Project A

The Polarization of Information on the Web

Charles Dickens, EE Dept. UH Manoa
Dr. Narayana Santhanam, EE Dept. UH Manoa, Project Mentor

-Objective-

Understanding and developing a more accurate model of modern day information aggregation is the primary goal of this study. Specifically we aim to answer the question: Across the internet, how polarized are opinions and information covering a topic?

-Approach-

Consider many users browsing the internet researching a controversial topic. Depending on their predispositions, one would begin their inquiry by choosing a familiar news portal. Then these sites would link to other news sites, blogs, etc. which a user will follow with a probability that reflects their current state of opinion. As described, users follow certain strains of information with higher probability, becoming confined to certain sections of the internet. This explanation implies that the way information is accessed on the web is a slow mixing Markov process, which would correlate with a polarization of opinion in the society.

Different users with biases can be identified by the states of the Markov process they are in. Then the polarization of information surrounding a topic can be characterized by the mixing rate among the ensemble of different users. We can summarize our approach with the following question: how much common information should different users see before they begin to agree?

-Data-

For this study we will be working with data from the web, that is any information that is accessible by a conventional user investigating a topic. For example we will see: articles, blog posts, tweets, etc.

The first challenge we face with this data is determining what information would indeed be relevant to influencing a user's opinion regarding a topic. Then, once this document is deemed relevant, another question we must answer is: Where on the spectrum of positions we are considering in our model does the relevant document lie? To better understand the issue consider a simple solution of labeling documents based on the presence or absence of controversial keywords and phrases.

To develop techniques to resolve the complexities described above, we can first harvest data from the web by developing an unbiased web crawler. Then we may begin our data analysis to come up with heuristics that will provide satisfactory solutions.

-Impact-

An understanding of the polarization of information on the web could be used to help policy makers identify topics of controversy or agreement to better serve their constituents. Another impact is in personalizing search suggestions that can break users out of confined regions of the internet.

Project B

Title: **Data shadows: exploring personal data**
Madi Whitman
Department of Anthropology
Purdue University

Key question: By diving into personal data, this project asks: how do people become data? And how do data become people? This open-ended project uses personal data to explore social complexity.

Data: The data for this project come from your data silhouettes via platform usage data. Potential data sources include Facebook, Google Takeout, scraping data from other platforms like Spotify, Tumblr, Twitter, etc. through their APIs. Because of consent issues associated with collecting data from social media users, we will use our own data, which, in theory, are data with which we should have a great deal of familiarity. We may see immense variation between us: how do our data account for our complexity? We may work to do some un-black boxing, or we might develop something novel to work toward greater transparency. A source of inspiration might be a data visualization tool like Data Selfie.

Potential impact: Predictive uses of data are fraught with assumptions about social life, personality, and technology use, in addition to what data purport to represent. The Cambridge Analytica "datastrophe" has demonstrated that how our data allege to reflect us has immense implications for how we move through the world. The time for data literacy, especially regarding our own data, is now. This project will help us to elucidate an area of obfuscation increasingly relevant to everyday life: the collection and use of personal data from platforms.

Project C

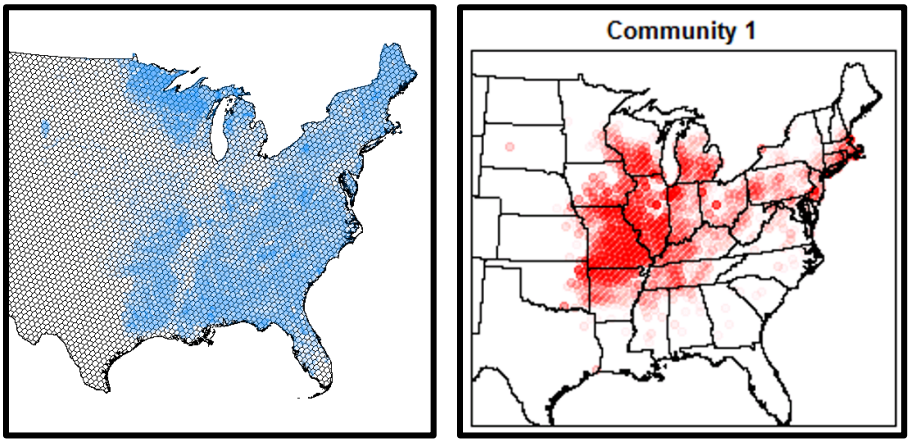
Title: Changes in forest communities of the eastern United States

Name/affiliation: Jonathan Knott, Dept. of Forestry and Natural Resources, Purdue University

Problem statement: Due to recent changes in climate, forest tree species have shifted their distributions to track with suitable conditions; however, little is known about how these shifts impact the larger forest communities that these species are a part of. This project aims to answer three main questions: 1) What are the main forest communities of the eastern U.S.? 2) Where are these communities located, and how have they shifted geographically? 3) What is causing these communities to change, specifically, is climate change influencing the distribution of these communities?

Data: This project utilizes the Forest Inventory and Analysis Program (FIA) data that was collected by the U.S. Forest Service in the 1980s and more recently in 2013-15. The FIA data contains abundance (stem count) of over 200 species from forests of the eastern U.S. and their associated size class (from which we can calculate things like basal area). The full data set contains ~75,000 FIA plots, but in order to reduce sampling bias and computation time, I have aggregated this data to around 2400 hexagon sampling units (which covers the same spatial extent as the original data). From this, I have used a statistical model (Latent Dirichlet Allocation, LDA) to identify 16 main forest communities. Therefore, the data for this project could be at two different stages, depending on the group's focus: either assessing changes to these already defined communities from LDA (Questions 2 and 3) or working with the raw and/or aggregated FIA data to explore the composition of the communities (Question 1).

Potential impact: Beginning in the early 1900s ecologists have argued about the definition of the "ecological community" with debates continuing today. In addition, maps published in the 1950s have been used to delineate these forest communities but were based on limited field observations. Others have tried to update these maps with the FIA data and other large-scale datasets, but to my knowledge, no one has looked at changes to these communities as a response to climate change. The outcome of this project will hopefully address recent changes to these communities, the drivers of these changes, and may provide an additional view point to the ecological community discussion. In addition, this project will help identify forest communities that may not be sustainable under future climate change.



(Left) Map of FIA plots (blue dots) and hexagon sampling units (black hexagons) and (Right) an example community and its abundance in the hexagons (darker red = more abundant).

Project D

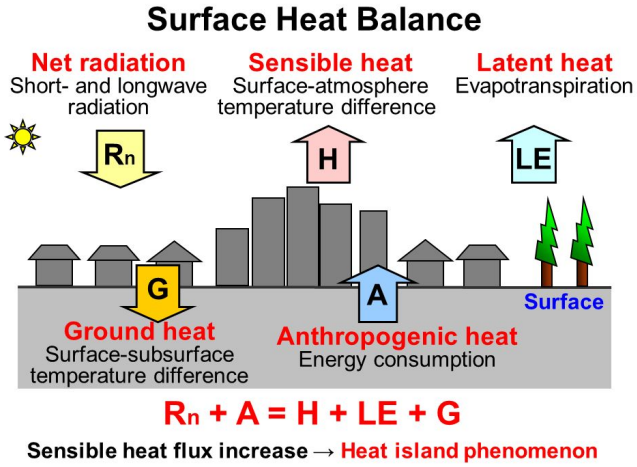
The estimation of surface heat fluxes using weather station data

Elahe Tajfar
Civil and Environmental Engineering Department
University of Hawaii at Manoa

Sensible and latent heat fluxes (*H* and *LE*) play a vital role in the earth-atmosphere system. These fluxes characterize the exchange of heat and moisture between the land surface and its overlying atmosphere. Therefore, accurate estimation of sensible and latent heat fluxes is of vital importance in different disciplines of hydrology, meteorology, ecology, agronomy, irrigation scheduling, and marine biology. Daily measurements of air temperature and specific humidity at weather stations contain implicit information about the magnitude of sensible and latent heat fluxes. As a result, we can take advantage of weather station data to estimate sensible and latent heat fluxes.

Half-hourly data including meteorological observations (e.g., air temperature, humidity, wind speed, atmospheric pressure) in addition to surface energy fluxes (e.g., sensible, latent, and ground heat fluxes, and net radiation) and land surface temperature are available (excel files) for four contrasted sites with different climatic and vegetative conditions: three sites in the United States (Audubon, Jun 19-Sep 16, 2006; Bondville, Jul 01-Aug 29, 2005; and Brookings, Jun 25-Sep 22, 2009) and one site in China (Daman, Jun 07-Sep 4, 2012).

In this project, the correlation between weather station data and surface heat fluxes can be assessed. In addition, the applicability of R packages (e.g., neural network) can be evaluated to predict sensible and latent heat fluxes.



Project E

1. Project Title: *The role of migrants in building city resilience for emergency response and disaster risk reduction (DRR).* Szymon Parzniewski, DPhil (PhD) candidate, Department of Political Science and International Studies (POLSIS) & Institute for Research into Superdiversity (IRiS), University of Birmingham, United Kingdom.

2. What is the role of migrants in building city resilience for emergency response and disaster risk reduction (DRR)? How can migrants be better integrated into resilience policy-making and implementation of emergency response and disaster risk reduction (DRR) initiatives?

3. As a team we will be looking at the Puerto Rican migration post Hurricane Maria to the continental United States (Florida in particular) based on the Emergency Management Agency (FEMA), the American Community Survey (ACS) and potentially additional data sources to analyze the role of migrants in resilience policy-making.

Two biggest challenges can be identified:
1. There are different paths of migration (including irregular and regular). There is no single record of Puerto Rican migration/displacement in the USA. Therefore, the analysis is often based on estimates and involves looking at indicators from various data sources.
2. Immigration is a highly political and sensitive issue. Building upon the previous studies (see below), we will be (among others) looking into the political dimension/activism of the Puerto Rican diaspora and the potential implications in the context of resilience policy-making/midterm elections that will take place in November 2018.

Media coverage (Dec 2017) highlighting the key issues that displaced Puerto Ricans face in Florida: https://youtu.be/-66VmnTfvwM

Several studies so far tryed to estimate the population displacement post Hurricane Maria and population simulation scenarios based on the monthly domestic arrivals and departures from Puerto Rican airports and the cumulative net domestic air passenger balance: https://medium.com/migration-issues/first-concrete-estimates-of-hurricane-maria-displacement-ad3e724a095f

More recent study analyzing changes registered in FEMA and changes in enrollment of students in school districts throughout the country estimated that 135,592 Puerto Ricans have settled on the continent and almost half a million could migrate to the continent by 2019: *Estimates of Post-Hurricane Maria Exodus from Puerto Rico* by Edwin Meléndez and Jennifer Hinojosa, Report by the Center of Puerto Rican Studies, October 2017: https://centropr.hunter.cuny.edu/sites/default/files/RB2017-01-POST-MARIA%20EXODUS_V3.pdf
https://www.citylab.com/equity/2018/03/exodus-the-post-hurricane-puerto-rican-diaspora-mapped/555401/
Some further visualizations and mapping based on the FEMA data:
https://edition.cnn.com/2018/02/21/us/puerto-rico-migration-data-invs/index.html

4. This research will generate a greater understanding of the contemporary features of resilience policy-making, in particular, the role of integrating migrants in resilience and disaster risk reduction (DRR) efforts at the city level. It will generate valuable insights which, among others, can be used to:
   - explain the cultural/geographical differences of decision-making and 'resilience thinking' in emergency and disaster risk reduction (DRR);
   - suggest and illustrate ways in which cities can improve the processes of resilience building in relation to the challenges deriving from the increasing role of human mobility and diversity;
   - support the work of local service providers by generating new insights into the resilience building processes underway.

Project F

Project Description

Zhen Li

5/3/2018

1. Project Title and your name, department, and university

Title: Towards cyber-physical vetting in critical infrastructures

Student: Zhen Li, major in Architecture Engineering, School of Civil Engineering, Purdue University

2. Problem Statement/Key Exploratory Question

Vet the potential cyber or physical attacks for smart buildings with the data flow in the systems through data mining and mathematical model.

3. Data

In this project, we are aiming to work with all the data can be collected in the smart building. Figure 1 gives a list of the smart appliances or meters in a smart residence which can provided the data for the customer and researchers.

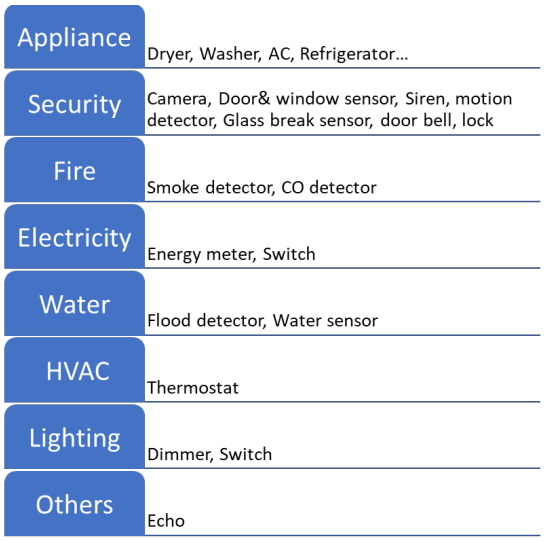


Figure 1. Appliances in a smart residence

In my project, we use a residence called ReNEWW House for the test bed. In this research level residence house, we can access some data through the pre-installed control system since 2017 Fall. Currently, we can access the data for the thermostat which reflect the temperature and relative humidity of the house, the energy meter and water meter. The complex part is how to vet the attacks by considering the data from different systems, which has different unit and different meaning.

4. Potential Impact/Reasons this project matters

Cyber Physical System is embraced by numerous applications in civil, energy, transportation and manufacturing fields. Meanwhile, the critical infrastructures are increasingly interwoven with

cyber components such as sensing, computing, and control devices. However, in recent years, some widely-known attacks to the CPIs have serious impact on the society and economic. An ad hoc issue in the Cyber-Physical Infrastructures (CPIs) is the security of the system: how to protect it away from the cyber and physical attacks. the It urgent demands the research of the strategies to vet a CPI from both “cyber” and “physical” perspectives and minimize the influence of the attacks.

Considering the gaps between the areas of cybersecurity and physical protection, and the difficulty to test the vulnerability of a real CPIs, this project aims to use a residence as a platform to develop the data mining technic in smart building and then to apply in on the common cyber physical infrastructure, which will have significant influence on system security.