"), they can create advertising for that area containing cats interacting with their products or release an updated UI including cats.

The project could also be used by governmental agencies to assign resources based on people's needs. For example, analyzing the searches for an area relating to impoverished conditions may enable an agency to shift resources to more fully support those in that area. Similarly, the current incarnation of the project would be useful for allocating health resources to affected areas.