# Resilience Analysis of Water Distribution Network

Dhruv Shah(202103017),* Vatsal Shah (202103022),† Pranav Patel (202103040),‡ and Vraj Thakkar (202103052)§
*Dhirubhai Ambani Institute of Information & Communication Technology, Gandhinagar, Gujarat 382007, India*

The main aim of this project is to focus on applying various concepts of Complex Networks to the Water Distribution Network of Colorado city. We begin by using various measures of network centrality and network similarities to identify potential nodes for reservoirs. We then construct the flow matrix for water distribution in the city and attempt to draw useful inferences into the network using it.

## I. INTRODUCTION

Analyzing water distribution networks is crucial to ensure efficient, reliable water delivery, detect potential vulnerabilities, and optimize resource management within the urban infrastructure.In this project, we apply the concepts of network science to analyze the water distribution network of a city. We focus in the following direction. Initially, we explored different centralities by plotting their phase transition. We want to explore two situations, one where were we all destroying edges from the network at once and a second where we want to look at the problem where we are adulterating the flow of water in the network. We realized that degree centrality is most efficient in breaking down the network completely by disrupting the fewest number of nodes while addressing the first problem. Ideally, we would want to observe the betweenness centrality for the second problem however we propose using the Current Flow Betweenness centrality instead. We then work on the problem of finding the nodes best suited for making reservoirs. We briefly propose two methods for the task which are closely related and quite effective. In the last section of the report, we look at the flow in the network assuming these nodes as reservoirs and other nodes as sink. We iterate over the network removing each node and measuring the disruption in the flow.

## II. WATER DISTRIBUTION NETWORK AND THE DATASET

A water distribution network (WDN) represents interconnected components that facilitate water transport from sources to consumers. Vertices in a water distribution network represent various components such as water sources (e.g., reservoirs, pumping stations), demand nodes (e.g., households, industries), junctions, valves, and storage tanks. Each vertex represents a point in the network where water flow can be controlled, measured, or manipulated. The edges in a WDN correspond to the pipes that connect vertices in the network. The properties of edges, such as diameter, material, length, and roughness, influence the flow characteristics and efficiency of the network.

The dataset represents **the water pipeline distribution network in the city of Colorado**. [1] This dataset comprises two main files as shown in the figure: the first file contains information regarding the node ID along with their corresponding spatial coordinates (x and y coordinates), while the second file shows the connectivity of the network by indicating the presence of edges between specific pairs of nodes and specifying the diameter of the pipes connecting them.

(1786, 3)

| | Node | xcord | ycord |
|---|---|---|---|
| 0 | 3001 | 3220655 | 1419888 |
| 1 | 6006 | 3222695 | 1416954 |
| 2 | 1003 | 3219027 | 1415374 |
| 3 | 6005 | 3221926 | 1417827 |
| 4 | 6011 | 3220433 | 1415395 |

FIG. 1: Node IDs and their respective coordinates.

(1995, 3)

| | Node 1 | Node 2 | Diameter [inch] |
|---|---|---|---|
| 0 | 12112 | 11200 | 8.0 |
| 1 | 12314 | 1172 | 8.0 |
| 2 | 12243 | 11206 | 8.0 |
| 3 | 1171 | 9022 | 24.0 |
| 4 | 1023 | 11097 | 8.0 |

FIG. 2: Pipes between nodes and their diameter.

*Electronic address: 202103017@daiict.ac.in
†Electronic address: 202103022@daiict.ac.in
‡Electronic address: 202103040@daiict.ac.in
§Electronic address: 202103052@daiict.ac.in

## III. CENTRALITY MEASURES

Using centrality measures such as Degree centrality, Closeness centrality, Betweenness centrality and Current Flow Betweenness centrality, we attempt to check the resilience of our network. We use the following idea. First, we remove node with highest centrality measure and note the size of the largest component that remains. We set a threshold according to the 80-20 rule, giving the condition that the network is said to be disrupted when 80% of the original nodes get removed,i.e. the same as the largest component having size of 20% of the original network.

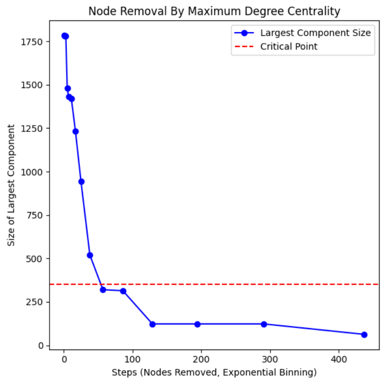


FIG. 3: Exponential Binning Plot for the Size of Largest Component vs Nodes Removed

We observe from Table-I and Fig.3 that, degree centrality performed the best in disrupting the network in the least number of nodes. This means that it is better to disrupt or break the connections at the nodes that have the higher number of connections in the network.

Since in our case the underlying network is a water distribution network, it is crucial to take betweenness centrality into consideration too. Because, the nodes with high betweenness centrality are essential for the flow across the network. If such a node or pipe fails or is disrupted, the impact on water distribution could be severe, as it might create bottlenecks or isolate certain areas of the network. Also, for the cases where we are looking at **contamination spread** through the water network, we would want to look at the nodes that have the highest betweenness centrality.

However, the standard betweenness centrality measure, even though effective, misses the cases when there is flow in the pipes from non-geodesic paths. As in real world water distribution network a node can have flows from different pipes and not just from the geodesic paths where it lies in between. To address this problem Freeman[3] proposed *Flow Betweenness* centrality [3]. The flow betweenness of a vertex $i$ is defined as the amount of flow through vertex $i$ when

| Centrality Type | Nodes Removed |
|---|---|
| Degree | 49 |
| Current Flow Betweenness | 91 |
| Betweenness | 119 |
| Closeness | 575 |

TABLE I: Nodes Removed for Each Centrality Measure

the maximum flow is transmitted from $s$ to $t$, averaged over all $s$ and $t$. This solution is also not favorable because even though it solves the problem of getting contributions from path other than the geodesic path, it brings in this assumption of knowing the maximal flow path. Countering this Newman proposed the idea of **Current Flow Betweenness** centrality also known as the **Random Walk Betweenness** centrality [2].

Now, we will give a brief description of the measure and highlight it's importance in the context of water distribution networks. We consider the water distribution network(n nodes) a linear flow network, i.e. the flow over an edge depends linearly on the gradient of a potential function across the edge. Therefore, concepts of Ohm's Law and Kirchoff's law have analogous counterparts in this contex.

Now, Kirchoff's law implies that the following equation is satisfied.

$$\sum_j A_{ij} (V_i - V_j) = \delta_{is} - \delta_{it}$$

Here $A_{ij}$ are the elements of the adjacency matrix, $V_i$'s are the voltages at each node and $\delta_{ij} = 1$ if $i = j$. We can rewrite these constraints for all the nodes as

$$(D - A)V = s$$

where $A$ is the adjacency matrix, $D$ is the degree matrix and $s$ is a column vector such that $s_i = 1$ if $i$ is the index of the source vector, $s_i = -1$ if $i$ is the index of the target vector and it is zero otherwise. We can impute the values of voltages of source($s$), sink($t$) and assume any vertex $v$ to have zero voltage and solve the system of equations. Given the voltages we can also find the currents flowing in each edge and consequently the current flow betweenness centrality as

$$b_i = \frac{\sum_{s<t} I_i^{(st)}}{\frac{1}{2}n(n-1)}$$

Newman[2] gives another interpretation as well as derivation to this measure of centrality in his paper using the idea of random walk. He defines the current flow betweenness of vertex $i$ to be the net number of times a walk passes. However, we leave the reader to explore the rest of the details in the paper.

Fig.4 is a correlation plot between different types of centrality measures. The key observation to note here is how current flow betweenness centrality correlates well with both betweenness and degree centrality. This shows that it is sort of in the **mid-way** between the two and captures both measures fairly better than any other measure.

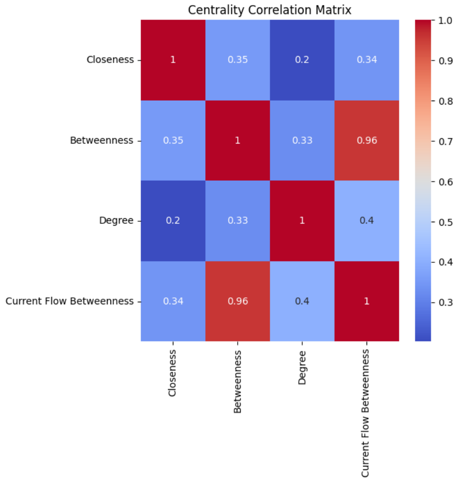


FIG. 4: Correlation Matrix of the Centrality Measures

## IV. IDENTIFYING RESERVOIRS IN THE NETWORK

In this section, we make an attempt to identify nodes which can act as reservoirs in our water distribution network. This is a crucial problem in city planning because proper location of reservoirs helps ensure an unhindered water supply throughout the network.

We assume that we are free to place the reservoirs at any node. Now, to solve this problem we set two objectives. Firstly, the reservoirs should be placed such that they are as close as possible to the homes to which they provide water. Secondly, the reservoirs should not provide water to the same homes i.e. the set of nodes to which they supply water should be disjoint or at least different up to some percentage. This percentage will also capture the resilience of the reservoir network and we can decrease to make the nodes in each reservoir set similar to make the network more robust to reservoir failure.

In the first method, we propose to first pick some n nodes with the highest closeness centralities and look at their structural equivalence(1-ring neighborhood). We would want to pick nodes with the highest closeness centralities and lowest structural equivalence. Extending this idea we can look rather at some r-ring neighborhood of each node and then find structural equivalence(r-structural equivalence) as the number of nodes common in this new set divided by the cardinality of the set of union of r-ring neighbors of both nodes. Through this we can find the nodes which are closest to the other nodes and do not share the same neighbors forming a cluster of sorts. We can choose these nodes as our reservoirs. The value of the threshold for choosing the r-structural equivalence(0 to 1) can be used as a measure of the robustness of the reservoir placement. This is similar to k-means clustering algorithm.

In the second approach, we first run the Girvan-Newman algorithm for community detection. Since this algorithm iteratively removes edges with high betweenness centrality, and given the implications of betweenness centrality in water distribution networks, the algorithm helps identify clusters that can function independently. Therefore, we can set reservoirs in each segment identified by the algorithm.

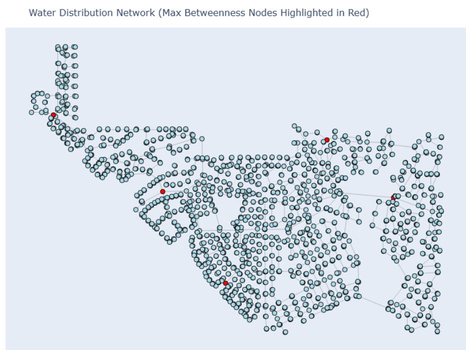


FIG. 5: Identification of Reservoirs(red colored nodes) by first identifying communities and then using the node with highest closeness centrality as the reservoir

## V. FLOW SIMULATION

As established in Section-3, we assume an analogy between water distribution network and the current flow in a circuit. We refer to the paper by *Kaiser et. al* [4] for the detail calculations of Flow through our network using Ohm's law, Kirchoff's law and Graph Laplacian. The difference however is that *Kaiser et. al* [4] did his calculations by removing each node and observing the change in dynamics whereas we observe the change in dynamics by removing each node. The method is described in brief as follows.

We simulate flow in the network by setting a few nodes as reservoirs(which we get from the previous section) and all other nodes as sinks. We assume that the sum of the flow reaching each node should be greater than some threshold. If not then we assume that those nodes are not getting enough flow and the nodes have failed. We try to see that removing which nodes cause the maximum node

| Centrality Type | Correlation with Node Failure |
|---|---|
| Degree | -0.17 |
| Current Flow Betweenness | -0.32 |
| Betweenness | -0.41 |
| Closeness | -0.24 |

TABLE II: Correlation between the importance of nodes obtained by flow simulation and centrality

failures, for each node we get a value n which is the number of disrupted nodes. Furthermore, we can use these values to make a heatmap highlighting the more important nodes. Assuming a threshold we computed the number of disrupted nodes by removing each node and we get a distribution as shown in the figure 6, where most nodes seem to disrupt a similar number of nodes, however, there are a few nodes which show larger/smaller fluctuations from the mean. We also found that betweeness centrality correlates(negatively) the most with the node failures as suggested initially that betweenness is most important while considering the flow. The proposed current flow betweeness centrality correlation value ranges between the degree centrality and the betweenness centrality. (table II)

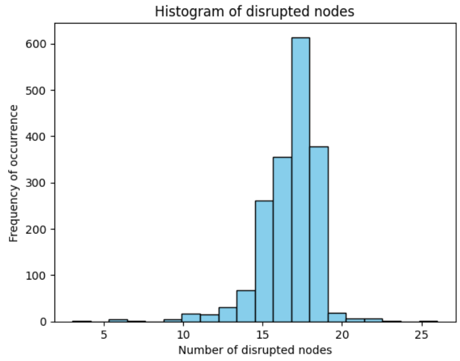


FIG. 6: Threshold of flow for failed node was assumed to be 0.0001(in appropriate dimensions)

[1] The dataset: Centre for Water Systems, University of Exeter, Data set: Colorado springs. Link, Access Date: January 13, 2021

[2] M. E. J. Newman, Social Networks, *A measure of betweenness centrality based on random walks* 27(1):39–54, 2005

[3] Linton C. Freeman, Stephen P. Borgatti, Douglas R. White, *Centrality in valued graphs: A measure of betweenness based on network flow*, Social Networks, Volume 13, Issue 2, 1991, Pages 141-154, ISSN 0378-8733,

[4] Kaiser, F., Latora, V., Witthaut, D. *Network isolators inhibit failure spreading in complex networks.* Nat Commun 12, 3143 (2021).