Complex Networks in Kelp Forest Ecosystems
Visualizing Big Data
Hamza Moulvi

Mark Novak, Mentor

Introduction

Kelp forests serve as an integral abode for aquatic species. Kelp forests are one of the most productive terrestrial systems, providing its denizens the conditions necessary for growth. Wave action allows for the constant motion of the seaweed, which provides maximum sunlight exposure and contact with nutrients (Mann).

The kelp forest ecosystem is a complex one. The understanding of the interactions between members of this unique ecosystem can be facilitated through the use of data visualization, a mode through which big data sets — such as the kelp forest — can be better understood.

This particular research project focuses on the kelp forests off the California coast, work that is funded by the NSF/NOAA CAMEO program, and is being conducted by OSU in conjunction with the UCSC, USGS, PISCO, and NOAA.

Objectives

Previous work has looked at the visualization of this data through the use of D3.js — a javascript library that allows for the creation of dynamic data visualizations — which led to the creation of a network visualization that illustrated the relationship amongst the diverse species found in the kelp forest. While this visualization highlights all the different types of interactions between species, the sheer number of species existent in the SQL database presented a large visualization, one that resulted in a slow load time, as well as a degree of difficulty in parsing the visualization to gain useful insights.

One of the goals with this work was to improve the load time with this particular visualization. Another was to expand the visualizations available through the exploration of a matrix visualization, which would contribute an increased understanding regarding the relationships between species.

Process

Because of the dynamic nature of D3.js, the resulting visualization is dependent on the data located in the database. A large effort to enter the data of various species interactions — both from observations as well as the literature — has taken place in the past few years, and the work continues to take place.

The work done in this particular project began following the existence of a large database containing these interactions.

The first task involved accomplishing faster load times for the existent network visualization (Figure 1). This was accomplished by taking a close look at the source code for the visualization. This examination of the code revealed the javascript framework — that is, the D3.js driving the visualization — to be located in the HTML code of the visualization. The D3.js was consolidated and moved to a separate javascript file, which allowed for an efficient allocation of resources thus resulting in improved load times. Further consolidation involved continued reduction of elements in

the HTML code; such was the case for the style declarations for the visualization which were moved to their respective CSS file. Consolidation resulted in a noticeable improvement.

The second task that was then taken was the creation of a matrix visualization. The first requirement in the creation of a D3.js visualization is a data source for the dynamic visualization. A SQL query was necessary to pull only the species interactions data from the larger data pool in the database. Following the creation of a query that produced the desired data, a PHP file was created. The PHP serves as the link between the data pulled from the SQL — which is located on the server — and the webpage — shown on an individual's own device. The PHP contains the SQL, and after querying the server for the data, returns the data in the form of a JSON file. A JSON file contains the desired data in an array and serves as the data file source for the consequent

Figure 1. Visualization (as of Dec. 11, 2014) mapping the species interactions in the California Kelp Forest.

D3.js visualization. The fact that this JSON file is created in real-time is the beauty of this type of dynamic visualization, as it allows for the visualization to be rendered with all pertaining data present at any given point in time. From here, the JSON file is inserted as the data source in the D3.js javascript code. The framework for the visualization itself comes from the D3.js website (d3js.org), which provides both documentation as well as a library of various visualizations that have been created.
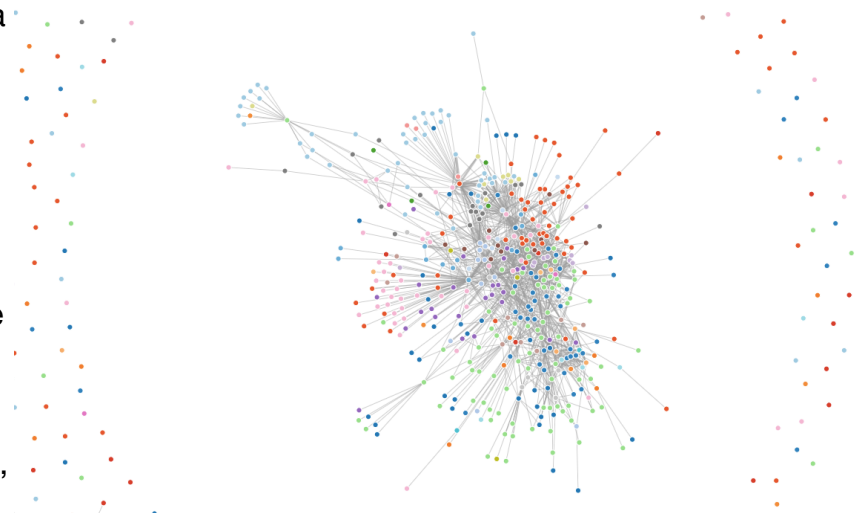
Challenges

Aside from the fact that the D3.js language has a bit of a learning curve, the largest challenge came with the creation of the matrix visualization. In this particular case, the groundwork was already established; that is, the SQL and the PHP already existed. The javascript for the visualization had been started, but was not functional.

The first iteration of the visualization — after the first round of work had been completed on the code — displayed nothing, which was unfortunate. The links were returning as null, a problem reflected by the error returned by the development error console in the browser.

Exclusion of this specific section of code displayed a grid along with the names of the species along the top as well as along the horizontal, with one crucial piece missing: the links themselves, which would fill in the cells in the middle of the grid.

It was here that it was decided to consult with a colleague, Mykle Hoban, a former student with the UCSC and now a PhD student at UH Manoa, who has experience with SQL/javascript and data visualizations. Mr. Hoban discovered that the problem related to the code derived from the example visualization from the d3.js website: the original code required number indices, whereas the code written involved
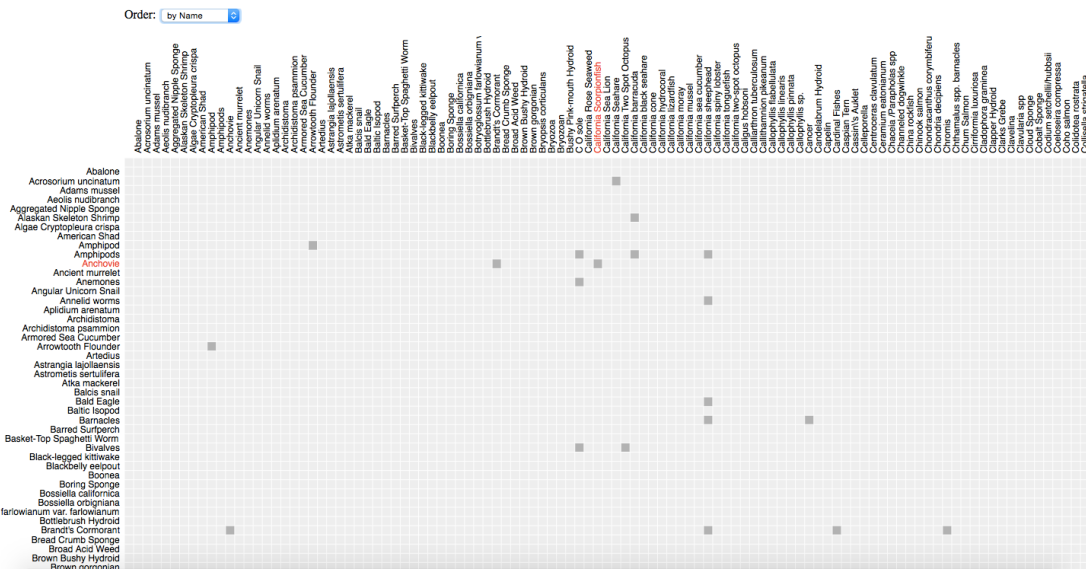
the working names of the species as indices. Mr. Hoban added an array that would map these working names to the numerical indices first, which eliminated the problem (this resulted in Figure 2; however, this is a memory-intensive visualization). He then optimized this for memory efficiency, which allowed for the resulting successful matrix

Figure 2. Matrix visualization (as of Dec. 11, 2014) mapping the types of species interactions in the California Kelp Forest.

visualization. Thus, a challenge was mitigated in collaboration with a colleague.

Products/Accomplishments

The species interaction network relating the interactions amongst species underwent an improvement in efficiency (fig. 1). **Figure 1** shows a static version of the network. The actual network (accessible at http://bit.ly/1sn0Jay) is dynamic, allowing the user to interact with the data, as well as adapting to new data added into the database.

The additional visualization created was a mapping of the various interactions (e.g. parasitic) amongst species on a matrix (fig. 2). **Figure 2** shows a small static version (the
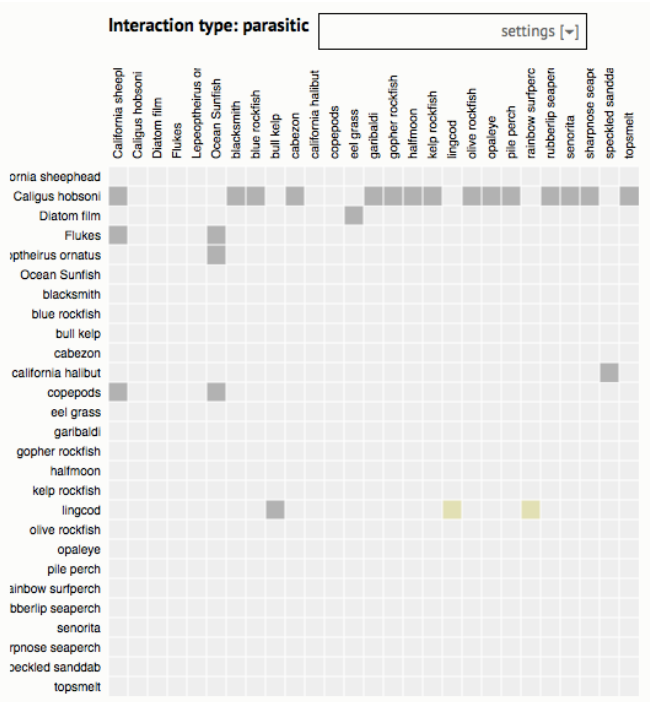

Figure 3. Improved matrix visualization (as of Dec. 11, 2014) mapping the types of species interactions in the California Kelp Forest. Developed with assistance from Mykle Hoban. .

visualization is accessible by navigating to http://bit.ly/HmMatrix). Again, the actual visualization incorporates a degree of interactivity. **Figure 3** is an improvement (reduced memory usage; removal of redundant links) of the visualization shown in **Figure 2**, and is the result of Mr. Hoban's collaboration (accessible by navigating to http://bit.ly/1urZqH0).

Skills Learned

D3.js is a different language, and like any languages, a learning curve is present. Much of the time spent in the course of this work involved developing an understanding of the programming language, which meant that prior to the creation of the matrix visualization, a fair amount of reading the documentation/looking at other examples was done to gain a better understanding. D3.js also, naturally, requires one to have an understanding of the javascript language, and this too was a bit of a learning curve.

Non-technical skills were also gained: communication of progress, time management, as well as communicating with colleagues when help was needed were all skills that were both developed and important in the course of this project.

References

Mann, K.H. (1973). Seaweeds: Their Productivity and Strategy for Growth. *Science*, *182*(4116). 975-981.