Bioinformatics Analysis: Visualizing Transmission Patterns of *Clostridium Difficile* Within an Urban Hospital

George Sivulka

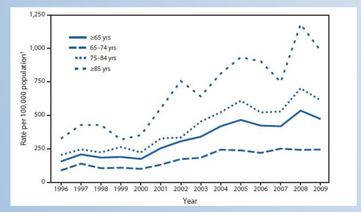
Mentor: Theodore Pak; Teacher: Jeffery Marcucio

Abstract

Computational analysis as a means of visualizing and quantifying the transmission of the potentially fatal *Clostridium Difficile* bacterium is a novel means of characterizing transmission pathways in healthcare facilities. To investigate the spread of *C. difficile* and link the transmission patterns to real world practices and contamination pathways, the electronic medical record (EMR) database of the Mount Sinai medical system was queried for a list of patients who contracted the disease nosocomially and their respective caregivers—who are possible 'vectors' of the disease. Using this 63,477 line dataset, JavaScript libraries *d3.js* and *Underscore.js* were used to parse the data and formulate a "Force Layout Diagram" to visualize and represent analytical trends. The layout produced significant amounts of connectivity, underscoring individual caregivers and departments more prone to spreading *C. difficile*, and potentially revealing problematic health care practices/practitioners.

Introduction

The spread of the *Clostridium difficile* bacterium is a growing problem in healthcare facilities, killing approximately 29,000 people in the United States in 2011. [1] In fact, the potentially fatal bacteria most commonly spreads in medical institutions and hospitals; the rate of C. difficile acquisition is estimated to be 13% in patients with hospital stays of up to two weeks, and 50% with stays longer than four weeks, resiliently thriving in the low native gut bacteria populations of patients on antibiotics [2,3]. Hence, the typical pathophysiology of the bacteria is that it replaces normal gut flora that has been compromised by an antibiotic treatment typically administered for an unrelated infection, typically overrunning the intestinal microbiome. However, through more efficient sanitation measures, hospitals have been attempting to curtail the spread of infection. Computational data analysis, especially data visualization, provides a valid means of streamlining this process, allowing trends in transmissibility to be revealed to the doctors and healthcare officials that they concern. Ultimately this can be used to attempt to reduce the rates of nosocomial infections in hospitals.

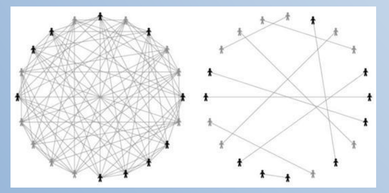


[4]

C. Diff As a Growing Issue
Rates of Clostridium difficile Infection Among Hospitalized Patients Aged ≥65 Years, by Age Group

Objective

The purpose of this research is to create web data visualizations with cross browser platform capabilities in d3.js using the hospital logs from the Mount Sinai Medical Center to clarify and shed light on the past transmission pathways and find trends and correlations not readily visible from the raw data. Ultimately proving a higher rate of nosocomial infection associated with different hospital caregivers using a visualization tool would lead to better infection control practices.



[5]

Transmission Pathways
Example of a mockup social network model of contact induced infections

*** An example of a nodes array**
Bottom left screenshot

Methodology

To acquire data relevant to the project the Mount Sinai databases were queried for all interactions between patients and caregivers with "008.45 INTESTINAL INFECTION DUE TO CLOSTRIDIUM DIFFICILE" as their primary or secondary diagnosis. The specific query received the parameters of each interaction that was hypothesized would have potential impacts on the rate of transmission, including the patients "LENGTH_OF_STAY", "AGE_IN_YEARS", "MASKED_MRN", "CHECK_IN_DATE" and their caregivers in each row by "FIRST_NAME" and "LAST_NAME" and categorization by their "CAREGIVER_ROLE". This generated a 63,751 row comma separated value or "csv" file of patient to caregiver interactions. However, in order to meet the specific requirements of the Force Layout in d3.js, Underscore.js was used to parse the data in each row into JavaScript Object Notation (JSON), creating an object of nodes with numerous arrays for each individual person involved in the interactions

Parsing the Data
Use of Underscore.js's [.each] function to parse through the csv and create the desired data arrays for nodes and edges discussed below

```
_.each(rows, function(r) {
    if (parseInt(r.id, 10) > 2000000 || parseInt(r.id, 10) < 1000000) {
        return;
    }

    if (_.isUndefined(allEdges[r.mrn + " " + r.first_name + " " + r.last_name])) {
        allEdges[r.mrn + " " + r.first_name + " " + r.last_name] = edges.push({
            source: r.mrn,
            target: r.first_name + " " + r.last_name
        }) - 1;
    }
    if (_.isUndefined(allMRNs[r.mrn])) {
        allMRNs[r.mrn] = nodes.push({
            name: r.mrn,
            group: "patients"
        }) - 1;
    }
    if (_.isUndefined(allCaregivers[r.first_name + " " + r.last_name])) {
        allCaregivers[r.first_name + " " + r.last_name] = nodes.push({
            name: r.first_name + " " + r.last_name,
            group: "caregivers",
            role: r.role
        }) - 1;
    }
```

Each row was swept using underscore.js's [.each] function for a list of elements, yielding each in turn to an iteratee function, in which the mrn number and other variable attributes of each patient and caregiver were extracted from the csv to further classify the attributes of their respective nodes later on. Thus each node array represented an individual on the visualization, complete with their attributes that would allow detailed analysis.

```
[{"name":"1654767840","group":"patients"},{"name":"Brian
Markoff","group":"caregivers"},
{"name":"1654767840","group":"patients
"},{"name":"Erin
Rule","group":"caregivers"},
{"name":"1654767840","group":"patients
"},{"name":"Peter
Taub","group":"caregivers"},
{"name":"1654767840","group":"patients
"},{"name":"Steven
Weinfeld","group":"caregivers"},
{"name":"1654767840","group":"patients
"},{"name":"O
Guttman","group":"caregivers"},
{"name":"1654767840","group":"patients
"},{"name":"David
Forsh","group":"caregivers"},
{"name":"1654767840","group":"patients
"},{"name":"Ageliki
Vouyouka","group":"caregivers"},
{"name":"1654767840","group":"patients
"},{"name":"Brian
Markoff","group":"caregivers"},
{"name":"1654767840","group":"patients
"},{"name":"Erin
Rule","group":"caregivers"},
{"name":"1654767840","group":"patients
"},{"name":"Peter
Taub","group":"caregivers"},
```

```
function color(d) {
    // TODO: check d.group ?
    return d.group == "patients" ? "#3182bd" : "#c6dbef";
}

function outline(d) {
    if (d.role == "Attending") {
        return "pink";
    }
    if (d.role == "Technician") {
        return "grey";
    }
    if (d.role == "Assessor") {
        return "red";
    }
    if (d.role == "radiologist") {
        return "green";
    }
    if (d.role == "Primary Surgeon") {
        return "blue";
    }
    if (d.role == "Attending Anesthesiologist") {
        return "black";
    }
    if (d.role == "Rn - Scrub") {
        return "orange";
    } else {
        return "#3182bd";
    }
}
```

Characterizing the Nodes
A series of if statements was used here to help identify caregivers based on their role, job, or patient interaction

In this way the code also acts as a table for understanding the vis.

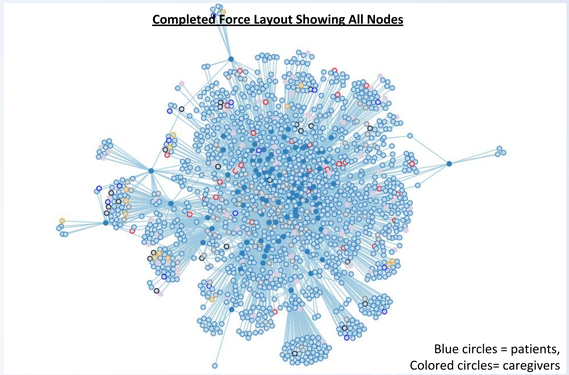
However, in order to show connectivity between patient and caregivers, a separate JSON object was created for edges to connect nodes. The two parameters for an edge were "source" node and "target" node, attributes that in d3 needed to correspond to the index value of the nodes "array" to connect in the nodes "object". In order to do this Underscore.js was again used to parse the mrn and name information from each row in the csv and later the source and target index (desired node array's number in the nodes object) for each array.
Once this data analysis was complete, a Force Layout was created in d3.js using code that created shapes for each array and connected them while maintaining a repulsive force. * In order to characterize nodes to facilitate visual analysis, node outline color changed depending on the role of the caretaker.

Results

Once run successfully, the visualization immediately revealed tremendous amounts of correlation in the 1171 bed hospital-- going above and beyond the fulfillment of our hypothesis.

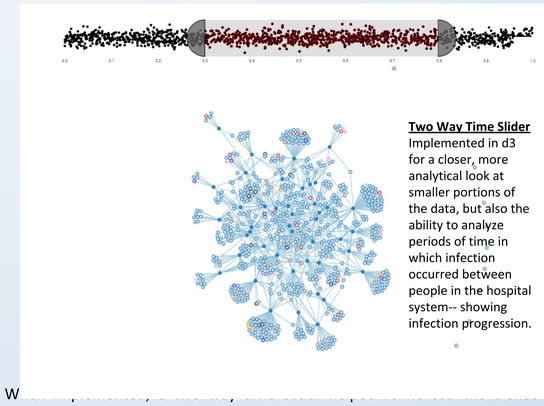
Even when repulsive force was maximized, the wide margin of the nodes in the center refused to expand away from each other, indicating connection to edges with other nodes that directly emphasized the incredibly high rates of transmissibility of *Clostridium difficile* among these 'nodes' in the community.

Completed Force Layout Showing All Nodes



Blue circles = patients,
Colored circles= caregivers

Especially relevant as seen above is the fact that the majority of the infected patients (the blue circles) are those in the center of the graph-- and thus the ones with the most connectivity. This high level of connectivity relates to contact with other nodes who received the disease, proving the first trend that transmission from a caretaker 'vector' is extremely prominent in a hospital setting. Moreover, assessors, attending nurses, and technicians (the lighter blue circles with red, pink, and grey outlines) were the most connected of all the caregivers.



Two Way Time Slider
Implemented in d3 for a closer, more analytical look at smaller portions of the data, but also the ability to analyze periods of time in which infection occurred between people in the hospital system-- showing infection progression.

W n above. Many caretaker nodes remained on the graph almost the entire time, disappearing at almost no point. Shown using this time selection, this trend in caretaker presence especially highlighted that a failure to achieve correct sanitation measures-- and thus success in infecting patients-- wasn't a random process. Certain mistakes in the system, even certain doctors (those recurring nodes that never disappeared), were highly at fault for many of the infections, and thus by not disappearing as time progressed are prime suspects for transmissions vectors.

Discussion

The Force Layout diagram created in d3.js most notably elucidated the hundreds of transmission pathways per patient in an urban medical center. It showed far higher levels of connectivity than hypothesized, and allowed for a better understanding of the frequency and magnitude of Clostridium Difficile's proliferation-- truly a bacterium thriving by the exploitation of low native gut populations due to antibiotics.

However, the data also outlined more interesting and specific trends directly related to specific entities and groups within the hospital system. Technicians showed the most amount of connectivity to infected patients, possibly due to incorrect machine sanitation practices, or lacing self-sanitation standards in place for technician employees. While Technician's relatively short amount of contact with most patients may seem to make them less likely candidates for Any contact with the virus, including touching a machine that someone else infected and then touching one's mouth for the slightest oral transmission can lead to the contraction of the disease. Assessors typically deal with patients not yet sanitized, and thus could be more prone to becoming C. diff's vector. Nurses are also more prone because they spend lots of time with patients in close contact, especially considering elderly patients they must help excrete. Also highly significant was the clear correlation between caregivers functioning as Attending Anesthesiologists and "Rn - Scrub"s and a lack of involvement in transmission. Probably interacting in short and sterile confrontation with patients, both caretaker positions would have a lower chance of acquiring the bacterium and unintentionally infecting others.

Sources of error were quite minimal due to a lack of human involvement as the computers dealt with the data and crunched the numbers, besides statistical probability that wasn't included in the experiment. The size of the hospital and the involvement of numerous types of caregivers-- across all fields-- allowed for a more accurate representation and understanding of the transmission network. While gastrointestinologists would more likely show up with connections to patients with *C.diff* that would need their expertise for treatment, this diversity in fields still accurately depicted visualization patterns across more than one doctor or caretaker position. Moreover, most patient hospital stays spanned multiple visits, and many of these patients were only diagnosed with *C.diff* on their second or third visit. Stomach doctors could also be the ones spreading the disease the most, as they have the most direct exposure-- just another balance and check in the data logs. Still, a specific inconsistency could be accounted for statistically per group in future visualizations. Other ways the visualization will be improved in the future include a quantitative analysis of the numbers of connections per selected nodes by counting the amount of edges attached to it and linking the result to the content of a separate HTML element outside the visualization. Also, grouping of caretaker nodes by their respective positions could even further outline transmission level disparity.

Conclusion

Ultimately the Force Layout proved a strong connectivity between caregivers and patients infected with C.diff., emphasizing a significant correlation of caregiver to patient transmission and highlighting certain fields that require reassessment.

References

[1] Lessa, Fernanda C *et al.* (26 February 2015). "Burden of Infection in the United States". *New England Journal of Medicine* 372
[2] Clabots CR *et al.* (September 1992). "Acquisition of *Clostridium difficile* by hospitalized patients: evidence for colonized new admissions as a source of infection". *The Journal of Infectious Diseases* 166(3):
[3] Halsey J (2008). "Current and future treatment modalities for*Clostridium difficile*-associated disease". *American Journal of Health-System Pharmacy : AJHP : Official Journal of the American Society of Health-System Pharmacists* 65 (8): 705–15.
[4] http://www.cdc.gov/mmwr/preview/mmwrhtml/mm6034a7.htm
[5] www.umdrightnow.umd.edu/news/researchers-propose-social-network modeling-fight-hospital-infections