Network Analysis of Inter-Organizational Communication Leading to the Flint Water Crisis

By Pete Souza [Public domain], via Wikimedia Commons

President Barack Obama sips filtered water from Flint following a roundtable on the Flint water crisis at Northwestern High School in Flint, Mich., May 4, 2016

Organizational Communication (CMN 412)

Justin Pence, Graduate Research Assistant
Socio-Technical Risk Analysis (SoTeRiA) Laboratory
Directed by Professor Zahra Mohaghegh,
Department of Nuclear, Plasma and Radiological Engineering
http://soteria.npre.illinois.edu/
PhD Student, Illinois Informatics Institute
jpence2@illinois.edu
Table of Contents

1. Introduction .......................................................................................................................... 1
2. Project Guidelines .................................................................................................................. 1
3. Case Materials ...................................................................................................................... 1
4. Network Analysis .................................................................................................................. 2
   4.1. Stating Data Assumptions ............................................................................................... 2
   4.2. Edge Weighting ............................................................................................................... 2
   4.3. Getting Started with Gephi ............................................................................................ 4
   4.4. Dynamic Network Features ............................................................................................ 5
   4.5. Network Data Analysis .................................................................................................... 7
       4.5.1. Modularity ................................................................................................................ 7
       4.5.2. Average Degree ....................................................................................................... 7
       4.5.3. Eigenvector Centrality ............................................................................................. 7
       4.5.4. Graph Density .......................................................................................................... 8
   4.6. Exploratory Analysis ........................................................................................................ 8
       4.6.1. Attribute Visualization ............................................................................................ 8
5. Interpretation of Results ......................................................................................................... 10
   5.1. Modularity ...................................................................................................................... 10
   5.2. Average Degree ............................................................................................................. 11
   5.3. Eigenvector Centrality .................................................................................................. 11
   5.4. Graph Density .............................................................................................................. 11
6. Authoring Your Paper ............................................................................................................. 12
   6.1. Referencing .................................................................................................................... 12

References .................................................................................................................................... 13

File Inventory

[Diagram of file inventory with files and folders]
1. Introduction

The Flint Michigan Water Crisis was selected as a case study due to the widespread inter-organizational communications that contributed to flawed decision-making, resulting in an unprecedented man-made contamination of a public water supply (See Case Materials for Additional Background). This case study is focused on the events that led up to the crisis. During this stage of pre-crisis, organizations played a critical role in contributing to systemic failure, and as organizational communication scholars, it is our opportunity to analyze the organizational contribution to this crisis to ensure that it does not happen again.

This project requires the student to read case materials, perform a network analysis of the organizations and individuals involved, relate the results of the model to the relevant organizational communication theories discussed in this course, and interpret findings in a comprehensive paper which brings references from relevant materials, uses visual outputs from the model as supporting evidential figures, and elaborates on the results of the network analysis to reach new theoretical and practical solutions.

2. Project Guidelines

This document provides step-by-step guidelines for conducting this project. The steps of this project include:

- Reading the case materials outlined in the following section (Section 3)
- Conducting a network analysis using Gephi (Section 4)
- Interpreting model results (Section 5)
- Authoring a paper (Section 6)

3. Case Materials

Students will read the following sections of the March 2016 Flint Water Advisory Task Force Final Report (FWATF, 2016):

- Executive Summary (Pages 1-3)
- Background and Summary Timeline of Key Events (Pages 15-21)
- Roles of Government Entities in the Flint Water Crisis (Pages 26-52)

Students are encouraged to take notes and develop a conceptual mental model of the organizational communication patterns in the Flint Water Advisory Task Force Final Report (FWATF, 2016). Upon completion of reading the case materials, students should be able to answer the following three questions:

1. Were concerns over water quality and safety communicated effectively?
2. What constraints contributed to a reduced flow of information between organizations?
3. Which organization is primarily responsible for the Flint Water Crisis?

In this project, you will take the role of an organizational communication expert. Based on the information you received from the Flint Water Advisory Task Force Final Report, you will elicit your own opinion about the contributions of distinct connections between individuals and organizations leading up to the Flint Water Crisis. Section 4 provides an overview of this weighting process, and how to use your data in Gephi (Bastian et al., 2009).
4. Network Analysis

**Before beginning this section of the project, students should have read the Flint Water Advisory Task Force Final Report located in the Case_Materials folder.**

In this step, students will perform a network analysis on a dataset relating to the case study. This section provides detailed instructions on using the network analysis software Gephi.

4.1. Stating Data Assumptions

The network data in this project was extracted from qualitative information from the Background and Summary Timeline of Key Events (Pages 15-21) of the Flint Water Advisory Task Force Final Report. In order to develop a model with limited scope and manageable features, the following assumptions were made:

- This model combines individuals and organizations into one data set.
- Individuals are associated to their respective organizations through a directed connection (i.e., individual → organization), where communications or information sharing can be directed from any individual or organization to another.
- Time in this model refers ONLY to the initial event time where the individual or organization became involved in the case study as indicated in the Summary Timeline of Key Events (Pages 15-21), therefore, historical relationships or built-in collaborations/contracts are not assumed.
- Nodes were created when an individual or organization introduced some new information in the case, or performed some type of communication or with one or more other individuals or organizations.

4.2. Edge Weighting

Derived from Pages 15-21 of the report, each edge in the network corresponds to a communication/event between individuals and organizations. The analyst should read through each of the Key Events with the excel file open, and add their data sequentially, making sure that the data is properly entered into the cell corresponding to each event. For quantification of these qualitative events, two factors are considered: (1) days to crisis (i.e., culmination to national and internationally recognized crisis in October, 2015), and (2) significance of the communication as it contributes to public health concerns (i.e., based on a likert ranking of; low significance = 1, medium significance = 2, high significance = 3, and critical significance = 5). Additional indicators could be considered, but for the simplicity of this exercise, these two are implemented in the study. When entering significance values, the analyst should provide a brief written justification for their score.

For this case study, days to crisis represents an inverse time pressure, where the more time a public health concern was communicated and left unresolved, it may be an indicator of poor quality decision making, and in combination with the significance of the communication to public health concern, a weight of importance for the case study can be estimated. Simply put, a multiplier for each communication/event significance will be combined with each time value leading up to the crisis to determine the weight of each edge between two nodes, considering the context of the case study.
The opinion process can be conducted entirely in Excel, and can combine as many analysts as is desired. The template for conducting this process is in the Network_Data > Weight folder (weight.xlsx).

The quantitative process is as follows:
1. Days are calculated in Excel from the time of the event to the time of crisis using a function (=DAYS(end_date, event_date)).
2. Analyst will rank each event as its significance to public health concerns on a scale of low (1), medium (2), high (3), and critical (5).
3. A combined indicator is calculated (=days_result * message_significance_score).
4. The rank of each value in the combined indicator is calculated as a percentage out of the entire column of data (=PERCENTRANK(array, x)), and this percentage will provide the final weight as a value from 0-1
5. Multiple weight columns can be copied into one workbook, and using the averaging function in Excel (=AVERAGE) an aggregated total weight can be calculated

Therefore, the combined indicator of these two factors will create a weight which can be applied to the edges in the network, and the aggregation of these scores creates a more comprehensive elicitation process.

The concept is shown in the following table:

<table>
<thead>
<tr>
<th>Id (edge ID)</th>
<th>Score (PERCENT RANK) = Weight</th>
<th>Key_Event_Date (Background and Summary Timeline of Key Events (Pages 15-21))</th>
<th>Days_to_Crisis (crisis_date – Key_Event_Date)</th>
<th>Crisis_Date (fixed)</th>
<th>Combined_Indicator (days * significance)</th>
<th>Message Significance to Public Health (Low=1, Medium=2, High = 3, Critical = 5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.864</td>
<td>1/1/13</td>
<td>1003</td>
<td>10/1/15</td>
<td>1003</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0.922</td>
<td>1/2/13</td>
<td>1002</td>
<td>10/1/15</td>
<td>2004</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>0.951</td>
<td>1/23/13</td>
<td>981</td>
<td>10/1/15</td>
<td>2943</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>0.99</td>
<td>1/23/13</td>
<td>981</td>
<td>10/1/15</td>
<td>4905</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>0.951</td>
<td>1/23/13</td>
<td>981</td>
<td>10/1/15</td>
<td>2943</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>0.864</td>
<td>1/1/13</td>
<td>1003</td>
<td>10/1/15</td>
<td>1003</td>
<td>1</td>
</tr>
</tbody>
</table>

Once these values are finalized they should be copied into a ‘clean’ worksheet in Excel, which includes: Source, Target, Type, Id, and Weight columns. This .csv file is provided with the case study materials in the Network_Data folder (edges.csv). In the Score column, select the entire column, right click and select “Paste Special...”, select “Values” and click OK. In the weight.xlsx file, columns A, B, C, D and E should be copied into a new sheet. The new sheet should be Saved As a .csv file (edge.csv). This will give a final static number which can be imported into Gephi. The instructions for using this file to populate the network are in Section 4.3.
4.3. Getting Started with Gephi

Gephi is an open source graph (network) visualization and manipulation software for data analysts to discover patterns in network data (Bastian et al., 2009). Gephi was selected because it can be installed on a personal PC or Mac, and has enough of the basic statistical tools needed for this case study. Furthermore, it allows students to explore the data set and develop new and interesting ways to visualize the organizational communication network.

Java may be required on your system prior to installing Gephi (https://java.com/en/download/).

For questions about Gephi, please first refer to the online forum (http://gephi.forumatic.com/)

1. Download Gephi: https://gephi.org/ (Mac or PC)
2. Launch Gephi:
3. Click: Open Graph File…
4. Navigate to: Flint_Case_Study > Network_Data, and open ‘Flint_Case_Study.gephi’
5. The data and a default window configuration should open, as seen here:
*IF the windows do not appear as above*, click ‘Window’ and make sure the following windows are selected: Appearance, Layout, Graph, Context, Filters, Statistics, and Timeline.

*IF the .gephi file does not open*, you can use the data files to generate the network. Simply go to the ‘Data Laboratory’ window and click ‘Import Spreadsheet’, click the [...] button and find ‘node.csv’ in ‘Flint_Case_Study > Network_Data > Node_Edge_Files’. Select node.csv. In the field for ‘As table:’ select “Nodes table”.

![Import Spreadsheet Window](image)

Keep All Defaults, Click Next, and Click Finish

Repeat the same process but instead select the edge.csv file, and in the field for ‘As table:’ select “Edges table”.

![Import Spreadsheet Window](image)

Keep All Defaults, Click Next, and Click Finish

### 4.4. Dynamic Network Features

This network is equipped with a temporal dimension to show how the network changes over time. To enable the temporal dimension of this network, you must click the green plus sign button on the bottom of the window, as seen here:
**IF you do not see the Timeline button**, you may need to enable the window by selecting Window > Timeline, as seen here:

Once the timeline is enabled, it should look like this:

*IF the dates of the timeline are not as seen above*, you may need to go into the ‘Data Laboratory’ window, Click the ‘Configuration’ button:

Now, back in the ‘Overview’ window showing the network, you can adjust the slider size and move it forwards and backwards to explore the network evolution over time.
4.5. Network Data Analysis

Gephi is equipped with basic statistical tools and allows students to explore the data set to derive new insights into how Organizational Communication theories relate to the Flint Water Crisis.

4.5.1. Modularity

In the Statistics window, Run Modularity, keeping default settings. Interpret the results.

“A value of 0 means that the links in the network fall between the nodes, without any regard to the group structure, i.e. randomly. This implies that the groups identified by the division of the network in that particular fashion are not very strong. A high modularity is desirable (1 is the maximum value) since this indicates that the identified groups are not likely to have occurred by chance.” (Uhr, 2009)

4.5.2. Average Degree

In the Statistics window, Run Average Degree for Network Interpretation: Directed. Once this has run, go to the ‘Data Laboratory’ window, and in the Node Table, new columns for the score of ‘Degree, In-Degree, and Out-Degree’ have been calculated. You can click the top row to order them from lowest to highest or vice versa. Now, in the ‘Overview’ window, under ‘Appearance’, ‘Node’ ‘Attribute’ includes Degree, In-Degree, and Out-Degree, visualize the graph using these parameters. Interpret the results.

“Highest average degree shows the most active organization, which is the one that has many connections or communications with more organizations regardless of direction of the connection.” (Abbasi and Kapucu, 2012)

“Out-degree computes the number of links sent to another actor, while In-degree refers to the number of links received by each actor.” (Valente, 1995)

4.5.3. Eigenvector Centrality

In the Statistics window, Run Eigenvector Centrality for Network Interpretation: Directed. Once this has run, go to the ‘Data Laboratory’ window, and in the Node Table, a new column of ‘Eigenvector Centrality’ has been calculated. Now, in the ‘Overview’ window, under ‘Appearance’, ‘Node’ ‘Attribute’ includes Eigenvector Centrality, visualize the graph using this parameter. Interpret the results.

“Balance theory has been a powerful approach to networks with positive and negative relations. Balance theory is based on the assumption that friends of friends are friends, enemies of friends and friends of enemies are enemies, and enemies of enemies are friends.” (Bonacich, 2007)

“Eigenvector centrality represents the degree measure of a node in terms of the centrality of adjacent nodes (Borgatti and Everett, 1997). Thus, an individual is higher in eigenvector centrality if he or she is directly connected to others who are high in degree centrality.” “Higher centrality scores correspond to higher levels of embeddedness.” (Chenwning, 2009)
4.5.4. Graph Density

In the Statistics window, Run Graph Density for Network Interpretation: Directed. Interpret the results.

“Density describes the general level of linkage among the actors in a network (graph) (Scott, 1991). The more actors connected to one another, the denser the network is. It is the proportion of existing links, to all possible links. The densest network is the one in which all points are connected with each other but such networks are very rare. The precise definition of density can be shown as below, where \( L \) is the total number of links (ties) among actors in the network and \( n \) is the number of actors and thus the denominator is the maximum possible number of links in the network.” (Abbasi and Kapucu, 2016)

\[
Density = \frac{L}{\frac{n(n - 1)}{2}}
\]

4.6. Exploratory Analysis

Gephi is a tool for exploring your network data. For example, in the Layout window, there are several algorithms for bringing visual emphasis to the data (e.g., Forced Atlas to ‘spatialize’ small-world, scale-free networks (Jacomy et al., 2014)).

Try some of the following techniques to see if there are any interesting patterns in this dataset.

4.6.1. Attribute Visualization

In addition to the visualization approaches in 4.5.1, 4.5.2, and 4.5.3, here are some additional techniques which can be used:

**Visualize Network by Attributes (e.g., Node Type):**
Show the edges only between types (e.g., Organizations):

Using the Filter window, Select Attribute > Equal > type (String (Node)). Below in the Equal (type) Settings, type “Organization”, Click OK and Click “Filter” button. See the associated screen shots below:

Students are encouraged to explore different ways of visualizing the case study data in Gephi. For more instruction, please see the Gephi tutorials on visualization: https://gephi.org/users/tutorial-visualization/

For questions about Gephi, please first refer to the online forum (http://gephi.forumatic.com/)
5. Interpretation of Results

This section provides insights from Organizational Communication literature which should be considered in the interpretation of the results. Theory of organizational communication and network analysis is still evolving, and therefore, students are encouraged to explore additional literature on the topic, for example: (Borgatti and H alb gín, 2011).

For the interpretation of results, students are encouraged to challenge the assumptions of this model, and can start by considering the following:

- Is this analysis approaching the Flint Water Crisis in an etic or emic way?

5.1. Modularity

- How modular is the Flint network (i.e., could the network be separated and recombined to work more effectively (Karim, 2006))?
5.2. Average Degree

- What does Average Degree indicate about the three mechanisms of isomorphism? (DiMaggio and Powell, 1983; DiMaggio and Powell, 2000)

- Which predictors of isomorphic change (DiMaggio and Powell, 1983) can be attributed to the Average Degree results? Interpret this for Degree, In-Degree, and Out-Degree.

5.3. Eigenvector Centrality

- If an organization is motivated by coordination over adaptation, is the influence of the node on others proportional to eigenvector centrality? (Calvó-Armengol et al., 2015)

- How can centrality be an indicator of integration and coordination in the network (i.e., (Katz and Kahn, 1978))?

- How does centrality indicate the degree of coordination-non-coordination as: “the extent to which the organization confronts a set of environmental entities whose actions are orchestrated or structured” (Scott, 1981)?

5.4. Graph Density

Students should consider the following insights and questions from Provan et. al. (2007) with respect to density:

- “What is the overall level of connectedness among organizations in the network? Are some networks more fully connected than others? And, more importantly, how much density is beneficial versus detrimental to effectiveness of the network?” (Provan et al., 2007)

- Compare and contrast the concept of density of Hannan and Carroll (1992) and density dependence of Hannan and Freeman (1987, 1988) as mentioned in (Hannan, 2005).

- How might density overlap—overlap in the resource positions of sets of organizations (Van de Ven, 1976), have affected the outcomes in Flint? Does density overlap increase over time as suggested in (Provan et al., 2007)?

- How does density change over time?

- How might density relate to the dimensions affecting uncertainty or dependence from Scott (Scott, 1981)?

- How might network density affect the availability of information in Flint (i.e., (Blau, 1977) as discussed in (Gulati and Gargiulo, 1999))?
6. Authoring Your Paper

Students should write a well-argued paper with full supporting evidence, citations from the course readings, and citations of any additional and relevant literature. The paper should follow a standard format: abstract, introduction, methodology, results, conclusion, and references.

6.1. Referencing

Search out relevant publications through the UIUC online library system: http://www.library.illinois.edu/

Be sure to reference any and all materials you use. See UIUC resources for information on citation management software: http://www.library.illinois.edu/learn/research/citation/#managing

Friendly research tip: use google scholar (https://scholar.google.com/) to directly export all citation information (just search the title, download the reference and cite): http://guides.is.uwa.edu.au/c.php?g=324807&p=2178720

Use any citation style you prefer, as long as it is consistent throughout your paper.

e.g., using APA-6th :


For citing this case study, please use the following:

References

