Evolution of Water Consumption in the USA: A Network Approach

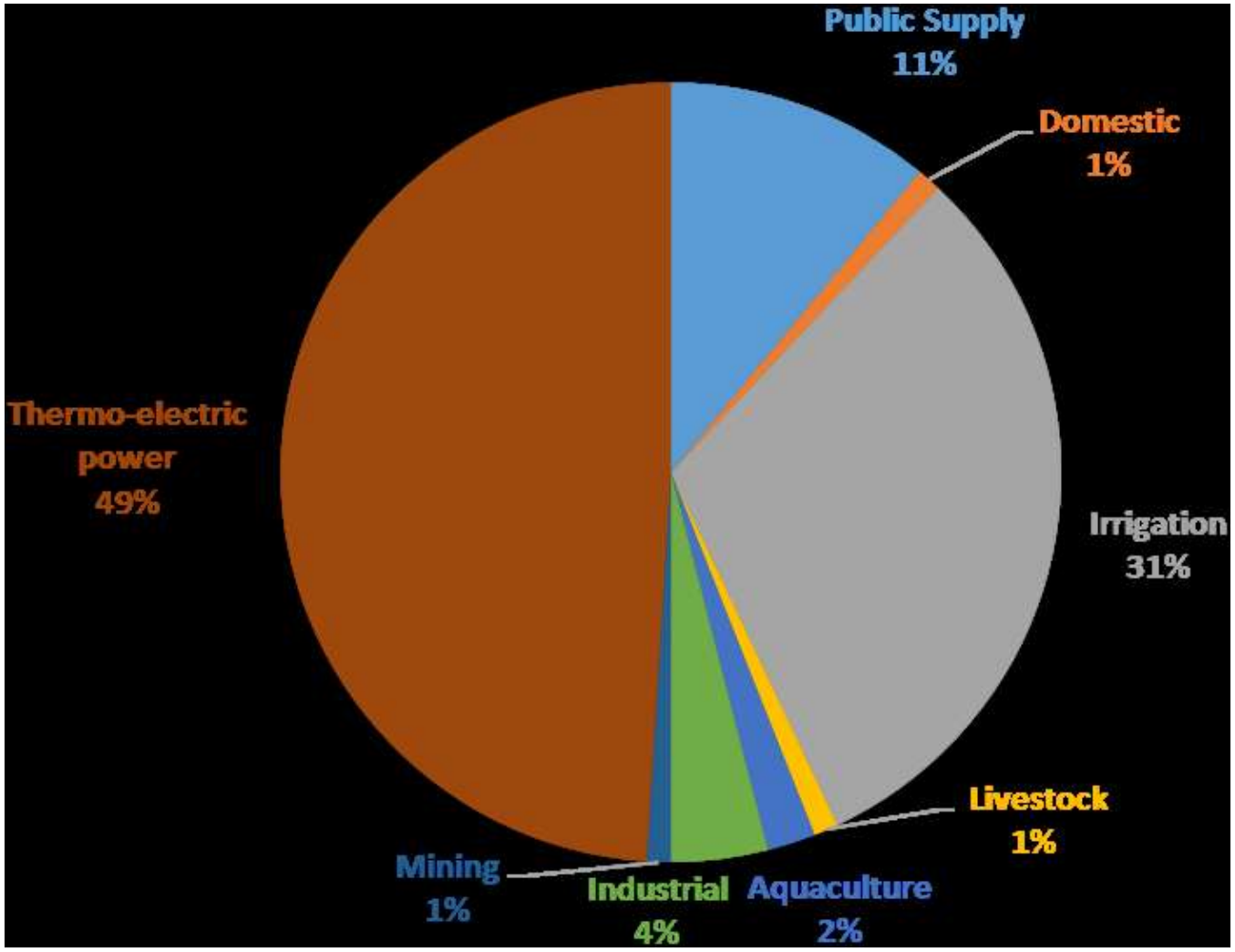
SK NASIR AHMAD

Rationale

- Water is one of the most vital element for human survival.
- Although, water covers approximately 70.9% of the earth's surface, only 1.8% of total water is available for use.
- Of this limited quantity of usable water, less than 0.3% is found in lakes and rivers, which is known as surface water, and the rest is found in the form of groundwater and polar ice.
- The USA withdraw 80% of its total water withdrawal from surface water source.
- Therefore, tracking trends of water consumption is critical.

So, are we consuming more or less water now than we did 30 years ago?

Total Withdrawals by Category



Public Supply

- Water withdrawn by public and private water suppliers that provide water to at least 25 people or have a minimum of 15 connections and is delivered to users for domestic, commercial, and industrial purposes, and also is used for public services and system losses.

Use of Statistical Indicators

- Traditional statistical indicators (e.g., mean and standard deviation) are commonly used to study and compare resource consumption trends in different regions.
- These indicators are also used to define and explain certain relevant properties, such as their distribution across cities, counties, states, nations.
- Additionally, these traditional indicators are most often used first-hand in models to predict the demand; e.g, how much water will we consume in 2030?

But!

Outliers

- In most datasets, some observations are distant from the general trends. These observations are known as outliers.
- Can be a real observation or may appear due to error.
- Most importantly, they can easily biased the outcome of the traditional analysis.
- It is difficult to detect them accurately.

Need: develop a new methodology to capture these outliers

Histogram

- For a graphical representation of the distribution, histograms are typically used .
- To construct a histogram, the range of the distribution, which is the difference between the maxima and minima, is first divided into separate bins. Then, the number of observations for each bin is recorded.
- Base of each histogram represents the bin width and the height represents the frequency for that particular bin.
- Finding an appropriate bin width turns out to be a major and highly non-trivial challenge because selection of different bin width can easily provide different types of distributions.

Network Science

- The foundation of Graph Theory, the “father” of Network Science, was laid by Leonhard Euler in 1736
- Later, Paul Erdős and Alfréd Rényi developed probabilistic theory in Network Science in eight papers on random graphs.
- Much later in the late 1990s, Duncan Watts and Steven Strogatz described the small world problem mathematically.
- At the same time, Albert-László Barabási introduce scale free network.
- Since then early 2000s, network science has been applied to myriads of fields, and it can be relevant in our case as well.

Homophily

- A propensity for human being to link with similar others is known as homophily.
- Persons in homophilic relationships share common attributes such as beliefs, education, and social status.
- In this study, the concept of homophily is applied to counties that share similar water consumption properties.

Network

- A network or a graph *G* is a collection of vertices/nodes *N* joined by edges/links *L*; *G*={*N*, *L*}.
- In network science distance refers to the least number of “hops” or nodes that must be visited to join any pair of nodes. It is also known as shortest path length.
- Average shortest path length is defined as the average of the shortest path lengths of that network.
- Largest of all these shortest path lengths is called the *diameter* of a network, which is used to determine the range of the network.

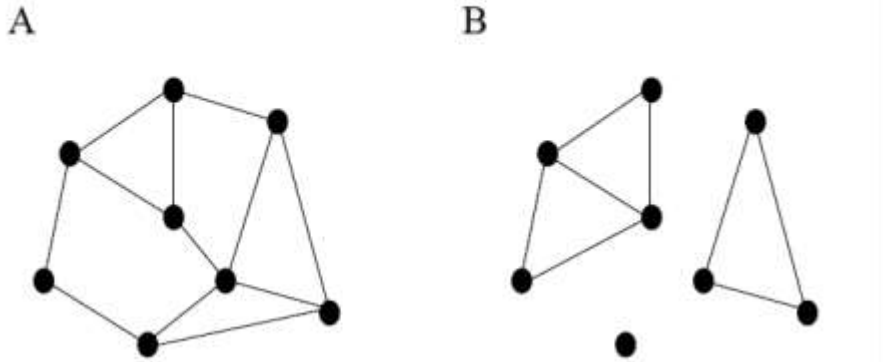
Connectivity

- A popular measure in network theory is related to the concept of connectivity and one measure of connectivity is called density.
- Density basically calculates the number of links in a network divided by the total number of potential links in this network. The equation for undirected networks is,

$$\rho = \frac{2L}{N(N-1)}$$

Giant Cluster

- All nodes in a network do not have to belong to one single network and this is referred to a connected versus disconnected network.



- Some networks, as they grow, i.e., as they accumulate more links, start to have one sub-network that tends to absorb most of the links, which is known as the giant cluster.

Formal Methodology

- A node i is linked to node j when:

Where,

$$\mu_i(1-\xi) \leq \mu_j \leq \mu_i(1+\xi)$$

μ_i is the consumption of node *i*

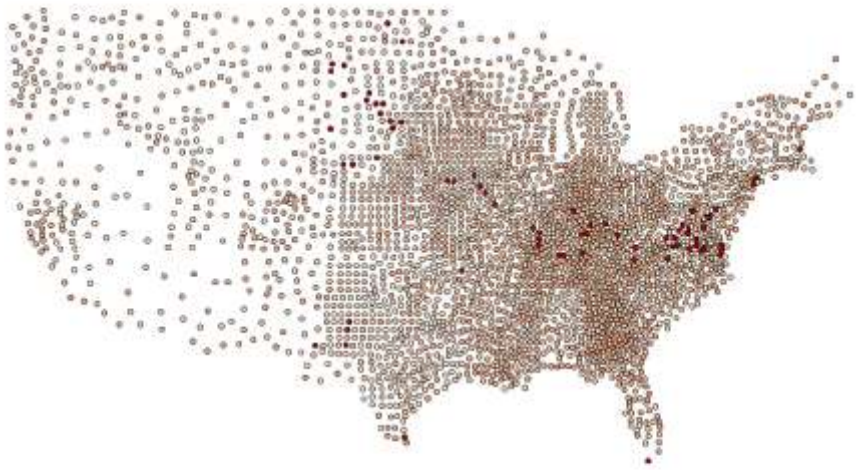
μ_j is the consumption of *j*

and ξ is called the cutoff percentage.

Selection of Counties

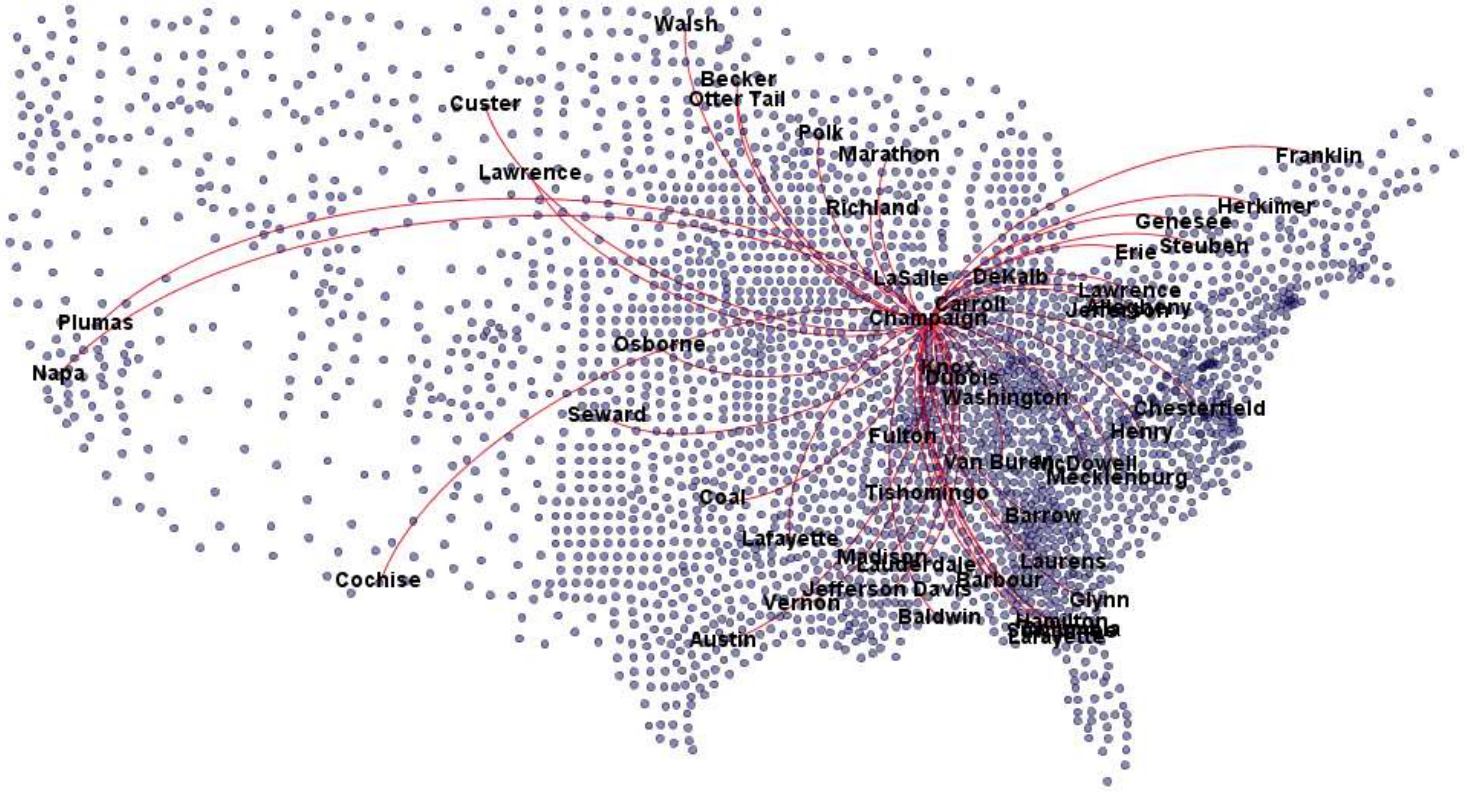
- 3109 counties are selected for this analysis, which excludes Hawaii, Alaska and other territories(e.g Puerto Rico)

Resulted Network

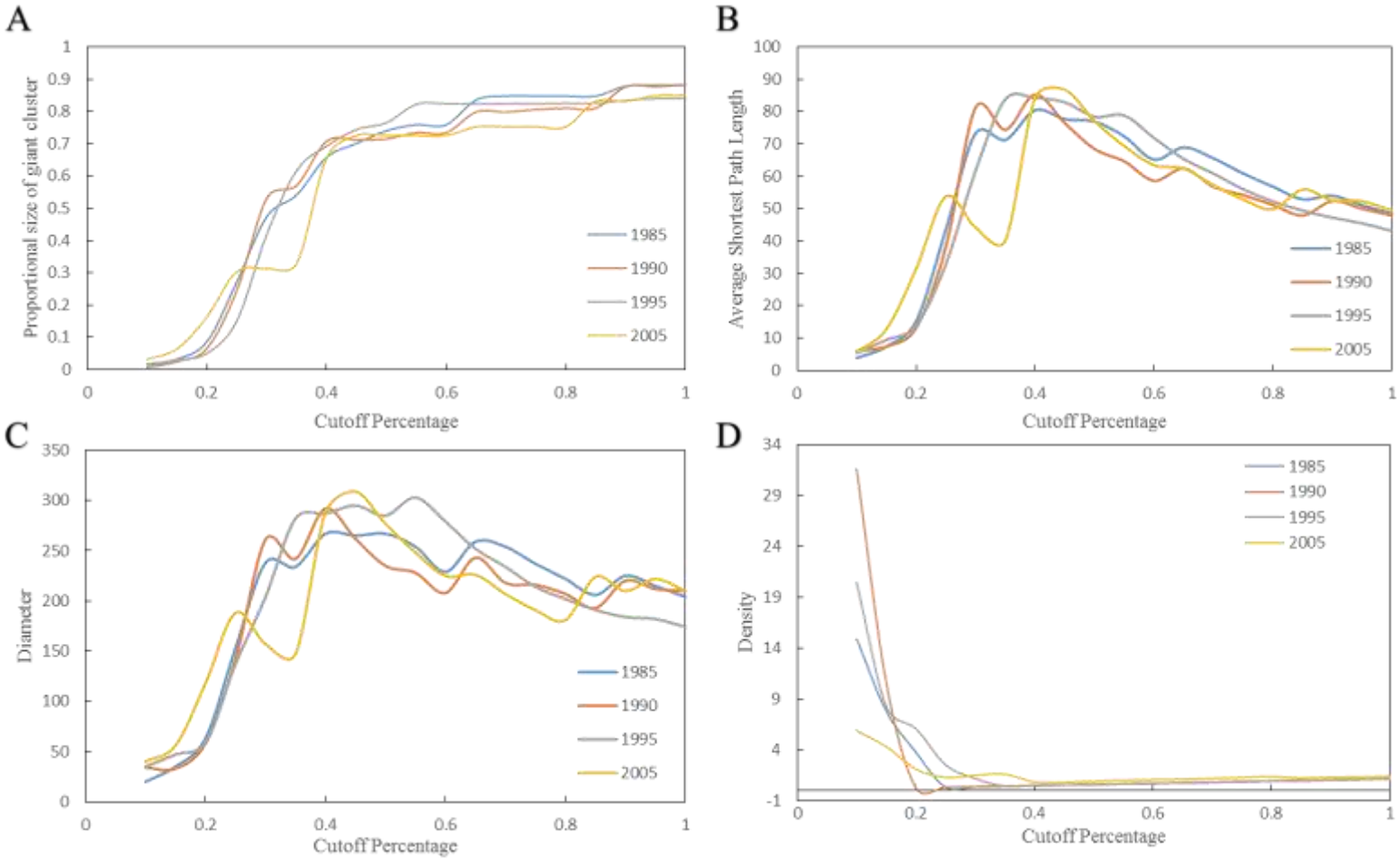


Year – 2005
ξ – 1%
No of Nodes – 3109
No of Edges – 51764
Graph Density – 0.01

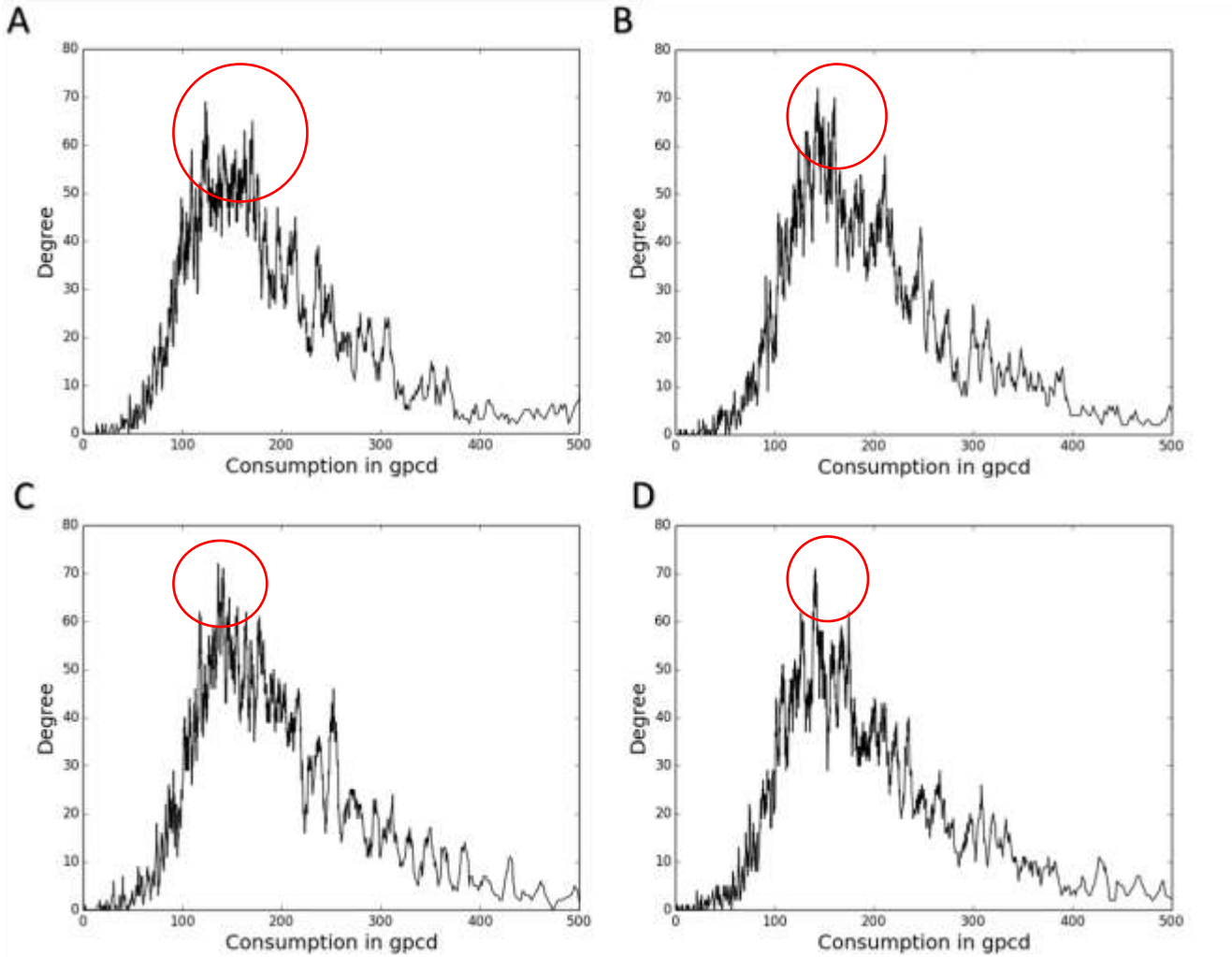
Champaign County



Network Properties



Degrees vs. Consumption



Visualization

- A full JavaScript visualization is also presented on the Complex and Sustainable Urban Networks (CSUN) Lab's website at http://csun.uic.edu.

Conclusion

- The main goal is to analyze the trend of water consumption in different regions.
- Traditional statistics, however, can fail to capture these current trends, and it was highlighted in this dissertation.
- Having these limitations in mind, a new approach is formulated based on network theory and already implemented to analyze the water consumption trend in the USA.
- Formalize this network approach and implement it to different dataset.

Thank you

Questions?