# Capstone Project 1: Displaying Large Organizational Networks – Final Report

## The Challenge of Communicating Large Organizational Networks

Networks can be a powerful tool for understanding, analyzing, and modeling the complex interactions that go on between organizations. They can even be deployed to quickly communicate broad information about the patterns that those interactions can take. But, most real-world networks are large, particularly those involved in making and implementing public policy, and those same network maps that can quickly communicate structures can just as easily hide both structure and detail once the map becomes large. This undermines their effectiveness as a communication tool.

That effectiveness can be reclaimed if the network is presented in the best way. One way that has shown signs of effectiveness is to break the network into subnetworks and only display a targeted portion of the network at any one time. But, this, too, presents a problem: an analyst is required who can take requests or anticipate what subnetworks will be needed, then build and supply those separate subnetworks. Instead, what is needed, is a tool that will allow anyone who accesses the tool to explore the network on their own, only asking for greater analysis when their interests go beyond what can be seen in the tool. In this project, I will build this tool using empirical data gathered from Vermont’s portion of the Lake Champlain Basin.

## Data

### Context

During that time, Vermont was the focus of actions by the federal Environmental Protection Agency (EPA). The Clean Water Act (CWA) requires jurisdictions to ensure that its waters are sufficiently free of pollution. Thresholds for how much of different pollutants can be handled in a waterbody are analyzed and set by the EPA through a Total Maximum Daily Load (TMDL). As the name indicates, the document reports the maximum amount of a pollutant that the waterbody can receive daily and remain clean. It assesses the natural processes in the waterbody and measures how much pollutant the waterbody can process in a day. Vermont has never met its obligations under the CWA in the LCB and, several years before the survey occurred, legal action overturned Vermont’s previous Phosphorus TMDL for the LCB. The legal challenge indicated that the previous TMDL overestimated the amount of pollution that the lake could process and failed to provide sufficient assurances that the targets could be met. This action forced state policy and how the CWA requirements could be met to the forefront, making water quality management a central policy issue in Vermont.

### Organization Data

The survey was executed as an online survey, with organizational representatives selected for their organization’s participation and approached through personal outreach. Table 1 lists the four organizational groups into which we sorted our respondents and reports the rate at organizations responded, by group.

Table 1: Survey Response Rates

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Organizational Group | Number of Contacts | | Completed Responses | | Response Rate (%) | | Observation Rate (%)[[1]](#footnote-4) | |
| 2015 | 2014 | 2015 | 2014 | 2015 | 2014 | 2015 | 2014 |
| Governmental Programs | 53 | 56 | 30 | 26 | 56.6 | 46.4 | 81.6 | 71.75 |
| Regional Actors and NGOs | 51 | 50 | 24 | 26 | 47.1 | 52.0 | 72.5 | 73.47 |
| Winooski Watershed | 52 | 52 | 29 | 11 | 55.8 | 21.2 | 80.9 | 38.16 |
| Missisquoi Watershed | 34 | 40 | 12 | 12 | 35.3 | 30.0 | 58.8 | 51.54 |
| Total | 190 | 198 | 95 | 75 | 50.0 | 37.9 | 75.1 | 60.26 |

### Link Data

### Impact

A range of different interests can be addressed through this project. Anyone who engages with network analysis can benefit from the approach to interactive network mapping that this project proposes. The immediate beneficiaries will be those working with the governance networks in the LCB, including the researchers who can better communicate their findings and the stakeholders who are embedded in the network. They will be better able to understand the context in which they work, allowing them to make better strategic decisions for their organizations, as well as to potentially diagnose and address situations where the network is impeding its own efforts to improve water quality in the LCB.

Many organizations and firms will find an interest in forecasting unobserved portions of a network from observed portions. The United States Department of Defense is interested in this technique for forecasting the structure of illicit networks, such as terrorist networks, where network targeting techniques are useful but the context prevents accessing information on the full structure of the network. Marketing firms may find this technique interesting for trying to use social media to market products. This project will build this visualization through organizational networks, but it will seek to build a demonstrate a tool that could be used in any situation where the data could be organized as a network.

### Deliverables

This project’s primary deliverable will be an application built in R’s *shiny* package. This package is specifically designed for making interactive online graphs and figures. R also offers a wealth of network analytic packages, including the *igraph* package and *statnet* suite of packages, as well as *ergm*, a package for doing ERGMs, included as part of *statnet*. This makes R an ideal venue to create a tool for dynamically manipulating network maps that can communicate both global and local network structures to non-specialists. Few examples of this approach currently exist, with one possible example being a network published by Resilient Infrastructure as Seas Rise (RISER) project, located at the University of California-Berkeley and the Center for Environmental Policy and Behavior at UC-Davis.[[2]](#footnote-6) This project will build an application with greater abilities to focus on small pieces of the network than that application allows.

Additional deliverables will include a code for a machine learning script to study link placement in the network as well as documentation of the results and, possibly, a comparison with current ERGM techniques.

1. Observation Rate records the percentage of non-directional network links that the survey was able to observe by obtaining a response from at least one of the two organizations involved in each link. See: Scheinert, S., Koliba, C., Hurley, S., Coleman, S., and Zia, A, 2015, The shape of watershed governance: Locating the boundaries of multiplex networks. *Complexity, Governance & Networks*, 2(1), 65-82. doi: 10.7654/15-CGN25. [↑](#footnote-ref-4)
2. The map can be seen here: http://riser.berkeley.edu/sea-level-rise-management-network/ [↑](#footnote-ref-6)