Estimating the Inspection Frequencies of Civil Infrastructures using Correlation Networks and Population Analysis

Prasad Chetti and Hesham H. Ali
College of Information Science and Technology
University of Nebraska at Omaha
Omaha, NE 68182, USA
pchetti@unomaha.edu and hali@unomaha.edu

***Abstract*— Many recent studies have shown that a large percentage of bridges in many parts of the world have a low safety rating. The national bridge inventory database contains data on more than 600,000 bridges, where each bridge has 116 parameters. Current safety inspections require bridge inspectors to manually inspect each bridge every few years. Manpower and budget constraints limit such inspections from being performed more frequently. More efficient approaches need to be developed to improve the process of bridge inspection and increase the overall safety of bridges and civil infrastructures. In this study, we propose a correlation network model to analyze and visualize the big data associated with more than 600,000 bridges in the national bridge inventory database. We use correlation networks based on various safety parameters, then apply the Markov clustering algorithm to analyze a sub-set of 9,546 steel-stringer/multi-beam or girder bridges. We use the produced clusters to propose a different maintenance schedule based on the bridges that show a higher chance of becoming deficient. Results show that of the top five clusters of bridges, three need to be serviced more frequently. We recommend that their inspection frequency be reduced to 12 months instead of 24 months.**

Keywords*— Correlation Networks; Markov Clustering Algorithm; Structural Health Monitoring; National Bridge Inventory database; Inspection Frequency.

I. INTRODUCTION

Every year, the U.S. Federal Highway Administration (FHWA) records the data of more than 600,000 bridges, with a total of up to 116 parameters in the National Bridge Inventory (NBI) database [1][2]. Detailed descriptions of these parameters can be found in the coding guide [3] developed by FHWA. This is a big data and the authors divided some of the parameters associated with these bridges as internal, external, and outcome parameters. For example, the overall fitness rating/safety rating is an outcome rating/parameter of a bridge and is well reflected by the sufficiency rating (SR). The SR ranges between 0 and 1000. The higher the rating, the better the bridge condition is. The Deck Rating (DR), Structural Evaluation Rating (or Structural Condition Rating (SCR)), and Average Daily Traffic (ADT) are some of the internal parameters that affect the outcome rating, such as, the SR. The DR ranges from 0 to 9 and is used to rate the condition of the bridge. The higher the rating, the better the bridge condition is. SCR is calculated based on ADT and other condition ratings and represents the overall structural fitness of the bridge as given in the FHWA coding guide [3]. Ownership (OW) indicates the owner of the bridge responsible for its maintenance, and Inspection Frequency (IF) is an interval usually given in months to indicate how frequently the bridge gets inspected. The latter two are some of the external parameters. Structural Deficiency (SD) is a status assigned to each bridge to indicate whether the bridge is structurally sound or not.

Structural Health Monitoring (SHM) involves implementing a damage detection and characterization strategy for engineering structures [4]. Current safety inspections using traditional SHM mechanisms require bridge inspectors to manually inspect each bridge every few years. Most of the bridges in the NBI database are assigned a safety inspection frequency of 24 months [27]. However, the 24 months' inspection frequency may not be suitable for bridges that require immediate or more frequent attention due to their age or design standards. Manpower and budget constraints limit inspecting the bridges more frequently. Clearly, more efficient approaches need to be developed to improve the process of bridge inspection and increase the overall safety of bridges and civil infrastructures. As a result, we developed a correlation network model (CNM), based on SR rating values of the bridges for 25 years (i.e., from 1992 to 2016) in our earlier conference paper submitted [1]. The main idea behind this work is to use population analysis to assess the health level of each bridge and predict potential health hazards of bridges before they happen. Population analysis means that analyzing different things based on some particular context. Our main hypothesis is that bridges with similar health fitness ratings are included in a common group/cluster in the CNM and have similar behavioral patterns as discussed in our conference paper [1]. As an extension, we further analyzed these individual clusters and assigned updated inspection frequencies based on their structural health and verified what ratings are being affected by ADT.

Our method takes a population of 9,546 steel-stringer/multi-beam or girder design bridges across three states of the USA, such as, California, Iowa, and Nebraska, which come from three different climatic regions [9] of the country as shown in Figure 1. The reason for selecting bridges from different climatic regions is to study whether temperatures play any significant role on the bridges' health in these states.

Initially, we created a correlation network of bridges based on their outcome rating, such as, SR, and then applied the Markov Clustering (MCL) algorithm [10] to produce clusters of bridges. The top 5 clusters were considered for further analysis to see what internal or external parameters enriched to each cluster. In addition, the last 25 years' ADTs were considered to determine their effect on SR and other ratings. The correlations between ADT and other input parameters were also calculated to see what input parameters significantly impacted by ADT in a given cluster. Finally, we visualized the clusters of bridges with respect to their structural deficiency.

As it was already mentioned above, this paper is an extension of our conference paper [1] and here we have conducted and added some other experimental results such as the effect of ADT on various ratings of the bridges, and assigning different inspection frequencies to the bridges in the top 5 clusters. The remainder of this paper is organized as follows. Section II gives the background of this research, where different studies on monitoring bridge health and inspection frequencies are introduced. This section also describes the importance of correlation networks in this study and how correlation networks are used in other domains and ends with a brief introduction of how correlation networks are used in monitoring the structural health of civil infrastructures. Section II also presents some of the key concepts used for creating correlation networks and clusters. Section III describes the methodology used to develop correlation networks and clusters. Section IV presents the experimental results of this study. The final section presents the conclusions and future recommendations.

II. BACKGROUND

Several recent research studies have attempted to estimate the inspection frequencies of highway bridges, and argued that there should be a rationale to set inspection frequencies for both aged bridges and new bridges. While newer bridges may only require inspection 24 months or more, older bridges may need to be inspected more often. Similarly, these studies also argued that inspection frequencies may be rationale when considering different structural condition ratings, design standards and risk factors and proposed an assessment procedure to create inspection intervals for steel bridges with fracture critical members [6]. Some authors have estimated inspection frequencies using life-cycle cost analysis for Stay-Cable replacement design [8]. The Bayesian network model used in [17] demonstrates the predictive and diagnostic capabilities of the model to estimate the load ratings of prestressed concrete bridges and is useful for bridge management. Deterioration models were also developed for Nebraska state bridges using input parameters of the bridges to estimate the deterioration of various condition ratings [5]. Similar studies were done in [24] for developing various deterioration curves using historical data of condition ratings of New York state bridges. A neural network with a novel data organization scheme and voting process used in [28] shows that it can identify damages in bridges with 86% accuracy. Studies also compared various distribution methods to estimating the inspection intervals of bridges using statistical analysis and showed that the Weibull distribution is likely the best fit for historical data of condition ratings [7].

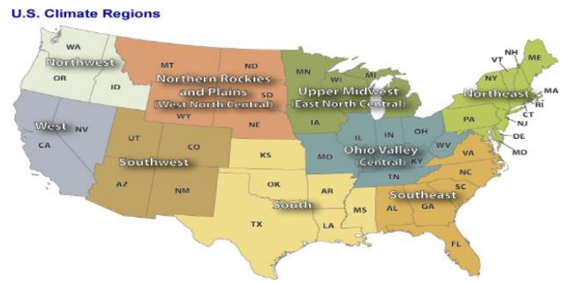

Figure 1. Map of nine USA climate regions (image courtesy : NOAA) [9].

However, all the studies done to date to estimate inspection intervals focused only on individual elements such as deck rating or superstructure rating. Hence, there is a need to estimate inspection intervals based on overall fitness ratings such as SR. If the given set of bridges are clustered based on correlations of historical or time-series data of SR, then individual clusters can be analyzed further to see what parameters are enriched for a cluster of bridges and accordingly the bridge owners may focus on those critical ratings/parameters, and update IF. For example, if a cluster is enriched mostly with structurally deficient bridges, then we can modify the IF of that cluster of bridges to a less than 24 months' interval. Similarly, if a cluster of bridge is enriched with structurally good bridges, then we can update the IF to more than 24 months.

A. Why to utilize a Correlation Network Model?

As the NBI database has data on more than 600,000 bridges, there is a need for powerful and efficient big data tools. CNM is such a powerful big data tool that can predict the structural health of civil infrastructures [1][19]. The key idea behind this work is to use population analysis to assess the health level of each bridge and predict potential health hazards of bridges before they happen. In the population analysis, we compare different clusters or groups of bridges with respect to a particular parameter, based on its enrichment. Our main hypothesis is that the bridges with similar health fitness ratings are included in a common cluster in the CNM and have similar behavioral pattern as shown in [1]. Analyzing these individual clusters will allow us to assign different inspection frequencies based on their deterioration patterns, and structural health.

B. Correlation networks in other disciplines

The ability to show generalization, visualization, and analysis capabilities made the correlation-based network analysis become a powerful analysis tool in biological studies and in various other disciplines [13]. Correlation network analysis was employed in studying biological systems to find the plant growth and biomass in Arabidopsis thaliana Recombinant Inbred Lines (RIL) and introgression lines (IL) [14][15].

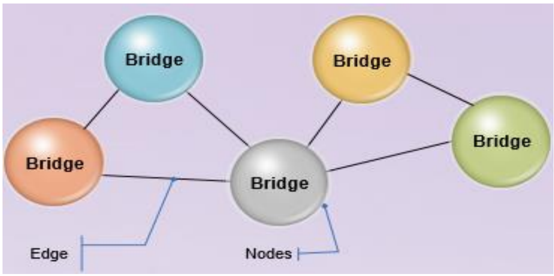

Figure 2. Correlation network representation with bridges.

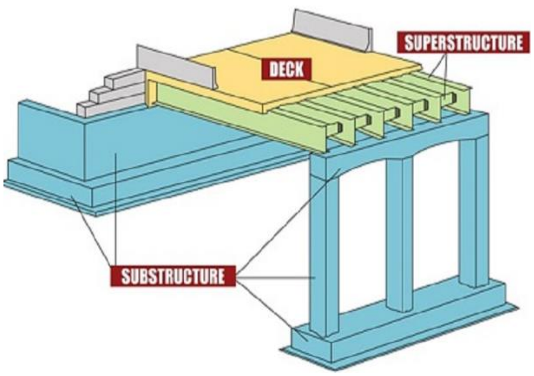

Figure 3. Structural elements of a bridge [22].

It was also successfully applied to evaluate the effect of hypoxia on tumor cell biochemistry [16]. Correlation networks are useful to measure the changes in temporal datasets and study the cluster enrichment by a few Gene Ontology (GO) terms [18].

C. *Correlation Networks for monitoring structural health*

Researchers in the recent past applied CNM to monitor the structural health of civil infrastructures, including analyzing safety issues of various bridges using inventory ratings and other parameters [19]. The basic advantage of using CNM is that the bridges can be clustered together based on some similarity and can be visualized as good and bad bridges [1][19]. As CNM is a new approach for monitoring various civil infrastructures, bridge owners may use CNM to display critical bridges and find an efficient way to improve the inspection schedules [19]. The advantage of this research over [19] is that it considers the temporal-data of SRs. Hence, it can accurately predict an overall fitness rating behavior of the bridges. So, creating a CNM that could deal with temporal-data is one of the objectives of this paper. The motivation of this paper is to process a CNM that could consider bridges' overall behavior (i.e., SR) over a period of time and analyze highly correlated clusters of bridges to predict bridges' behavior and alter the inspection frequencies accordingly. The research question of this paper is to determine what ratings are affected by ADT for each cluster of bridges in the population, if the bridges are clustered using the correlations of temporal data of SRs. The research objective of this paper is to provide a CNM-based decision support System for bridge owners to enable them to find out which bridges need to be serviced first and alter the inspection frequencies. As a result, we developed a novel CNM that considers the temporal data of SRs of the bridges for the last 25 years (from 1992 to 2016), so as to exactly characterize the overall fitness behavior of the bridges over a period of time [1] and see what ratings are effected by ADT, and update the inspection frequencies.

D. *Key concepts used*

The key concepts for this paper are the graph model developing the correlation network, and the Markov clustering algorithm to obtain the group or clusters of bridges.

a) *Graph model*

The CNM is basically an undirected and unweighted graph-based model. The graph is defined as set of vertices and edges, G = (V, E), where V is a set of vertices and E is a set of edges. Each vertex (sometimes called nodes) represents a bridge/civil infrastructure. Two vertices are connected by an edge if and only if their Pearson's correlation coefficient [11] $\rho >= 0.90$, where ρ is a real value, and p-value less than .05. A correlation between any two variables is a value between -1 and +1, which expresses the strength of linkage or co-occurrence. This strength is called Pearson's r or Pearson product-moment correlation coefficient if the correlation is between two continuous–level variables [11][12]. We have used the Pearson's correlation coefficient since the SR data follows normality. Figure 2 represents the undirected and unweighted graph model with 6 nodes and 6 edges. This paper uses bivariate (Pearson's) correlation analysis to show the relationship between any two bridges.

b) *Markov Clustering (MCL)*

In any clustering mechanism, the objects are clustered together in such a way that the distances among clusters are maximized while the distances among the objects are minimized [30] as shown in Figure 4. The Markov Clustering (MCL) algorithm [10] used in this paper is based on the random walks property of the graphs. MCL is a fast and efficient algorithm and is designed for undirected and unweighted graphs. A random walk in a strongly connected cluster usually visits almost all the nodes in the cluster. MCL was already applied on various protein-protein interaction networks and proved to be extraordinarily robust to graph changes and superior in mining complexes from interaction networks [29]. The correlation network that we created for bridges or civil infrastructures is also similar to the protein-protein interaction networks, we employed MCL to mine the groups or clusters of bridges that act similarly.

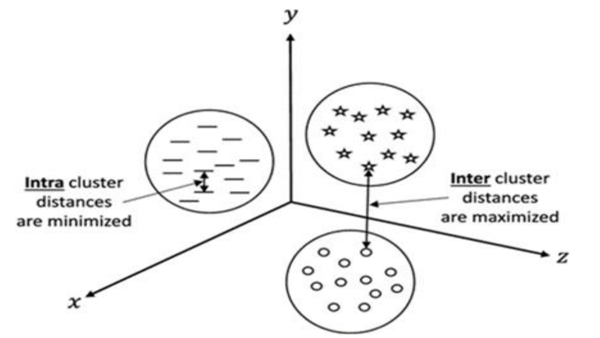

Figure 4. Representation of clustering.

III. METHODOLOGY

The following are the four phases of the CNM that we are proposing.

i. Data acquisition and filtering
ii. Creating a correlation network and applying MCL algorithm
iii. Analyzing various clusters with respect to both input parameters, and output parameters, and comparing various clusters (population analysis)
iv. Developing a decision support system

A. Data Acquisition and Filtering

The highway bridge data is obtained from the NBI database. A total of 25 years (1992-2016) of highway bridge data from three different states of the USA, i.e., California, Iowa, and Nebraska, are considered for this analysis. There are 9,546 highway bridges from these three states, and which are constructed with "Steel material and having Stringer/Multi-beam or Girder design". These are recorded with the numeric 302 in the structure type as given in the FHWA coding guide [3]. The bridges are considered in such a way that their SR is available throughout these 25 years. Hence, each bridge has a minimum age of 25 years. Inconsistent entries, such as the bridges that are recorded as culverts for some years and then non-culverts are removed from the consideration as explained in [1]. Our conference paper explains the data acquisition and filtering in detail [1].

B. Creating a Correlation Network

The SR data of the 9,546 bridges is collected as a matrix along with their 25 years SRs (i.e., from 1992 to 2016). This is called an SR matrix, SR matrix, with each row (i.e., for each bridge) of the matrix having 25 years' SRs in it as a vector. So, there are 9,546 rows in the matrix, A Pearson's correlation coefficients matrix (say, Correlation-matrix) along with the p-values matrix are then obtained over the SR matrix. Each of these matrices are of size 9,546 by 9,546.

Each bridge is then assumed as a node (vertex) in the graph model, and two nodes are connected by an undirected edge if and only if their correlation coefficient $\rho >= 0.90$ and significance value $p < .05$. This creates a correlation network with bridges as nodes along with highly correlated nodes connected by edges as shown in Figure 5. MCL clustering algorithm is then applied in Cytoscape [19] on the previously obtained correlation network to produce clusters. The inflation parameter in MCL clustering can be modified in such a way that the higher inflation value produces clusters of small sizes in terms of nodes. However, we restricted our experiments to the best inflation parameter, such as 1.8, as given in [29]. The clusters produced by MCL algorithm are basically sub-networks of nodes and edges. MCL has produced 8610 nodes in various clusters and 3,865 nodes are present in the top 5 clusters and shown in Figure 6. These top 5 clusters are considered for further analysis. Various experiments are conducted on the top 5 clusters produced by the MCL algorithm, and the results are shown in Section IV below.

C. The parameters considered for analysis

The authors divided the NBI data parameters into three categories, such as, input parameters, output parameters and external parameters—based on their effect on SR as given in the FHWA coding guide [3] to calculate the SR value. SR is an output parameter.

The following are the input and external parameters:

a) *Input parameters:* Average Daily Traffic (ADT), Deck Rating (DR), Superstructure Rating (SpSR), Substructure Rating (SbSR), Structural Condition Rating (SCR), and Water Adequacy Rating (WAR).
b) *External parameters:* State is mentioned as a Location (Loc) (to represent the climatic regions as shown in Figure 1), Owner (OW), Age Category (Age-Cat) derived from the Age (based on Year Built) of the bridge, Inspection Frequency (IF), Rebuilt (RB), and Structural Deficiency (SD) derived from the Status of the bridge.

Some of the above parameters are described as below in the FHWA coding guide [3].

1. Item 1- State Code: considered as the Location (LOC) of the bridge, as the bridges for our analysis were scattered across three states, i.e., California (CA), Iowa (IA), and Nebraska (NE). These three states are from three different climatic regions [9]. California is from the West, Iowa is from the Upper Midwest (East North Central), and Nebraska is from the Northern Rockies and Plains (West North Central) as shown in Figure 1.
2. Item 22- Owner (OW): Maintenance responsibility (Item 21) is used to represent the type of agency

that is the primary owner of the bridge. For example, code '02' in Item 21 is a county highway agency.

3. Item 27- Year Built- records the year that the structure was built. It is used to calculate the age of the bridge. In this study bridges are categorized into three categories based on their age. Ages ranging from 1 to 50 years are in Category A, 51 to 100 years in Category B, and more than 100 years in Category C.
4. Item 29- Average Daily Traffic(ADT): It represents the most recent average daily traffic volume on the bridge.
5. Item 91- Designated Inspection Frequency (IF): It represents the designated inspection frequency of the bridge in months. This interval could be varied from inspection to inspection based on the condition of the bridge. IF=24 indicates that the bridge inspection frequency is for every 24 months as shown in Figure 7.F.
6. Item 106- Year Reconstructed/Rebuilt (RB): It represents the year of reconstruction to keep the bridge operational. RB=0 means that the bridge is not rebuilt. RB=1 indicates that the bridge is rebuilt as shown in Figure 7.E.
7. Status of the bridge: There are four possibilities for the status of the bridges. Status 'N' indicates that it is "Not Applicable". '0' signifies that the bridge is "Not Deficient". '1'indicates that the bridge is "Structurally Deficient (SD)", and '2' means that the bridge is "Functionally Obsolete." A condition rating of 4 or less for Item 58, or 59, or 60, or an appraisal rating of 2 or less for Item 67 or 71, make the bridge structurally deficient [23] as given in [3].
8. Sufficiency rating (SR): It is an outcome measure/rating that is calculated from four factors as given in [3]. It represents the overall fitness rating of the bridge and ranges between 0 and 1000. The lower the rating, the lesser the overall fitness rating is.

IV. EXPERIMENTAL RESULTS

This section demonstrates various experimental results with respect to different network properties, various input, output, and external parameters of the top 5 clusters.

A. *Network Properties of Top 5 Clusters*

The correlation network (correlation $\rho>=0.90$) is presented with 9546 nodes, 767542 edges, and 101 connected components. This is basically a scale-free network. In a scale-free network the degree distribution of network follows a power-law. In a power-law node degree distribution, there are many nodes with fewer degrees and fewer number of nodes with more degrees. The nodes with higher degrees could be acting as hub nodes. The study of a hub node is very important with respect to network properties as this hub node is connected to many other similar nodes or bridges. However, studying those hub nodes is beyond the scope of this paper. Figure 6 shows the top 5 clusters (yellow colored clusters) produced by the MCL algorithm. These clusters' statistics are shown in TABLE 1, with the topmost cluster having the highest number of nodes, which is 1,496 and 354,939 edges, and the smallest cluster having 255 nodes and 13,922 edges. The higher the clustering coefficient [23], the higher the degree to which nodes in a graph are inclined to cluster together. The higher values of the average clustering coefficient for each cluster / subnetwork indicate that the nodes inside each cluster tend to be part of that cluster only. Therefore, the top 5 clusters with higher clustering coefficients are considered for further analysis. TABLE 1 shows that cluster 5 has the highest clustering coefficient, which is 0.838. The cluster density describes the potential number of edges present in the sub-network compared to the possible number of edges in the sub-network. From TABLE 1, we see that cluster 3 has the highest density (0.533) among all the top 5 clusters.

B. *Population analysis with respect to external parameters*

Figures 7.C through 7.F are a comparison of the top 5 clusters with respect to some external parameters such as Age-Category (AGE-CAT), Owner of the bridge (OW), whether the bridge is Rebuilt (RB) or not, and Inspection Frequency (IF). Figure 7.C is a bar chart for comparing the age categories of various clusters. Category A is the set of bridges whose age is 1 to 50 years (labeled blue). B-category bridges are from the age group 51 through 100 (labeled green), and finally the last category, which is Category-C bridges (labeled yellow), with an age of more than 100 years. Cluster 2 is highly enriched with Category-B bridges as shown in Figure 7.C, while the remaining clusters are mostly dominated by both Category-A, and Category B bridges. So, this could have affected the structural deficiency of the bridges as shown in Figure 7.B, where most of the cluster 2 bridges are structurally deficient.

TABLE 1. NETWORK STATISTICS OF TOP 5 CLUSTERS PRODUCED BY THE MCL ALGORITHM.

Cluster Number	**#Nodes**	**#Edges**	**Avg. Degree**	**Density**	**Avg. Clust. Coeff.**	**SR Avg.**
Cluster1	1496	354939	474.51	0.317	0.775	623.7
Cluster2	1180	99000	167.79	0.142	0.674	489.3
Cluster3	634	106955	337.39	0.533	0.823	801.9
Cluster4	300	13377	89.18	0.298	0.812	818.5
Cluster5	255	13922	109.19	0.43	0.838	577.5

Figure 7.D shows that there are six categories of owners of the bridges. However, all the clusters are highly dominated by Owner-2 (labeled green), which is a county highway agency [3]. So, we can infer that all the clusters are enriched and maintained by county highway agency. Figure 7.E is a bar chart that shows whether the bridges were rebuilt (labeled green) or not (labeled blue). If we observe cluster 4, out of 300 bridges, 83 bridges were rebuilt, and most of them were rebuilt in the recent past. The average ratings given in Figures 8.B, and 8.C, show that cluster 4 is a special cluster having lower average ratings in the beginning year 1992 and subsequently increased to high average ratings (especially structural condition rating shown in Figure 8.B) as the number of rebuilt bridges increased in those subsequent years.

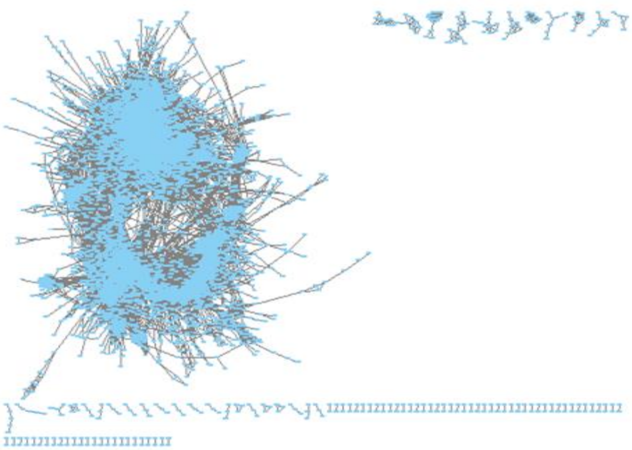

Figure 5. Correlation network (correlation $\rho >= .90$) with 9,546 nodes, and 767,542 edges (Average degree=89.14, and 101 connected components).

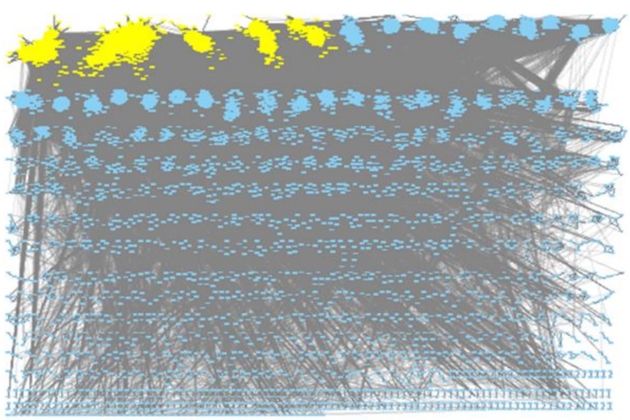

Figure 6. Top 5 clusters (yellow colored clusters) produced by MCL algorithm. (Figure 5 and 6 were generated using Cytoscape [20]).

We also observe from Figures 7.E that most of the cluster 2 bridges are rebuilt, but that have not increased any ratings in Figures 8.B and 8.C. From Figure 7.F, we see that cluster 2 is highly enriched with 24- month inspection frequency. So, we recommend that these bridges' IF's should be lowered, and maintenance must be done more frequently to increase the ratings. An inspection interval of 12 months is more suitable for these bridges.

C. Population analysis of Top 5 clusters with respect to input rating parameters

The input ratings that are considered for population analysis are, DR, SPSR, SBSR, WAR, and SCR of the NBI-dataset-2016. These five different input ratings are compared with respect to their average values in the top 5 clusters as shown in Figure 7. All the average ratings of clusters 3 and 4 are higher compared to all other clusters in top 5. This clearly indicates that these bridges do not have any maintenance issues in terms of any condition ratings in near future. This could be an indication that these bridges' IF's could be updated and increased to either 36 months or 48 months instead of 24 months. Another interesting finding about this figure is that cluster 2 has lower average ratings in all the clusters. For example, cluster 2 has average SBSR value of 4.59. Which is less than 5. According to coding guide [3], these bridges will very soon become structurally deficient. But from Figure 7.F, we see that majority of these bridges' inspection frequencies are 24 months. Hence, we recommend that this cluster's IFs must be lowered to 12 months. Similarly, from Figure 7, if we see the average SCR rating of cluster 2, it is 3.78. which is even below 4. Hence, most of these bridges fell in structurally deficient category. This can be seen from Figure 7.B, as this cluster is highly enriched with SD bridges.

Figures 11 through 15 show the enrichment of individual average input rating values for different clusters. Figure 10 shows that both cluster 1 and 2 are highly enriched with DR = 5. As per the coding guide [3], these bridges are in "Fair Condition" as per the deck is concerned. They are just 1 rating above structural deficiency rating, such as rating 4. This calls the frequent maintenance, such as making IF=12 months for the decks of the bridges in these clusters.

Figures 12 and 13 show the average condition ratings of superstructures and substructures, respectively. From Figure 11, we see that cluster 2 is enriched with SpSR <= 5. Once these bridges' SpSR's drop from 5 to 4, then most the bridges will fall into the SD bridge category. Hence, the improvement in the SpSR rating in terms of dropping the live load is required. This can be done by reducing Average Daily Traffic and implementing required recovery services on these bridges. Cluster 2 from Figure 12 also shows that the substructure rating (SBSR) is critical, as most of the bridges' enriched with SBSR ratings <= 5. Figure 13 shows that all the clusters are highly enriched with water adequacy rating (WAR) >=6. Hence, any improvement in terms of WAR rating is not required for the next couple of years.

From Figure 14, we find that the bridges in clusters 1, 2, and 5 are highly enriched with average structural condition rating (SCR) <=4, which shows that the maximum of the bridges in these clusters are either structurally deficient (SD) or very soon will they fall into this SD category. Hence, more frequent (such as, IF=12 months) maintenance is required for these clusters.

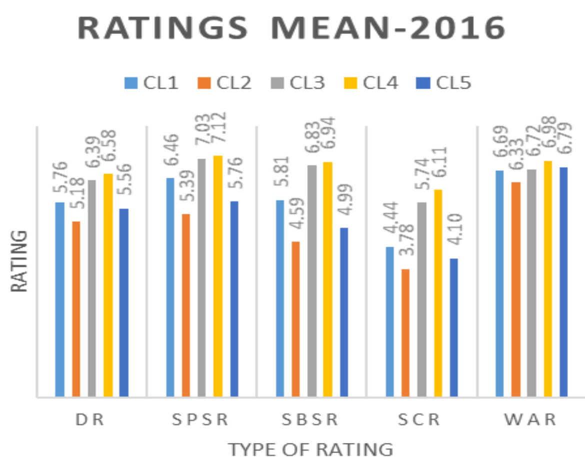

Figure 7. Ratings comparison for top 5 clusters (year 2016).

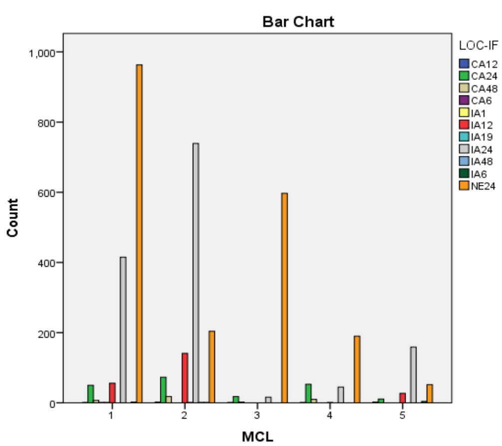

Figure 7.A. Location-Inspection Frequency (Loc-IF).

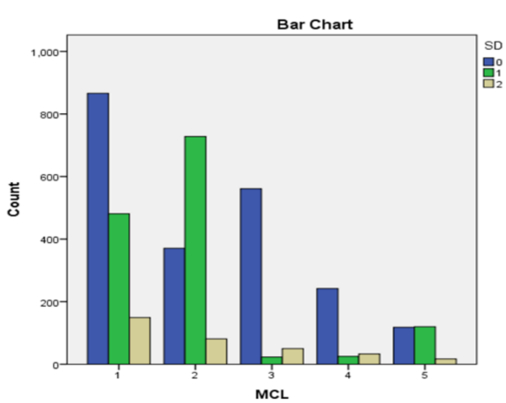

Figure 7.B. Structural Deficiency (SD).

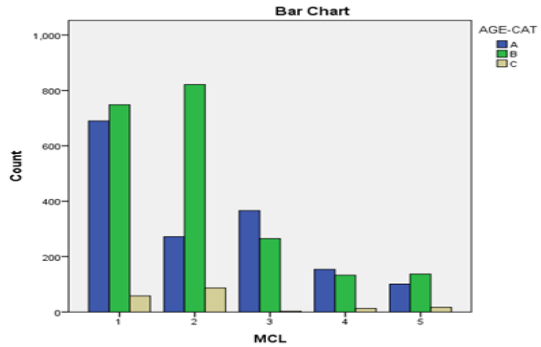

Figure 7.C. Age-Category (AGE-CAT).

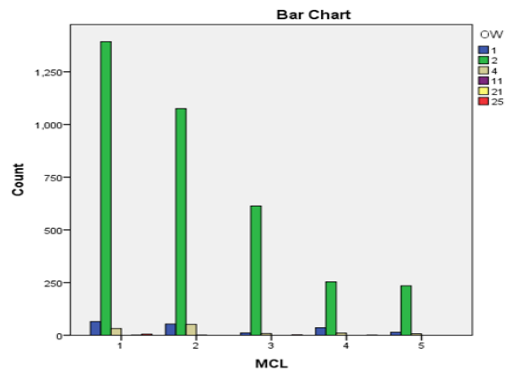

Figure 7.D. Owner(OW).

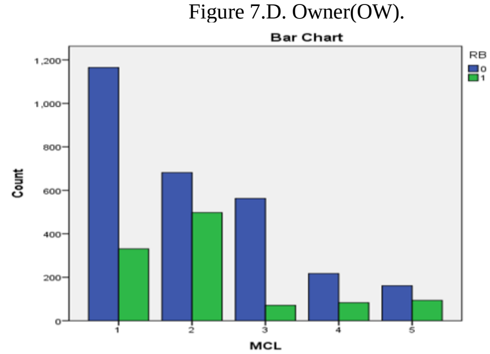

Figure 7.E. Rebuilt (RB).

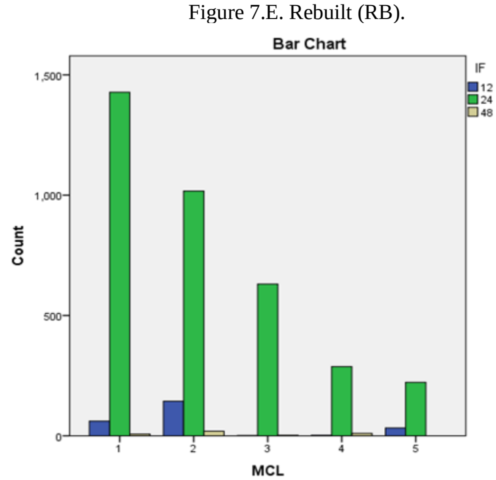

Figure 7.F. Top 5 clusters' Inspection Frequencies (IF). (All these bar charts from Figures 7.A through 7.F were produced in SPSS [25]).

D. Effect of the Average Daily Traffic(ADT) on various ratings

Figures 8.A, through 8.C show the averages of Averages Daily Traffic (ADT), Structural Condition Rating (SCR), and Sufficiency rating (SR), respectively, for the years 1992, 1997, 2002, 2007, 2012, and 2016 (i.e., for every 5 years except for the year 2016, which has a gap of four years from 2012). The ADT graph shown in Figure 8.A indicates that the traffic rate is increasing year by year for any cluster of bridges except for the cluster 2 (between the years 1997 and 2002) and for the cluster 5 (between the years 2007 and 2012). Which clearly indicates that some measures regarding lowering the ADT must have been taken in those clusters in those years to increase the life of the bridges. However, that is not enough to increase the ratings of these bridges at the end year 2016. But at the same time, we see that the increase in ADT, decreases the ratings in case of remaining clusters as we compare Figures 8.A, with 8.B, and 8.C. As these input ratings are reduced, it automatically affects the SR (the overall fitness) and hence, it gets reduced. For clusters 1, 2 and 5, the effect of ADT is huge on both input ratings (for example, SCR), and on output rating SR. As most of the bridges in these three clusters are already structurally deficient, the effect of ADT would be instant. From Figure 8.D, we see that mean ADT volumes increased only due to category D (ADT>=5000) in cluster 1. This indicates that the very high traffic needs to be controlled for the bridges in cluster 1.

E. Analysis with respect to Structural Deficiency

The Structural Deficiency (SD) of the top 5 clusters is shown in Figure 7.B and their individual visualized comparison is shown in Figures 9.A through 9.E, with a total of 61.69% of cluster 2, 47.06% of cluster 5, and 32.15% of cluster 1 bridges are structurally deficient. At the same time, further analysis on cluster 2, as shown in Figures 9.B and 9.F, shows that this cluster is enriched with 62.63% Iowa and 17.29% Nebraska bridges, and both have 24-month inspection frequencies and a mean sufficiency rating below 500. The same is the case for cluster 1. Here, more than 60% of the bridges are from Nebraska with 24-month inspection frequency and having a mean sufficiency rating of just above 600. Hence, we suggest that the 24-month inspection frequency for these structurally deficient bridges needs to be modified to a 12-month inspection frequency to provide rehabilitation services more frequently. Similarly, only 3.63% and 8.33% of cluster 3 and cluster 4 bridges are structurally deficient, as shown in Figures 9.C and 9.D. Cluster 3 and Cluster 4 are enriched with more than 94% and 63% of Nebraska bridges, respectively, with 24-month inspection frequency (as shown in Figure 7.A). Also, the average sufficiency rating of these two clusters is above 800 as shown in Figure 15. Hence, these bridges' 24-month IF can be increased to either 36 months or 48 months as these bridges need not be serviced more frequently.

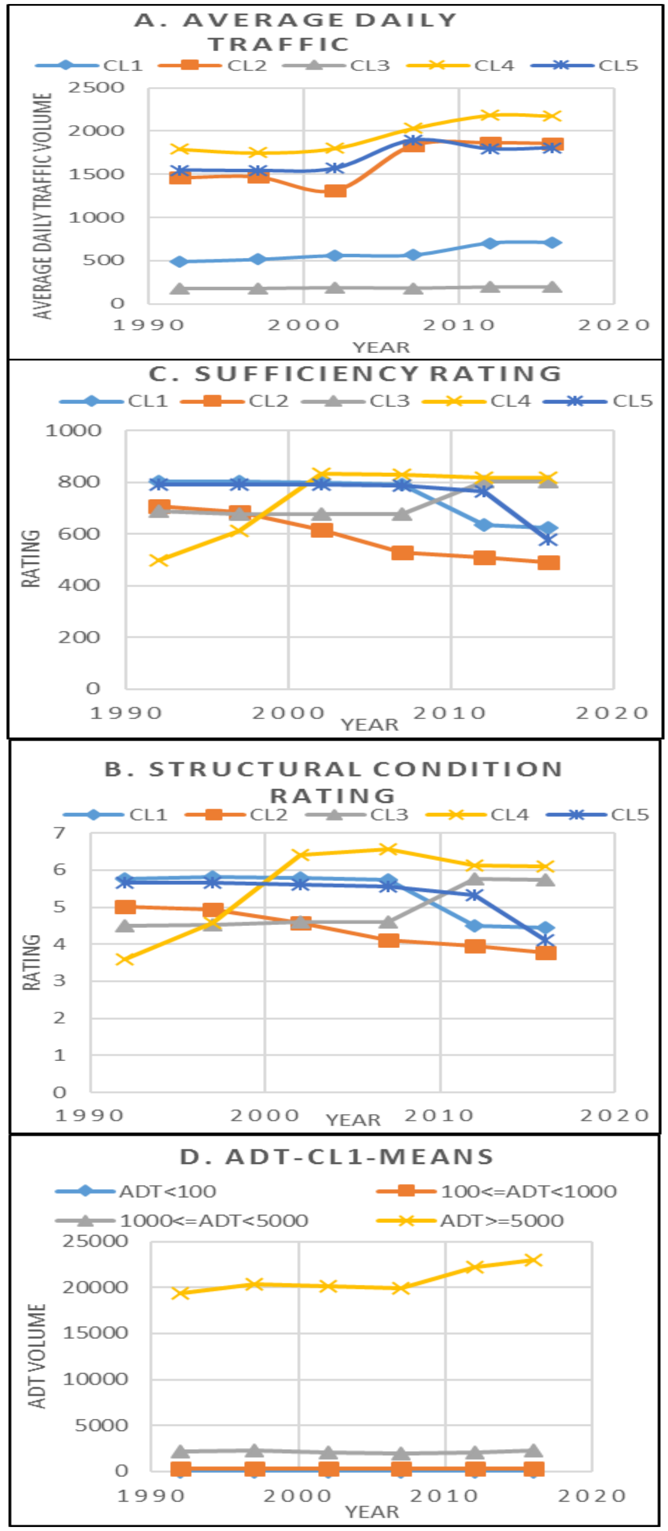

Figure 8. Comparison of top 5 clusters' averages (dataset years 1992,1997,2002,2007,2012 and 2016) with respect to various ratings and average daily traffic. A: Average Daily Traffic (ADT). B: Structural Condition Rating (SCR). C: Sufficiency rating (SR). D: Cluster 1 means of ADT, category wise (four different colors indicating four different categories. Category-A: ADT<100, B: 100<=ADT<1000, C: 1000<=ADT<5000, and category D: ADT>=5000).

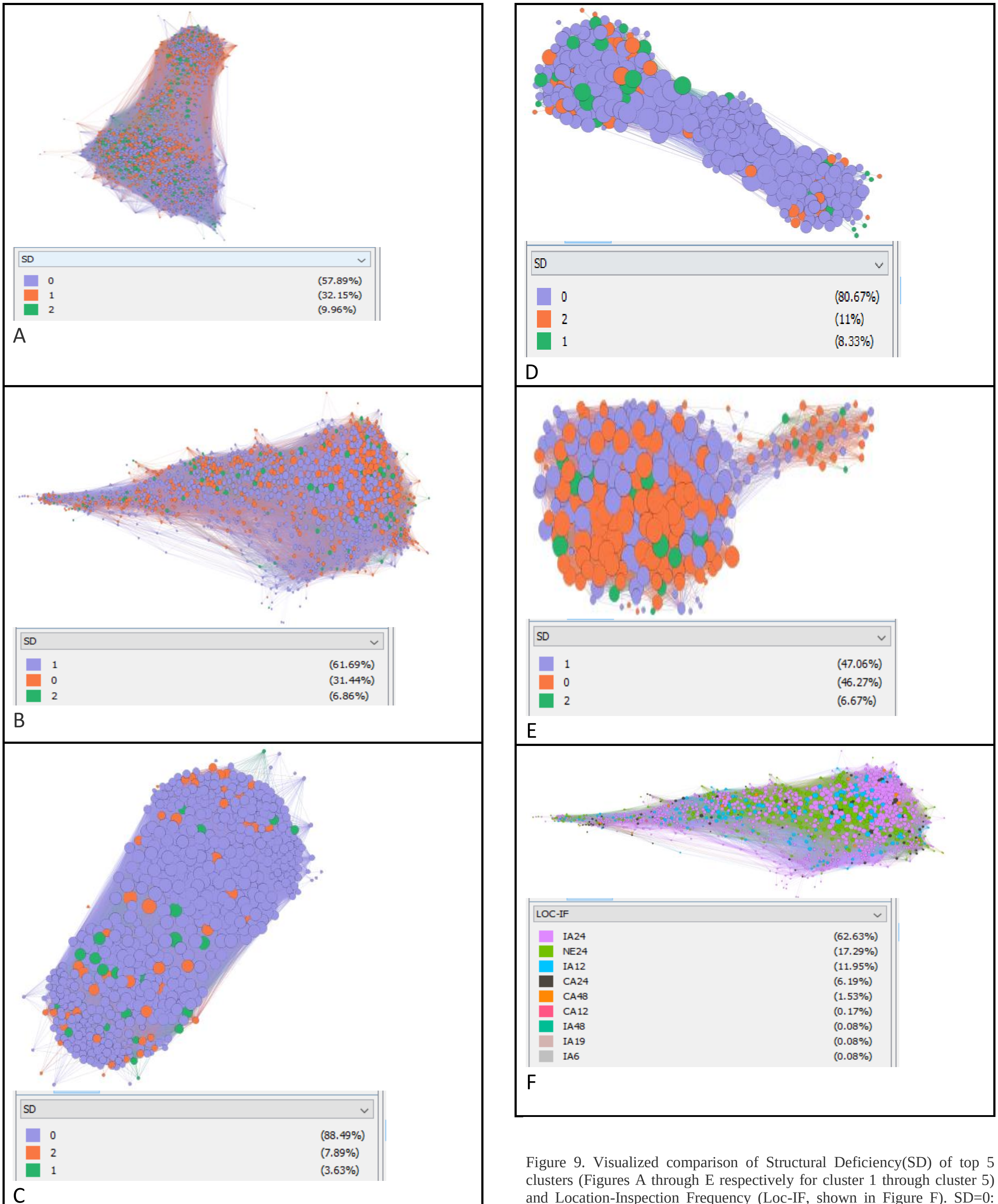

Figure 9. Visualized comparison of Structural Deficiency(SD) of top 5 clusters (Figures A through E respectively for cluster 1 through cluster 5) and Location-Inspection Frequency (Loc-IF, shown in Figure F). SD=0: Structurally Good, SD=1: Structurally Deficient, SD=2: Functionally Obsolete. (The above images were generated using Gephi [21]).

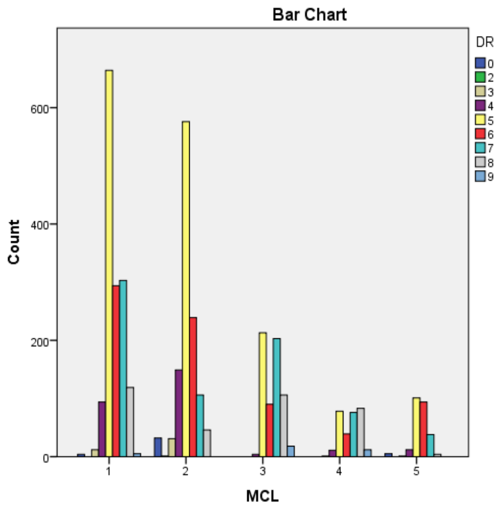

Figure 10. Comparison of top 5 clusters with respect to Deck Rating (DR).

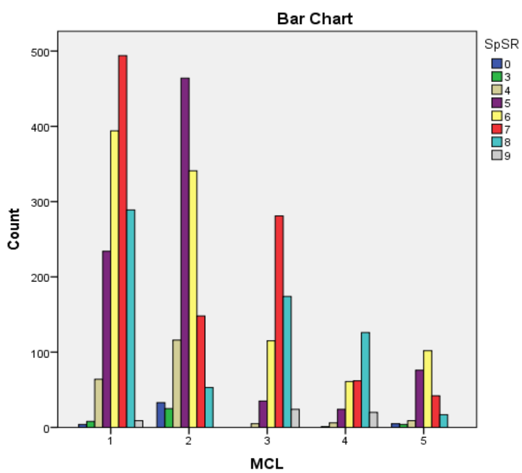

Figure 11. Comparison of top 5 clusters with respect to Superstructure Rating (SpSR).

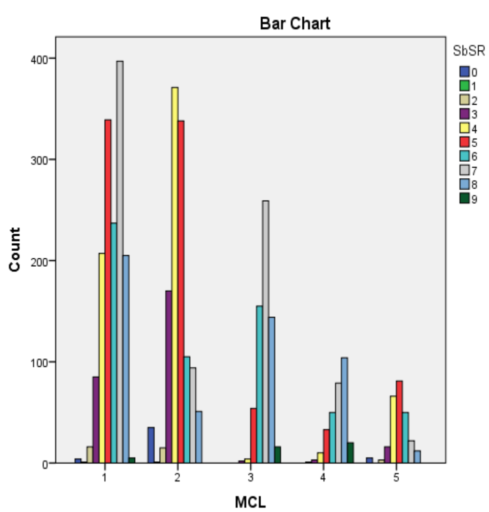

Figure 12. Comparison of top 5 clusters with respect to Substructure Rating (SBSR).

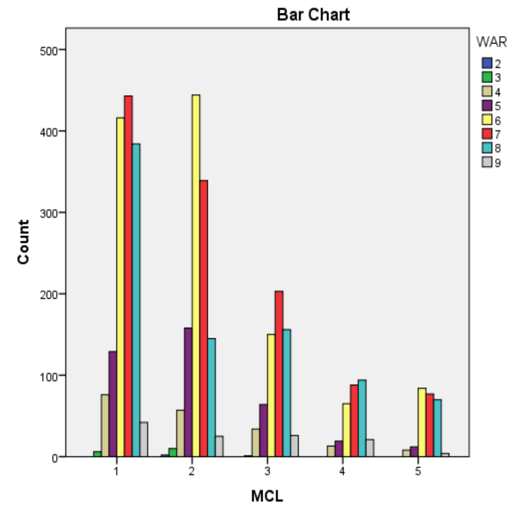

Figure 13. Comparison of top 5 clusters with respect to Water Adequacy Rating (WAR).

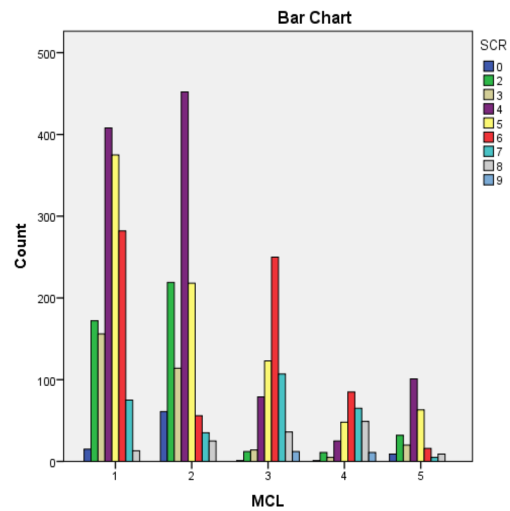

Figure 14. Comparison of top 5 clusters with respect to Structural Condition Rating (SCR).

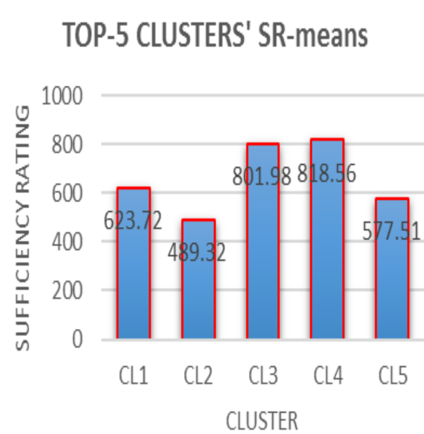

Figure 15. Averages of SRs of top 5 clusters for the year 2016.

V. CONCLUSION

We have presented a model for identifying which bridges need to be serviced first and suggested intervals of inspection using correlation networks and population analysis along with cluster enrichment parameters. Different clusters are enriched with different parameters and/or ratings. Each cluster is analyzed and has given sufficient evidence for our original hypothesis that bridges with similar characteristics/parameters must be part of the same cluster. We found that there are three clusters of bridges (clusters 1,2 and 5), which have low values of sufficiency ratings as shown in Figure 15, and to which immediate rehabilitation services are required. Hence, they need to be serviced first. Therefore, their inspection frequency should be adjusted from 24 months to 12 months. Out of these three clusters, cluster 2 is eligible for federal funding as the average SR for this cluster is below 500 (SR value below 500 makes a bridge eligible for federal funding as per FHWA guidelines [2]) as shown in Figure 15. From Figure 15, we also found that the bridges in clusters 3 and 4 have high overall fitness ratings. Therefore, they do not require immediate attention and their inspection frequencies could be adjusted to 36 months or 48 months instead of 24 months.

We have further analyzed the clusters to see how increased ADT leads to decreased overall fitness (in terms of SR) of the bridges. We have also presented visualizations of all the top 5 clusters with respect to SD to allow different bridge owners to clearly distinguish what bridges are deficient, or functionally obsolete, or in good condition.

From various visualizations and statistics with respect to different input ratings and external parameters, our decision support system could visualize that most of the bridges (built with steel material and having stringer/multi-beam or girder design) from clusters 1, 2 and 5 are in both Iowa and Nebraska states, and with IF = 24 months, and aged above 50 years. These bridges need to be serviced first and their inspection frequencies need to be adjusted to 12 months instead of 24 months.

With all these results, our correlation network model enables various bridge authorities to clearly distinguish between the structurally good and deficient bridges. SHM inspectors can now estimate which bridges' IFs need to be adjusted to 12 months instead of 24 months. Rehabilitation services should be provided accordingly, and authorities can distribute funds on priority basis which could result in saving money and many human lives.

One shortfall of this method is that the big data associated with the information of thousands of bridges may consume more time to create the correlation matrix and correlation network, but with the power of existing supercomputers and their huge memories this couldn't be a big problem.

Presently we have considered only the bridges that are constructed with steel material and with stringer / multi-beam or girder design. As a future work, we may study CNM to see the clustering of various bridges constructed with a different design and/or with different material, and for different states. We would also like to study these clusters further and assign risk rankings to these top 5 clusters, to prioritize the clusters that need immediate rehabilitation services. Further studies may also focus on studying the remaining clusters instead of top 5 clusters. One can also use either DR or SbSR instead of SR, to create the correlation network and provide the inspection frequencies based on the temporal data of these ratings. We can also study the network properties of these clusters in detail to get more insights about these groups of bridges. This work may further be extended for verifying the temperatures role on the bridges as they come from different climatic regions.

ACKNOWLEDGEMENTS

The authors of this paper would like to thank the administration of the University of Nebraska at Omaha, USA as this research work is sponsored by the university under the GRACA research grant with the application number 16796.

REFERENCES

[1] P. Chetti and H. H. Ali. "Analyzing the Structural Health of Civil Infrastructures Using Correlation networks and Population Analysis", Proceedings of the Eighth International Conference on Data Analytics, Porto, Portugal, 2019.

[2] Federal Highway Administration (FHWA), https://www.fhwa.dot.gov/bridge/nbi/ascii.cfm , 2018.08.15

[3] Federal Highway Administration (FHWA), https://www.fhwa.dot.gov/bridge/mtguide.pdf, 2018.08.15

[4] C. R. Farrar and K. Worden, "An introduction to structural health monitoring", Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences 365.1851, pp. 303-315, 2007.

[5] A. Hatami and G. Morcous, "Developing deterioration models for Nebraska bridges", No. M302, 2011.

[6] J. Parr, Michael, J. Conner, Robert, and Mark Bowman. "Proposed method for determining the interval for hands-on inspection of steel bridges with fracture critical members", *Journal of Bridge Engineering*, 15.4 , 2009, pp. 352-363.

[7] M. Nasrollahi and G. Washer, "Estimating inspection intervals for bridges based on statistical analysis of national bridge inventory data," Journal of Bridge Engineering, Vol. 20, No. 9, 04014104, 2014.

[8] A. B. Mehrabi et al., "Evaluation, rehabilitation planning, and stay-cable replacement design for the hale boggs bridge in Luling, Louisiana", *Journal of Bridge Engineering*, 15.4, 2010, pp. 364-372.

[9] T. Karl and W.J. Koss, "Regional and national monthly, seasonal, and annual temperature weighted by area, 1895-1983", 1984.

[10] S. Dongen, "A cluster algorithm for graphs", CWI (Centre for Mathematics and Computer Science), 2000.

[11] K. Pearson, "On the coefficient of racial likeness", Biometrika pp. 105-117, 1926.

[12] J. Benesty et al., "Pearson correlation coefficient", Noise reduction in speech processing, Springer Berlin Heidelberg, pp. 1-4, 2009.

[13] A. Batushansky, D. Toubiana and A. Fait, “Correlation-based network generation, visualization, and analysis as a powerful tool in biological studies: a case study in cancer cell metabolism”, BioMed research international, 2016.
[14] J. Lisec et al., “Identification of metabolic and biomass QTL in Arabidopsis thaliana in a parallel analysis of RIL and IL populations”, The Plant Journal, Vol. 53, No. 6, pp. 960-972, 2008.
[15] R.C. Meyer et al., “The metabolic signature related to high plant growth rate in Arabidopsis thaliana”, Proceedings of the National Academy of Sciences, Vol. 104, No. 11, pp. 4759-4764, 2007.
[16] H.L. Kotze, E.G. Armitage, K.J. Sharkey, J.W. Allwood, W.B. Dunn, K.J. Williams and R. Goodacre, “A novel untargeted metabolomics correlation-based network analysis incorporating human metabolic reconstructions”, BMC systems biology, Vol. 7, No. 1, pp. 107, 2013.
[17] K. LeBeau and W.F. Sara, "Predictive and diagnostic load rating model of a prestressed concrete bridge", *Journal of Bridge Engineering*, 15.4, 2010, pp. 399-407.
[18] K. Dempsey, I. Thapa, D. Bastola and H. Ali, “Identifying modular function via edge annotation in gene correlation networks using Gene Ontology search”, Bioinformatics and Biomedicine Workshops (BIBMW), IEEE International Conference, pp. 255-261, 2011.
[19] A. Fuchsberger and H. Ali, "A Correlation Network Model for Structural Health Monitoring and Analyzing Safety Issues in Civil Infrastructures", Proceedings of the 50th Hawaii International Conference on System Sciences, 2017.
[20] P. Shannon et al., “Cytoscape: a software environment for integrated models of biomolecular interaction networks”, Genome research, Vol. 13, No. 11, pp. 2498-2504, 2003.
[21] M. Bastian, S. Heymann, and M. Jacomy, “Gephi: an open source software for exploring and manipulating networks”, ICWSM, Vol. 8, pp. 361-362, 2009 (retrieved on August 15, 2017 from https://gephi.org/)
[22] Daily Civil, http://www.dailycivil.com/structural-elements-bridge/ 2018.08.15
[23] Federal Highway Administration (FHWA), https://www.fhwa.dot.gov/bridge/britab.cfm 2018.08.15
[24] A.K. Agrawal, K. Akira, C. Zheng. "Deterioration rates of typical bridge elements in New York." *Journal of Bridge Engineering* 15.4 (2010): 419-429.
[25] A. Field, “Discovering statistics using IBM SPSS statistics”, Sage, 2013.
[26] http://journal.r-project.org/archive/2013-1/kahle-wickham.pdf
[27] C.F.R., *National Bridge Inspection Standards*, in *23 CFR part 650*. 2004: USA. p.7441974439.
[28] Li Zhe and R. Burgueño, "Using soft computing to analyze inspection results for bridge evaluation and management", *Journal of Bridge Engineering*, 15.4, 2010, pp. 430-438.
[29] S. Brohee and J. V. Helden. "Evaluation of clustering algorithms for protein-protein interaction networks." BMC bioinformatics, Vol. 7, No. 1, pp. 488, 2006.
[30] A.K. Jain, “Data clustering: 50 years beyond K-means”, Pattern recognition letters, Vol. 31, No. 8, pp. 651-666, 2010.