

Statistical Methods for Water Distribution Networks

Estéban Nocet-Binois

Github Page

URL: <https://github.com/DezrannCAS/CEE599B-Project-2024.git>

- See [docs](#) for full proposal and current progress.
- See [src](#) for source code (containing graph in R, methods, and eventually analysis).
- See [data](#) for original NJ1.inp file, as well as my data pipeline and the resulting .csv files.

Research Purpose

The goal is to explore statistical methods on graphs such as model selection and graph parameter estimation, but also uncertainty propagation, correlation and clustering in graphs. Just like distribution fitting, graph estimation involves identifying the best model for a set of observations, where those observations are network connections rather than time series, hence modeling large-scale interdependencies as realizations of random processes on graphs. For example, the Erdős-Rényi model can be estimated with the link density of the network, the Watts-Strogatz model can be estimated with the network's average degree, clustering coefficient and average path length, etc. The estimated model can be used for scenario generation, deviation estimation, and spectral theory in general can help gain insights into the network's structural properties (e.g. finding bottlenecks).

Most advances in this type of work focus on biological and brain networks (see reference 1.), for tasks such as identification and correlation of functional sub-networks, or network classification (eg, sound or diseased). The goal of this project is to explore the applicability of these tools to larger-scale, spatial networks such as water distribution systems (WDS). I have three points of interest:

- Hydraulic modeling in WDS that integrates real-world data and methods from graph theory. This will be the main focus of this project, but depending on the course this project takes, I may also explore the following questions.
- Network design and expansion: tools from graph limits can model how the network might evolve with extension. Precedent point is also important for this task.
- One other interesting research question could also be that of subgraphs and boundary effects: while zonal management of large scale distribution systems is a common approach, it is crucial to estimate the information loss when sampling subgraphs from larger network (see [Rheinwalt et al., 2012](#)).

For our study, we will be considering the Dover Township (NJ) distribution system [05 NJ 1](#). For more details on the dataset, checkout the [research log](#).

Limitations

There is one major limitation to this work, that is, the analysis is mainly structural. Although time-series demand (and induced pressure) data are available for said dataset, the range (80 hours) is just enough to provide information on network flow. The reason for it is that the analysis focuses on stability and perturbation theory, with respect to the flow over the network, aiming at sensitivity quantification (as ordinal ranking rather than prediction) and decomposition tasks. However, if we want to explore, and predict, the actual behavior of the system under disruption or failure (leakages, pipe breaks, floods or extreme events), the data we have is insufficient. The second problem is that there are few such datasets: we have the [LeakDB](#) generated dataset (of abrupt vs. incipient leaks), but no real-world examples and no data under extreme events.

These articles from researchers at the Università di Firenze [Arrighi et al. \(2017\)](#), [Tarani et al. \(2019\)](#) and [Arrighi et al. \(2021\)](#) can be interesting with that respect, however they also seem to perform quasi-static sensitivity analysis. One option could be to find a dataset of water precipitation in the same region as the distribution network, and estimate the network's response to potential flooding (by setting a threshold water level, past which disruption would be assumed). Yet the results can not be validated without real data on the state of the network (or at least on supply shortfalls at demand nodes).

References

1. [Fujita et al. \(2020\)](#): “A semi-parametric statistical test to compare complex networks” – focus on PPI networks of enteric pathogens
2. [Adraoui et al. \(2023\)](#): “Towards an Understanding of Hydraulic Sensitivity: Graph Theory Contributions to Water Distribution Analysis”
3. [Yu et al. \(2024\)](#): “A review of graph and complex network theory in water distribution networks: Mathematical foundation, application and prospects”
4. [Fragkos et al. \(2021\)](#): “Decomposition methods for large-scale network expansion problems”
5. [Anchieta et al. \(2023\)](#): “Water distribution network expansion: an evaluation from the perspective of complex networks and hydraulic criteria”
6. [Rebrova & Salanevich \(2024\)](#): “On Graph Uncertainty Principle and Eigenvector Delocalization”, with [code available](#)
7. [Arola-Fernandez et al. \(2020\)](#): “Uncertainty propagation in complex networks: from noisy links to critical properties”
8. [Tello et al. \(2024\)](#): “Large-Scale Multipurpose Benchmark Datasets for Assessing Data-Driven Deep Learning Approaches for Water Distribution Networks”