CS486C – Senior Capstone Design in Computer Science
Project Description

Project Title: Internet Scraping for Power Outages from Utility Company Trackers	
Sponsor Information:	**Dr. Mikhail Chester**, Professor School of Sustainability and the Built Environment Arizona State University mchester@asu.edu **Ryan Hoff**, PhD Candidate School of Sustainability and the Built Environment Arizona State University ryan.hoff@asu.edu

Project Overview:

Infrastructure resilience researchers are people who try to figure out if an infrastructure system can handle failures, how well – and how fast – they recover from failures, and how to minimize the negative impact on any infrastructure (i.e., power systems, water systems, Internet, etc.).

These researchers need open-access methods to gather realistic failure data for power systems. In the modern environment, more and more power outages occur from a variety of hazards that continue to increase in frequency and intensity. Nearly every infrastructure system is interdependent with energy systems. And these infrastructure systems are threatened by multiple hazards.

Thus, more studies are required on the resilience of power systems, which will enable not only an understanding of the resilience of power systems themselves but also of the interdependent relationships that power systems have with other infrastructures.

The largest barrier to these studies is the lack of data that is publicly available for power outages. This lack of data prevents validation and leaves researchers without a viable way to model and predict future power system failures.

The resilient infrastructure lab has been attempting to capture the dynamics of interdependent infrastructure failure by designing fictional, or model, infrastructure networks for power and water, linking them with physical interdependence via water pumps, and then simulating how power failures can propagate to pump power loss than water pressure losses.

These fictional networks are called “Synthetic Networks.” An example for the city of Phoenix is provided in Figure 1.

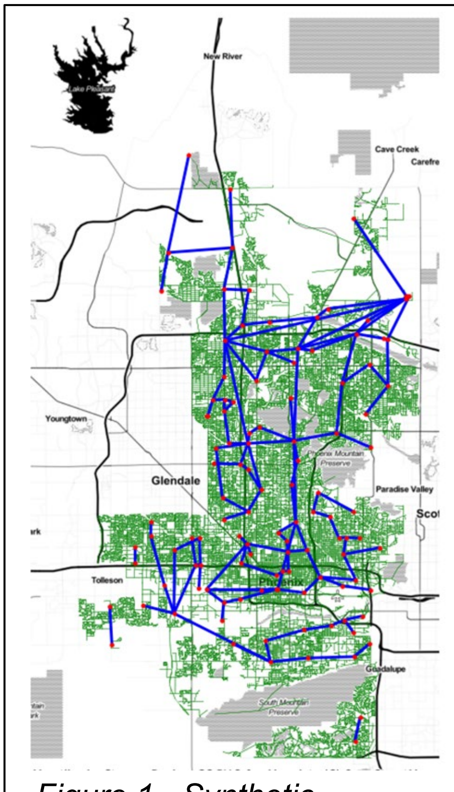

Figure 1 - Synthetic Transmission and Distribution Model for Phoenix, AZ. Little is known about how failure should propagate across these synthetic models.

These model networks use real geospatial constructs to simulate their real-world counterpart networks. The geospatial aspect of these networks is revolutionary in that it allows researchers to pair many different types of infrastructure networks together with the four types of interdependencies: physical, where the systems are directly dependent (e.g., water pump uses electricity); geographic, where the systems are co-located (e.g., water, stormwater, and sewer often travel under roadways); cyber, where information exchange between networks is crucial to their operation; and logical, where networks depend on some exchange via some other means, such as networks that rely on the exchange of goods and services.

Using these networks to simulate failures may allow researchers and city stakeholders to perform theoretical resilience assessments on infrastructure systems in the built environment without violating privacy or security policies.

These resilience assessments will aid in failure prediction, and potentially support ways to reduce or eliminate the identified failures.

Recently, a lot of progress has been made in the field of designing synthetic infrastructure networks. But **there is still a lack of data describing how failures propagate or "grow" within power networks**.

Synthetic infrastructure failures cannot be directly compared to real-world failures in a meaningful way. Most utility companies will not share past records of outage propagation. However, many companies provide real-time outage maps with locations and number of customers affected.

Figure 2 demonstrates how Georgia Power provides this information. But past data for outages cannot be found on the Georgia Power website.

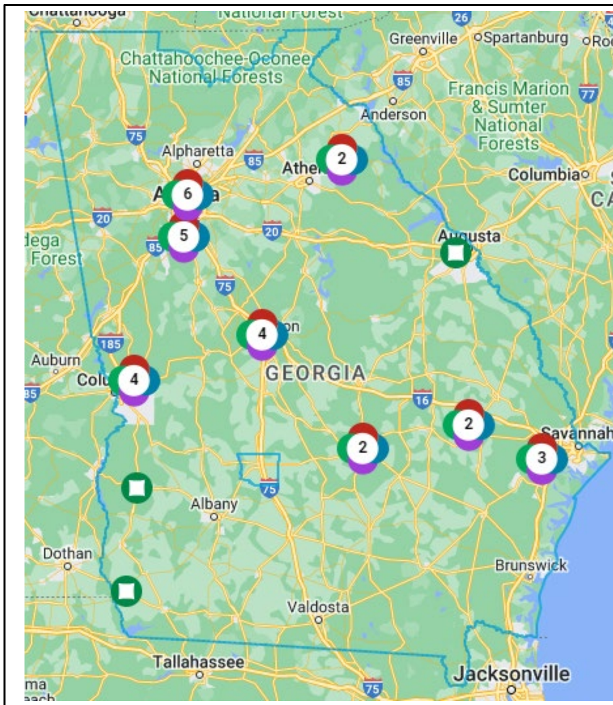

Figure 2 - Example Outage Map for Georgia Power. Locations are specific and affected customers are updated as the failure progresses.

Thus, the resilient infrastructure lab is seeking a database that captures ("scrapes") real-time outages from utility company websites and stores their metadata and geographic information in a database for research purposes. There is a commercial product that has attempted to collect this data in the past. It collects generalized county and township information, and it also includes timestamps as the number of customers without power increases and decreases during an event. But this product does not include specific geographic information, which is crucial for specific validation of cascading failure for synthetic networks.

The Envisioned Product

A live and growing database fed by a power utility scraper that captures power outages and their characteristics from public utility websites and APIs. For testing and development, the scraper and database can initially focus on a small set of utilities or regions, but be designed with the intention for scaling. Ideally the database would capture the utility name, onset of failure, extent of failure (often reported as number of affected customers, anticipated duration, and remediation of failure (i.e., when the failure event ceased to be reported).

Our vision of the minimum viable product (MVP):

- A software management product for searching for, and identifying unique outage maps, and configuring the system to capture (scrape) the sites for the time-stamp, population impact, and other needed informational components

- The software product must also return the data back to a user in a dashboard or comparable interface format for review and analysis

- The ability to capture (scrape) the specified data and store it in a scalable database

- A system of API query operations so that others can access and potentially display and/or use the data in their own systems

Some interesting stretch goals for this project:

- Extensions on the outage map searching tools that would improve access, and potentially reduce the amount of user-driven searching for the outage maps
  - this could include some form of AI searching strategies that finds, and reports finding, outage maps in real time without user involvement

- Extensions on the display of outage information that could include map overlays and/or other more user-friendly presentation formats

- The ability to conduct the scraping management operations (i.e., the main program "dashboard") from a mobile device

Knowledge, skills, and expertise required for this project:

- Web page development, including back-end (database, etc.) management

- Cloud-based data management

- Web-based scraping

- API development and management

Equipment Requirements

- Most tools for this project should be freely available, including database access and communication operations

- For cloud management, access will be provided by the sponsor or the capstone facilitator

Software and other Deliverables

- The applications as described above, deployed and tested successfully with real data
- A strong as-built report detailing the design and implementation of the product in a complete, clear and professional manner. This document should provide a strong basis for future development of the product
- Complete professionally-documented codebase, delivered both as a repository in GitHub, BitBucket, or some other version control repository; and potentially as a physical archive on a USB drive upon client request
- User Manual for configuring and operating the software, as needed