Water Distribution-Transportation Interface Connectivity Responding to Urban Geospatial Morphology

Noha Abdel-Mottaleb[1] and Qiong Zhang[2]

ABSTRACT

Water distribution and transportation systems are geospatially co-located, forming a network of connections. This network of connections is referred to as an *interface network*. Investigation of interface network connectivity can help understand and minimize failure propagation from water to transportation systems. Water distribution–transportation interface networks consist of nodes, which can be either pipes or roads, and edges, which represent the geospatial co-location of a pipe and road. The purpose of this study is twofold: to topologically represent geospatial co-location by characterizing the connectivity of water distribution– transportation interface networks for multiple cities, and to identify the nodal attributes that are most predictive of a given connectivity profile. A total of forty interface networks from eight cities of varying geospatial morphology are extracted and analyzed using network analysis and machine learning. Using network analysis, we investigate if the topological connectivity between water and transportation is consistent across different cities. Then we use a random forest model to ascertain which nodal attributes may have predictive power to identify the connectivity cluster of the city to which a node belongs. Results indicate that cities of different geospatial morphology may vary in their interface network connectivity, and average shortest path length of

[1]Ph.D. Student, Department of Civil and Environmental Engineering, University of South Florida 4202 E. Fowler Ave., ENB 118 Tampa, Florida, 33620. Email: nohaa@mail.usf.edu

[2]Associate Professor, Department of Civil and Environmental Engineering, University of South Florida 4202 E. Fowler Ave., ENB 118 Tampa, Florida, 33620. Email: qiongzhang@usf.edu

a given node is the major nodal feature contributing to a given city's interface network connectivity. These findings hold implications for urban planning and water distribution design to mitigate potential cascading failures.

Keywords: water distribution networks, interdependency, transportation networks, urban morphology, complexity

INTRODUCTION

Water distribution and transportation systems are both critical infrastructures, crucial and integral to the functioning of cities. All people need access to water and mobility from and to different locations in cities. In fact, a lack of performance of these two types of infrastructures can even be life-threatening. To compound the complexity of these two infrastructure systems, they cannot be analyzed independently because the failure of water distribution networks (e.g. pipe and main breaks) has a direct impact on transportation networks as has been illustrated by traffic disruptions and their subsequent socio-economic impacts (Reed 2017; Chen et al. 2018) and as alluded to in urban design and architecture literature (Alexander 1977; Ahern 2011; Meerow et al. 2016). Further, in addressing or repairing component failures within water distribution, an impact to transportation occurs by way of accessing the underground piping which is often located beneath roads (Mair et al. 2017). This type of interdependency between water distribution and transportation networks, due to co-location, is referred to as geospatial interdependency.

There are many different models to study interdependent infrastructures, depending on the purpose, systems' specifications, and data availability. Some of the more common approaches include: agent-based models, inoperability input output models, systems dynamics, probabilistic methods, multi-objective optimization, network-based approaches, and combinations thereof (Ouyang 2014; Johansen and Tien 2018). Network-based approaches include both structural (i.e. as a snapshot of topology) and dynamics on the network (defining stress/failure states and involving stress/failure propagation probabilities). Further, the coupled networks can be modeled in a variety of ways using a network approach. The choice of model and analysis is in part influenced by data availability/understanding of the system, and also the type of interdependency between the infrastructures, as interdependency can be physical, geo-spatial (i.e. co-location/proximity), cyber, or logical (Gillette et al. 2002; Rinaldi et al. 2001). Because network science is an interdisciplinary field, variations of the same method and type of multi-layer networks have been called by different names in the literature (e.g. influence model, deterministic/probabilistic Bayesian network, coupled network, interlinked network, in addition to multi-layer/multiplex). Multi-layer networks are often used to study the dynamics on interdependent networks, where each infrastructure network layer has its own properties and connections (and thus cascading failure dynamics), in addition to a layer of distinct connections between the two layers (Kivelä et al. 2014; Boccaletti et al. 2014; Bianconi 2018). Researchers have used both real, virtual, or generic network data to construct these multi-layer network models. Findings often show that what constitutes a robust network from analyzing it alone, is different when considering coupled infrastructures together (i.e. interconnected infrastructure networks each can respond uniquely to external events, internal failures, and operation errors) (Haimes et al. 2005; Winkler et al. 2011; Ouyang et al. 2012; Casalicchio and Galli 2008; Rinaldi et al. 2001). Abdel-Mottaleb et al. (2019) have also previously shown that whether or not considering interdependency in management decisions can be dependent on the size/magnitude of failures between water distribution pipe network and transportation road network. Due to the coupling between infrastructures, studies have treated two (or in rare cases more) networks together as multi-layer/multiplex networks, where dependencies are assumed to be either unidirectional (only one infrastructure depends on the other) or bidirectional (both infrastructures depend on each other in some way) (Dueñas-Osorio et al. 2007; Svendsen and Wolthusen 2007; Ouyang and Dueñas-Osorio 2011a; Ouyang and Dueñas-Osorio 2011b; Das et al. 2014; Danziger et al. 2016; Wang et al. 2019; Buldyrev et al. 2010; Korkali et al. 2017; Guidotti et al. 2016; Rahnamay-Naeini and Hayat 2016; Johansen and Tien 2018; Abdel-Mottaleb et al. 2019). Most research on interdependent infrastructures has focused on the failure dynamics of the coupled infrastructure layers, without an analysis of the structure of the interface (i.e. the network consisting of components that directly influence each other from both infrastructure layers) between them. While Winkler et al. (2011) did study the structure of the potable water-power interface for a case study, the interdependency between the two infrastructures is more functional than it is geo-spatial (or due to co-location). Whereas the interdependency between water distribution pipes and transportation roads is heavily co-location based. Furthermore, there have not been studies on how interface structure may vary for different urban development schemes (i.e. geo-spatial morphology). To address these gaps, this study characterizes structural network properties of water pipe-road interface (a network of connections) instead of simulating dynamics on a multi-layer network. We start with a multi-layer network consisting of pipes and roads, but only the pipes and roads that are co-located are extracted to generate a network of the connections of the two layers. We seek to find if the structure of the interdependent linkages between water and transportation is some sort of universal property, independent of a city's geo-spatial layout or if it can vary for cities that developed in different patterns. To date no study has characterised the complex network structure of the *interface* between water distribution pipes and transportation road networks. The interface network, similar to the definition in Winkler et al. (2011) and Ouyang and Dueñas-Osorio (2011a), refers to the network of connections between water distribution and transportation. Nodes in the interface represent infrastructure components from either water distribution or transportation (i.e. pipes or roads) and edges represent potential failure propagation, due to geospatial interdependence.

The water-transportation interface as a logical network is derived from the physical water distribution infrastructure and surface transportation infrastructure. The topology and structure of a given infrastructure network is constrained by geospatial boundaries. Thus, geospatial morphology is a contributor to the form an infrastructure network takes (Louf and Barthelemy 2014). Gastner and Newman (2004) identified geography as a main driver of distribution networks' overall layout and shape. Water distribution networks, specifically, are spatially constrained by the environments in which they are located due to both geology and spatial impediments. Hence, they often trace the same path as other utility networks such as transportation, urban drainage (sanitary sewer and stormwater), and power, which also suffer the same geospatial constraints (Zischg et al. 2017; Yang et al. 2017). Spatial impediments are not only a result of a city's geography, but its morphology (i.e. how it's form developed and continues to do so over time). The layout of a city's infrastructure networks evolves as a spatial and temporal event, in so much as it is driven yet constrained by a city's morphology. This evolutionary morphologic phenomenon yields distinct urban forms throughout the world. Porta et al. (2006) identified six general city forms in their research on networks in urban design: medieval, grid-iron, modernist, baroque, mixed, and lollipop (post-war, low density sprawled suburbs). Different urban forms cities may take, and the properties or charactertistics of each, have been identified and studied by (Boeing 2017; Boeing 2019; Alexander 1977)

In this vain, cities with different urban forms around the world will have infrastructure networks with varying layout. However, this does not guarantee that infrastructure interfaces (i.e. network of connections between two or more infrastructure networks) vary in their network characteristics. Given that water distribution and transportation systems are as interdependent as they are regarding potential failure propagation from water to transportation, it is of importance to investigate structural properties of the interface network between them. Network structure is integral to understanding interdependent and complex infrastructure systems; information such as connectivity can be extrapolated from structure. Similar to the limitations imposed on entire cities' layout and shape by geospatial form, infrastructure networks' efficiency (i.e. resistance to failure) has been linked to structural properties in both individual and interdependent infrastructures (Winkler et al. 2011; Porta et al. 2006).

Within network analysis, observations can be made at different, often compartmentalized into three scales: micro, community, and macro (Soundarajan et al. 2014). Soundarajan et al. (2014) classified the different scales based on operation level. For example, micro scale refers to nodal measurements (i.e. measurements of the constituent parts), while macro refers to network wide, global measures. Community scale refers to the nodal neighborhood level. Complex systems cannot be understood by solely examining and analyzing their constituent parts – or micro-scale analyses; they are defined more by their internal relationships than by their constituent parts. Network connectivity quantitatively captures the extent of relationships between constituent parts (i.e. nodes) of a network. Moreover, using a combination of observation across scales, patterns can be found that relate micro and community scale measures to macro- scale phenomena, and thus allow for predictions to be made regarding network-wide properties from measured nodal characteristics. This is relevant to infrastructure operation and management, because decisions and design are enacted on a component-level, with aim of yielding particular network-wide or global characteristics. For example, the work of Porta et al. (2006) showed that micro-scale streets in transportation networks that were locally efficient yielded an overall transportation network that also tended towards global efficiency. Also in the case of water distribution-transportation interfaces, a decrease in failure propagation between the two networks is a desirable network-wide goal. To understand such network structure properties, complex network theory and modeling has been applied.

Complex network analysis has been used to study the structure of transportation road networks (Louf and Barthelemy 2014; Boeing 2017; Zhang et al. 2015) and water distribution networks (Yazdani and Jeffrey 2010; Yazdani and Jeffrey 2012; Giustolisi et al. 2017; Giudicianni et al. 2018; Di Nardo et al. 2018). Such an analysis has not been conducted for the water distribution-transportation interface.

To address this limitation, this study seeks to investigate and characterize connectivity of the interface between water and transportation networks among different cities, which may be indicative of potential failure propagation from water distribution to transportation. Then, this study uses that knowledge to identify micro-scale features that best predict macro-scale network wide connectivity of the interface between water and transportation networks. To this aim, network analysis is fortified with machine learning to be able to identify patterns emerging from local (micro-scale) features into global (network-wide) properties.

METHODOLOGY

In this study, higher network connectivity between water distribution and transportation infrastructure networks is assumed to indicate higher propensity for failure propagation. Due to the inevitable co-location of water distribution and transportation road networks, it was not clear if the interface network between the two infrastructures could have different connectivity for a city in which the infrastructures were located. To investigate this, several different measures of global, network-wide connectivity are used. Four topological connectivity metrics are computed for 5 samples of a set of 8 cities. Five samples are selected for statistical robustness, to account for the variability driven by different possible layouts for a water distribution network within a single city.

The workflow of the methodology is delineated in Figure 1. The study begins with the selection of cities and results in assessment of network-wide interface connectivity and identification of a nodal predictor that predicts interface connectivity.

FIG. 1

City Selection

City selection ensures the analysis of a representative set of geospatial morphological conditions constraining water distribution networks' layout. A literature review is conducted to choose cities that are identified as differing in age and geomorphology, and also to allow for extending research and knowledge that is already present (i.e. to deepen the body of knowledge) (Alexander 1977; Boeing 2017; Porta et al. 2006). Within the urban planning and transportation body of literature, city shapes are distinguished: older and circular; newer grid like; coastal, linear, tree like; mix of dense old circular city and sparser extensions, thereby allowing the selection of eight cities of varying spatial geometry and evolutionary development. The eight selected cities are: Boston, Chicago, Portland, and Atlanta within the USA; Rome, Italy; Paris, France; Dubai, UAE; and Alexandria, Egypt.

Maps of the studied cities are shown in Figure 2. Polar histograms of street orientations are inlaid in the maps using the opensource Mapbox tool by Agafonkin (2018) with map data from (Mapbox 2018; Contributors 2012). Though street orientations do not completely represent the complexity of the studied urban areas' layouts, they highlight some of the significant differences in their development patterns. For example, Portland is the most "grid-like" city, and that is shown in its polar histogram because the rainbow bars are concentrated along four directions. However, Rome is the most "circularly" laid out, as can be seen from its histogram indicating that streets are oriented in all directions.

FIG. 2

Data Collection

Transportation road network data for each respective city is extracted from Openstreetmap (Contributors 2012). Five water distribution network models for each city of study are generated using an open source virtual water distribution network generator (DynaVIBe) developed by Sitzenfrei et al. (2010). Each of the five water distribution pipe networks from DynaVIBe comprise the different samples for each city, because they vary in their spatial distribution (Abdel-Mottaleb and Zhang 2019). Input for DynaVibe includes digital elevation models, demand volume, and cyclicity percentage. For each city, digital elevation models are downloaded from the Consortium for Spatial Information database. For hydraulic consistency, a total demand of 1083.3 l/s is used for all of the cities' network generation (as was used in previous studies since Quindry et al. (1981)), and a constant cyclicity value of fifty percent is input. For a given city and water network layout, the water distribution network is converted to a shape file and imported into the ArcGIS user interface, and the transportation network for the same city is also imported into GIS for geoprocessing.

Interface Network Extraction and Model Construction

After importing the two shape files (water distribution and transportation) into the same ArcGIS map, a buffer is added to account for lane width (road lane information is present in Openstreetmap GIS files). After adding the lane width buffer, both networks are intersected using the intersect geoprocessing tool, resulting in a new feature layer containing a subset of the original pipes and roads. After obtaining the intersect of the overlaid pipe and road networks (shown in Figures 3 and 4), the interface network is generated from the logical implications in the intersection: a pipe failure will cause a road failure if it is below the road (or over a longer time, vehicular loads on roads can stress water pipes). Thus, a node represents the pipe, another node represents the road, and the causal link is represented by an edge between them. The attribute table of the interface network consists of an edge list (i.e. a list of the edges, the connections between water and transportation, where one column contains pipes, and another contains the roads that are co-located to them), as shown in Figure 3. This network represents unidirectional propagation of failure from the water distribution pipes to the transportation roads in the event of a pipe failure. The attribute table is exported as a spreadsheet for analysis in the following step.

FIG. 3

FIG. 4

Network Analysis

Classification of similar networks is challenging but using multiple measures for comparison eases and enhances the accuracy of the process (Soundarajan et al. 2014). There are network-wide measures representing facets of connectivity that are assumed to indicate failure propagation, and there are nodal measures for each node in a given network. Thus, two distinct network analyses are conducted: global, network-wide and local, nodal-scale. An example of the distinction is the difference in measurement scale between centralization and centrality; centrality refers to the nodal or component scale (a micro-level measurement) whereas centralization is a global scale network-wide metric (a macro-level measurement).

To conduct network analysis, the interface network for each city is imported as an edge list into Cytoscape (an open source network analysis software by Doncheva et al. (2012)). Within the software, the connected components of each network are automatically identified. Because all of the tested interface networks contain a giant component (containing $>> 50$ percent of the networks' nodes), only the giant component is used in the analysis as is common in network analysis (Kolaczyk 2009). The network analyzer and CytoCluster plugins are used to compute four network-wide connectivity measures: modularity, heterogeneity, centralization, and average number of neighbors (Doncheva et al. 2012; Li et al. 2017). Average number of neighbors is based on the aggregate connectivity of network neighborhoods. It is calculated as the average of the number of nodes that each node in the network is connected to. Centralization and heterogeneity are closely related. Heterogeneity is the tendency of a network to contain hub nodes. Similarly, centralization is the tendency of a network to have central structure, meaning the presence of central nodes relative to how central other nodes are in the network (Freeman et al. 1979). Thus, higher values of heterogeneity and centralization indicate that connectivity is not homogeneous across the entire network. Modularity examines connectivity from the vantage of community structure; it is the strength of network division into communities. High modularity indicates high connectivity between nodes in the same community or module, and sparser connections between nodes of different modules.

The network analyzer plugin within Cytoscape is also used to compute the nodal attributes: in-degree, out-degree, neighborhood connectivity, betweenness centrality, and average shortest path length (Doncheva et al. 2012). The nodal degree is the number of edges incident to a given node, meaning more connected nodes have higher in or out degrees. Neighborhood connectivity is the mean of the number of neighbors for all neighbors of a given node, and so higher values indicate higher nodal connectivity. Betweenness centrality is the fraction of all shortest paths in the network that contain a given node, while average shortest path length is the average length of all shortest paths between a node and other nodes. The higher the betweenness centrality, the more crucial a node is to the connectivity of a network (Freeman et al. 1979; Soundarajan et al. 2014; Newman 2018). The shorter the average shortest path length of a node is, the more highly connected it is.

Instead of using Cytoscape, the networkx package can be used within python for convenience. After obtaining network wide measures, ANOVA and post hoc Tukey HSD analyses are conducted to test if connectivity measures are statistically different among cities using the Scikitlearn package in python, and to identify where the differences are located.

Clustering on Global Measures

The hierarchical clustering of network-wide connectivity in this study is analyzed to discern similar and different networks using a distance matrix based on four network-wide connectivity measures. To eliminate the effects of scale on the analysis, connectivity metrics are normalized prior to calculating a distance measurement between each pair of measurements for all of the interface networks using dynamic time warping. Dynamic time warping is a robust technique that is used in signal processing applications and other classification problems (Mueen and Keogh 2016). The mlpy (machine learning python) was used for this method (Albanese et al. 2012).

Using the Scipy package within python, clustering was conducted to group cities' interfaces based on the distance matrix of the global network-wide measures. A clustering criterion defines obtained clusters, imposing structure on the observations — it is therefore important to evaluate whether the clusters do in fact group similar objects, and distinct clusters are more different (Cesar Jr and da Fona Costa 2009). Ward's linkage was used to cluster the cities' interfaces based on the network connectivity, because it is a dispersion-based clustering approach and is regarded as superior or the best hierarchical clustering technique (Cesar Jr and da Fona Costa 2009). Though Ward's linkage was used, other linkage techniques were tested and provided similar results.

Roads Indegree Distribution

In order to provide some physical understanding of information the interface network provides, and to better demonstrate the data being input into the random forest model, this study quantifies the indegree distribution of roads in the interface network. The indegree distribution is the distribution of the number of roads with a given number of pipes intersecting them. This is an interesting property because irrespective of type of mechanical failure of a pipe, the co-located road(s) must often be disturbed as a consequence. The degrees of the transportation roads are extracted to identify a best fit distribution, via methods explained in Clauset et al. (2009), using the networkx, matplotlib, scipy and numpy packages in python. The scaling parameters are also determined as they reveal unique properties regarding the distributions' shape. In particular, a power law distribution is tested for. If the distribution follows a power law, $y = x^{-\alpha}$, then the scaling parameter, α, can provide insight about the hierarchy of connections within the interface networks (e.g., is there a small proportion of roads interacting with many pipes?).

Random Forest Model and Feature Extraction

Random forest is a powerful machine learning algorithm that uses a combination of many decision trees (hence the name, forest) based on given predictors or classifiers to make a prediction or classification. Using a training set, the model "learns" from past data, and reduces the variance. It is of importance to ensure the diversity of the input training set such that the random forest model is not biased. Due to the non-linear and non-normal nodal network data, and the large number of samples, and multiplicity of classifiers (i.e. the nodal features), a random forest model is used for nodal feature extraction. The total samples for the random forest exceeded 1 million (nodes). The large amount of data allowed for using fifty percent of the data for training the model. The relative importance of a feature is measured by calculating the increase in the model's prediction error after shuffling the feature (Molnar 2019). A feature is important in the model if shuffling its values increases the model error, because the model relied on the feature for the prediction. If shuffling feature values leaves the model error unchanged, then the model ignored the feature for the prediction, indicating the feature is unimportant (Molnar 2019). This shuffling feature importance measurement is described in Breiman et al. (1984) and in Breiman (2001) and Breiman (2002) for random forests. It is called the "Gini importance" or "mean decrease impurity", because it is a measure of the decrease in node impurity, averaged over all trees of the ensemble (i.e. the random forest). The method is implemented in scikitlearn, the python package used to calculate it for this study.

RESULTS AND DISCUSSION

Measures of Connectivity

For each interface sample, network-wide, global measures of connectivity are taken. Figure 5 shows plots of the ANOVA with post-hoc Tukey HSD results for each city. As shown in Figure 5, cities have statistically similar connectivity measures where the bars overlap and where the bars do not overlap, they are statistically different with p value $\leq$ 0.05. From the plots shown in Figure 5, Rome and Portland appear at opposite ends of the spectrum for each connectivity measure. It is also interesting to note that values of sample centralization showed the highest variability among connectivity measures, whereas the average number of neighbors showed the least variability. For Heterogeneity, the samples for Dubai are between Rome, and Chicago and Atlanta. Paris is between Chicago and Atlanta, and Boston and Alexandria. In terms of Modularity, it was surprising that Rome's samples were highest and Portland the lowest. Due to the grid like shape of Portland's infrastructure its individual physical infrastructures (i.e., water distribution and transportation) appear to be more modular than those of Rome. However, the grid-like pattern is no longer present in the interface network when the connections between two infrastructures are examined. Just as surprising is that Portland's interface samples showed the highest centralization, whereas Rome, Dubai, Chicago, and Atlanta have the lowest values for centralization. The centralization values for Paris and Boston are higher, followed by Alexandria. It is interesting that Alexandria's samples are closer to Portland. Even though Alexandria is an older city, it is more coastal and has had newer expansion that may be cause for the similarity. The physical water and transportation infrastructures of Rome have centralization values three orders of magnitude higher than those of Portland (see Appendix), with higher connectivity in the center – so it is interesting that this characteristic does not translate to the interface network. Boston is between Rome and Portland, and the remaining cities were all more similar and closer in value to Rome than Portland. For the average number of neighbors, Alexandria's samples show the lowest values, and Portland's samples have the highest values. The value for Dubai's samples is between Alexandria's, and Rome, Paris, and Atlanta. Chicago's samples fall higher on the spectrum, and Boston's samples are between Chicago and Portland. This result signifies that characteristics of the physical infrastructures network are different from that of the logical interface network and not all cities have the same interface connectivity. As a result, clustering of the cities based on properties of the interface network is useful to investigate infrastructure interdependency and grouping them based on all four connectivity measures is necessary, because the similarities of cities' interface samples was not consistent across the four measures. For example, Alexandria was closer to Portland for two measures, but closer to Rome for the other two measures.

FIG. 5

Interface Connectivity Clusters

The results of the clustering analysis are shown in Figure 6. Two things are clear from Figure 6. Cities can be clustered into three groups (in Figure 6, they correspond to cluster 1 on the left (containing interface samples from Rome, Dubai, Chicago, and Alexandria), cluster 2 in the middle (containing city samples from Alexandria, Atlanta, and Paris), and cluster 3 on the right (containing city samples from Paris, Boston, and Portland)) based on the four measures of connectivity used in the analysis of variance. The three clusters are selected for two reasons: to maximize the distance (i.e. ensure that distinct cluster labels are actually different from each other), and to ensure that the distances obtained from a clustering technique accurately reflect the original data. This is often checked using the cophenetic correlation coefficient, introduced in Sokal and Rohlf (1962). It measures the linear correlation coefficient between dissimilarity between each pair of observations (i.e. original connectivity data) and their cophenetic distance (i.e. the vertical distances between where clusters diverge shown in the dendrogram). The cut dendrogram does have a cophenetic correlation coefficient ≥ 0.75, indicating that the clustering technique accurately represents the data. A few cities' samples (e.g., Alexandria and Paris) are divided between two different clusters. These two cities have a higher variability in their resulting interface network based on the input water distribution configuration. This observation indicates individual infrastructure system design and subsequent layout are driving forces of a city's interface network properties. Consequentially, engineers and urban planners may influence failure propagation from water infrastructure by accounting for a city's interface connectivity when designing water distribution systems. It also appears that cities with similar geospatial morphology tend to cluster together; Alexandria and Rome are closer to each other than they are to Portland. Portland, being the most grid-like shaped city, is closer to more grid-like cities such as Boston. Similarly, Rome, an older, circular city is clustered closer to Chicago which is one of the cities that has previously been identified to be modeled after older European cities in Louf and Barthelemy (2014). Rome, Dubai, Chicago, and Alexandria are clustered together, and are categorized under Cluster 1. Cluster 2 includes Alexandria, Atlanta, and Paris, and Cluster 3 includes: Paris, Boston, and Portland. Thus, for the nodal network measures, each node is labelled with either connectivity cluster 1, 2, or 3 depending on which connectivity cluster its sample is a part of, enabling use of random forest.

FIG. 6

Road Indegree Distribution

In Figure 7, one of the studied nodal attributes (Indegree) that are used in conjunction with the clusters for constructing the model is shown in detail. The significant power law distribution of indegrees for all of the tested interface networks indicates that there are some roads with very high connectivity to water distribution pipes, while most roads are not as highly connected to water pipes. Roads with higher indegree tend to be "secondary", connecting main highways to the neighborhoods. The power law distribution confirms this physical pretext, that has already been demonstrated for water distribution networks (Zischg et al. 2017). But the scaling parameter, the power law exponent, provides nuance. It is significant within the random forest model (as determined from the feature extraction), suggesting the hierarchy of the infrastructure as guided by urban development is an important factor in determining interface connectivity. Hierarchy here is quantified by the power law exponent, because the larger its magnitude, the more gradual the indegree distribution falls. The smaller the exponent magnitude is, the more unevenly distributed the pipe-road connectivity is among the roads, and the less "balanced the interdependency is". For example, Rome has the largest magnitude of the power law exponent ($\alpha = 2.4$), and Portland has the smallest magnitude ($\alpha = 0.95$). This is interesting as Rome and Portland are very different in their spatial layout as shown in their polar histograms in Figure 2, where Rome is most circular, and Portland is most grid-like. Another observation is that for all of the interface networks, the power-law tail begins at an indegree equal or larger than 10 pipes. This means that the power law relationship holds strongest for roads with greater than 10 pipes intersecting them at some location. The characterization of the power law exponents implicitly denotes the relationship between the road size/capacity and their corresponding connectivity to water pipes. A less evenly distributed hierarchy of connections between water pipes and transportation roads (i.e. lower value scaling parameter) suggests that it is possible for large road failures to be induced by seemingly inconsequential failures in the water distribution network. This result would be consistent with the findings regarding urban drainage and transportation from Wang et al. (2019), that localized floods can induce catastrophic road failures for some networks.

FIG. 7

Random Forest Model Prediction

From the random forest model, it is determined that, with 90 percent accuracy, a node's connectivity cluster can be predicted using the input nodal features (betweenness centrality, in-degree, out-degree, average shortest path length, and neighborhood centrality). A confusion matrix contains information pertinent to the accuracy of the model in identifying a connectivity cluster of nodes from nodal properties. The confusion matrix shown in Figure 8 contains more detail on the distribution of model accuracy in distinguishing different clusters from each other. The x-axis of the matrix represents the predicted city connectivity cluster of a sample by the random forest model. The y-axis represents the actual city connectivity cluster. The matrix is read by row, and the diagonal represents the normalized number of samples accurately classified (actual cluster matches the predicted cluster) by the random forest model. The diagonal in Figure 8 has the highest number of normalized number of samples owing to the high accuracy of the model. The matrix also indicates that the accuracy in identifying samples from clusters 1 and 3 is higher than identifying samples from cluster 2. It can also be observed from the 90 percent accuracy that nodal properties can predict a given connectivity cluster; this relates to the micro-nodal scale measures being able to predict global network-wide patterns. The random forest also enables feature extraction of the nodal property contributing the most to the prediction/classification, by shuffling attributes and evaluating the change in model accuracy.

FIG. 8

Feature Extraction

From the feature extraction, for which results are shown in Figure 9, average shortest path length accounts for 60 percent of the relative importance in the identification/prediction of the connectivity cluster a given node belongs to. This is much higher than any other nodal attributes. Average shortest path length has previously been shown to correlate with an individual water distribution network's performance. Specifically, for individual water distribution networks, average shortest path length positively correlates to redundancy and consequently network resilience (Giudicianni et al. 2018). This research extends the predicative utility of average shortest path length of nodes from single infrastructure networks to predict connectivity between interface networks of coupled, co-located infrastructures.

FIG. 9

CONCLUSION

In this work, we conclude that interface networks representing potential failure propagation from water distribution to transportation infrastructures may differ in their network connectivity depending on the water distribution layout for a given transportation network in a city. Practically, WDN design, despite the WDN being co-located, still has sufficient leverage to alter the connectivity profile of the interface between road and water. This indicates that the location/structure of redundancies in a WDN can significantly influence co-location (and thus propagation of failure from one infrastructure to another), as seen for two of the studied cities.

The variation in connectivity allowed the clustering of city interface networks based on connectivity. Further, the study determined that nodal properties, such as neighborhood connectivity and average shortest path length, can predict the connectivity cluster of an interface network that a given node belongs to. The study finds average shortest path length to be the most powerful attribute in predicting connectivity. Future work can validate connectivity clusters with simulation performance data and formulate an optimization model to minimize the failure propagation by changing interface network layouts using extracted nodal feature(s).

The study also found that the directional connectivity of water pipes to roads for the studied cities follows a power law, but with varying scaling or hierarchy. These insights are a step at answering 1) how to mathematically represent connectivity of the coupled interface structure and 2) how much does a given urban development pattern influence the connectivity of water-transportation interfaces. Such understanding can have practical implications in terms of designing the desired interfaces by modifying both urban form/development and some aspects of water distribution networks so that a lower failure propagation can be achieved from a water distribution network to a co-located road network. Future work will investigate network layouts with lower interface connectivity, as that is assumed to propagate less failure from water distribution to transportation infrastructure.

The interdependency between urban water distribution and transportation networks is doubtlessly present, but the extent is not fully understood nor quantified; hence, characterizing the interdependencies between urban water distribution and transportation systems can aid in the optimization of both networks by informing design, operations and maintenance.

DATA AVAILABILITY STATEMENT

- Some or all data, models, or code generated or used during the study are available in a repository online in accordance with funder data retention policies ("Water distribution–transportation interface network data", DOI: 10.5281/zenodo.3381596)
- All data, models, or code generated or used during the study are available from the corresponding author by request (generated interface networks, analaysis of variance, clustering, random forest, and feature extraction).

ACKNOWLEDGEMENTS

This material is based upon work supported by the National Science Foundation under Grant Number 1638301. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. Preliminary results were orally presented at the AWRA GISX Conference in Orlando, FL (April, 2018).

REFERENCES

Abdel-Mottaleb, N., Ghasemi Saghand, P., Charkhgard, H., and Zhang, Q. (2019). "An exact multiobjective optimization approach for evaluating water distribution infrastructure criticality and geospatial interdependence." *Water Resources Research*, 55(7), 5255–5276.

Abdel-Mottaleb, N. and Zhang, Q. (2019). "Water distribution–transportation interface network data, <DOI: 10.5281/zenodo.3381596>.

Agafonkin, V. (2018). "Visualizing street orientations on an interactive map, <https://mourner.github.io/road-orientation-map/>.

Ahern, J. (2011). "From fail-safe to safe-to-fail: Sustainability and resilience in the new urban world." *Landscape and urban Planning*, 100(4), 341–343.

Albanese, D., Visintainer, R., Merler, S., Riccadonna, S., Jurman, G., and Furlanello, C. (2012). "mlpy: Machine learning python." *arXiv preprint arXiv:1202.6548*.

Alexander, C. (1977). *A pattern language: towns, buildings, construction*. Oxford university press.

Bianconi, G. (2018). *Multilayer Networks: Structure and Function*. Oxford university press.

Boccaletti, S., Bianconi, G., Criado, R., Del Genio, C. I., Gómez-Gardenes, J., Romance, M., Sendina-Nadal, I., Wang, Z., and Zanin, M. (2014). "The structure and dynamics of multilayer networks." *Physics Reports*, 544(1), 1–122.

Boeing, G. (2017). "Methods and measures for analyzing complex street networks and urban form." *arXiv preprint arXiv:1708.00845*.

Boeing, G. (2019). "Urban spatial order: Street network orientation, configuration, and entropy." *Configuration, and Entropy (February 1, 2019)*.

Breiman, L. (2001). "Random forests." *Machine learning*, 45(1), 5–32.

Breiman, L. (2002). "Manual on setting up, using, and understanding random forests v3. 1." *Statistics Department University of California Berkeley, CA, USA*, 1, 58.

Breiman, L., Friedman, J., Olshen, R., and Stone, C. (1984). “Classification and regression trees. wadsworth int.” *Group*, 37(15), 237–251.

Buldyrev, S. V., Parshani, R., Paul, G., Stanley, H. E., and Havlin, S. (2010). “Catastrophic cascade of failures in interdependent networks.” *Nature*, 464(7291), 1025.

Casalicchio, E. and Galli, E. (2008). “Metrics for quantifying interdependencies.” *International Conference on Critical Infrastructure Protection*, Springer, 215–227.

Cesar Jr, R. M. and da Fona Costa, L. (2009). *Shape classification and analysis: theory and practice*. Crc Press.

Chen, A., Evans, B., Prior, A., Djordjevic, S., Savic, D., Butler, D., Goodey, P., Stevens, J. R., and Colclough, G. (2018). “Mapping urban infrastructure interdependencies and fuzzy risks.

Clauset, A., Shalizi, C. R., and Newman, M. E. (2009). “Power-law distributions in empirical data.” *SIAM review*, 51(4), 661–703.

Contributors, O. (2012). “Openstreetmap.” *URL www. openstreetmap. org*.

Danziger, M. M., Shekhtman, L. M., Bashan, A., Berezin, Y., and Havlin, S. (2016). *Vulnerability of interdependent networks and networks of networks*. Springer.

Das, A., Banerjee, J., and Sen, A. (2014). “Root cause analysis of failures in interdependent power-communication networks.” *Military Communications Conference (MILCOM), 2014 IEEE*, IEEE, 910–915.

Di Nardo, A., Di Natale, M., Giudicianni, C., Greco, R., and Santonastaso, G. F. (2018). “Complex network and fractal theory for the assessment of water distribution network resilience to pipe failures.” *Water Science and Technology: Water Supply*, 18(3), 767–777.

Doncheva, N. T., Assenov, Y., Domingues, F. S., and Albrecht, M. (2012). “Topological analysis and interactive visualization of biological networks and protein structures.” *Nature protocols*, 7(4), 670.

Dueñas-Osorio, L., Craig, J. I., Goodno, B. J., and Bostrom, A. (2007). “Interdependent response of networked systems." *Journal of Infrastructure Systems*, 13(3), 185–194.

Freeman, L. C., Roeder, D., and Mulholland, R. R. (1979). "Centrality in social networks: Ii. experimental results." *Social networks*, 2(2), 119–141.

Gastner, M. T. and Newman, M. E. (2004). "Diffusion-based method for producing density-equalizing maps." *Proceedings of the National Academy of Sciences*, 101(20), 7499–7504.

Gillette, J., Fisher, R., Peerenboom, J., and Whitfield, R. (2002). "Analyzing water/wastewater infrastructure interdependencies.." *Report no.*, Argonne National Lab., IL (US).

Giudicianni, C., Di Nardo, A., Di Natale, M., Greco, R., Santonastaso, G., and Scala, A. (2018). "Topological taxonomy of water distribution networks. water 10 (4): 1–19.

Giustolisi, O., Simone, A., and Ridolfi, L. (2017). "Network structure classification and features of water distribution systems." *Water Resources Research*, 53(4), 3407–3423.

Guidotti, R., Chmielewski, H., Unnikrishnan, V., Gardoni, P., McAllister, T., and van de Lindt, J. (2016). "Modeling the resilience of critical infrastructure: The role of network dependencies." *Sustainable and resilient infrastructure*, 1(3-4), 153–168.

Haimes, Y. Y., Horowitz, B. M., Lambert, J. H., Santos, J. R., Lian, C., and Crowther, K. G. (2005). "Inoperability input-output model for interdependent infrastructure sectors. i: Theory and methodology." *Journal of Infrastructure Systems*, 11(2), 67–79.

Johansen, C. and Tien, I. (2018). "Probabilistic multi-scale modeling of interdependencies between critical infrastructure systems for resilience." *Sustainable and Resilient Infrastructure*, 3(1), 1–15.

Kivelä, M., Arenas, A., Barthelemy, M., Gleeson, J. P., Moreno, Y., and Porter, M. A.

(2014). “Multilayer networks.” *Journal of complex networks*, 2(3), 203–271.

Kolaczyk, E. D. (2009). *Statistical Analysis of Network Data: Methods and Models* (01).

Korkali, M., Veneman, J. G., Tivnan, B. F., Bagrow, J. P., and Hines, P. D. (2017). “Reducing cascading failure risk by increasing infrastructure network interdependence.” *Scientific reports*, 7, 44499.

Li, M., Li, D., Tang, Y., Wu, F., and Wang, J. (2017). “Cytocluster: A cytoscapeplugin for cluster analysis and visualization of biological networks.” *International journal of molecular sciences*, 18(9), 1880.

Louf, R. and Barthelemy, M. (2014). “A typology of street patterns.” *Journal of The Royal Society Interface*, 11(101), 20140924.

Mair, M., Zischg, J., Rauch, W., and Sitzenfrei, R. (2017). “Where to find water pipes and sewers??on the correlation of infrastructure networks in the urban environment.” *Water*, 9(2), 146.

Mapbox, L. (2018). “Mapbox studio, <https://www.mapbox.com/mapbox-studio/>.

Meerow, S., Newell, J. P., and Stults, M. (2016). “Defining urban resilience: A review.” *Landscape and urban planning*, 147, 38–49.

Molnar, C. (2019). *InterpretableMachineLearning*, <https://christophm.github.io/interpretable-ml-book/>.

Mueen, A. and Keogh, E. (2016). “Extracting optimal performance from dynamic time warping.” *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM, 2129–2130.

Newman, M. (2018). *Networks*. Oxford university press.

Ouyang, M. (2014). “Review on modeling and simulation of interdependent critical infrastructure systems.” *Reliability engineering & System safety*, 121, 43–60.

Ouyang, M. and Dueñas-Osorio, L. (2011a). “An approach to design interface topologies across interdependent urban infrastructure systems.” *Reliability Engineering &*

System Safety, 96(11), 1462–1473.

Ouyang, M. and Dueñas-Osorio, L. (2011b). "Efficient approach to compute generalized interdependent effects between infrastructure systems." *Journal of Computing in Civil Engineering*, 25(5), 394–406.

Ouyang, M., Dueñas-Osorio, L., and Min, X. (2012). "A three-stage resilience analysis framework for urban infrastructure systems." *Structural safety*, 36, 23–31.

Porta, S., Crucitti, P., and Latora, V. (2006). "The network analysis of urban streets: a primal approach." *Environment and Planning B: planning and design*, 33(5), 705–725.

Quindry, G. E., Liebman, J. C., and Brill, E. D. (1981). "Optimization of looped water distribution systems." *Journal of the Environmental Engineering Division*, 107(4), 665–679.

Rahnamay-Naeini, M. and Hayat, M. M. (2016). "Cascading failures in interdependent infrastructures: An interdependent markov-chain approach." *IEEE Transactions on Smart Grid*, 7(4), 1997–2006.

Reed, D. (2017). *Water, Security and US Foreign Policy*. Taylor & Francis.

Rinaldi, S. M., Peerenboom, J. P., and Kelly, T. K. (2001). "Identifying, understanding, and analyzing critical infrastructure interdependencies." *IEEE control systems magazine*, 21(6), 11–25.

Sitzenfrei, R., Fach, S., Kleidorfer, M., Urich, C., and Rauch, W. (2010). "Dynamic virtual infrastructure benchmarking: Dynavibe." *Water Science & Technology*, 10(4), 600.

Sokal, R. R. and Rohlf, F. J. (1962). "The comparison of dendrograms by objective methods." *Taxon*, 11(2), 33–40.

Soundarajan, S., Eliassi-Rad, T., and Gallagher, B. (2014). "A guide to selecting a network similarity method." *Proceedings of the 2014 SIAM International Conference on Data Mining*, SIAM, 1037–1045.

Svendsen, N. K. and Wolthusen, S. D. (2007). “Connectivity models of interdependency in mixed-type critical infrastructure networks.” *Information Security Technical Report*, 12(1), 44–55.

Wang, W., Yang, S., Stanley, H. E., and Gao, J. (2019). “Local floods induce large-scale abrupt failures of road networks.” *Nature communications*, 10(1), 2114.

Winkler, J., Dueñas-Osorio, L., Stein, R., and Subramanian, D. (2011). “Interface network models for complex urban infrastructure systems.” *Journal of Infrastructure Systems*, 17(4), 138–150.

Yang, S., Paik, K., McGrath, G. S., Urich, C., Krueger, E., Kumar, P., and Rao, P. S. C. (2017). “Functional topology of evolving urban drainage networks.” *Water Resources Research*, 53(11), 8966–8979.

Yazdani, A. and Jeffrey, P. (2010). “A complex network approach to robustness and vulnerability of spatially organized water distribution networks.” *arXiv preprint arXiv:1008.1770*.

Yazdani, A. and Jeffrey, P. (2012). “Water distribution system vulnerability analysis using weighted and directed network models.” *Water Resources Research*, 48(6).

Zhang, X., Miller-Hooks, E., and Denny, K. (2015). “Assessing the role of network topology in transportation network resilience.” *Journal of Transport Geography*, 46, 35–45.

Zischg, J., Klinkhamer, C., Zhan, X., Krueger, E., Ukkusuri, S., Rao, P., Rauch, W., and Sitzenfrei, R. (2017). “Evolution of complex network topologies in urban water infrastructure.” *World Environmental and Water Resources Congress 2017*, 648–659.

APPENDIX I. CENTRALIZATION AND ROAD INDEGREE DATA

The centralization of Rome's individual infrastructure networks is $3.5 * 10^{-5}$. The centralization of Portland's individual infrastructure networks is $2.0 * 10^{-8}$. In contrast, the centralization of Portland's interface network is higher than that of Rome's interface network. These values are obtained using networkx within python. Centralization is normalized by the maximum possible centralization for a graph with the same number of nodes (i.e. a star configuration). It is not surprising that the physical infrastructure networks of Rome would have higher centralization values than Portland because Rome's layout more closely resembles a star while Portland's layout resembles a grid. For more on centralization, see Freeman (1979) and Newman (2018).
Figure 8 shows the road indegree distributions for the studied cities. Table 1 contains more information regarding the cities' road indegree distributions (values of the scaling parameter (α), and p-values).

TABLE 1: Power Law Data for Road Indegree Distributions

City	α range	p-value range
Alexandria	$[1.43, 1.53]$	$[10^{-17}, 10^{-13}]$
Atlanta	$[1.80, 1.88]$	$[10^{-26}, 10^{-22}]$
Boston	$[1.28, 1.33]$	$[10^{-33}, 10^{-30}]$
Chicago	$[1.88, 1.96]$	$[10^{-52}, 10^{-48}]$
Dubai	$[2.01, 2.14]$	$[10^{-16}, 10^{-14}]$
Paris	$[1.77, 1.84]$	$[10^{-24}, 10^{-23}]$
Portland	$[0.93, 0.98]$	$[10^{-22}, 10^{-21}]$
Rome	$[2.39, 2.48]$	$[10^{-19}, 10^{-16}]$

Figure 1

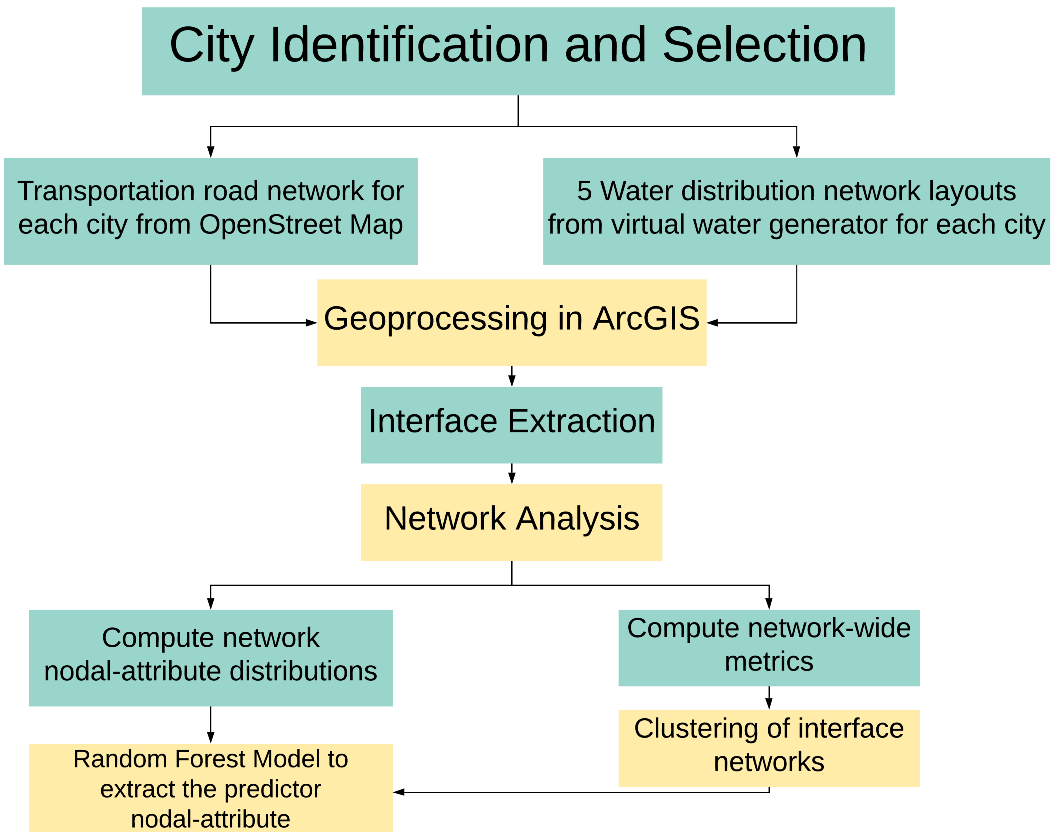


FIG. 1: Flowchart of study methodology starting with the city selection and ending with the extraction of a predictor nodal attribute

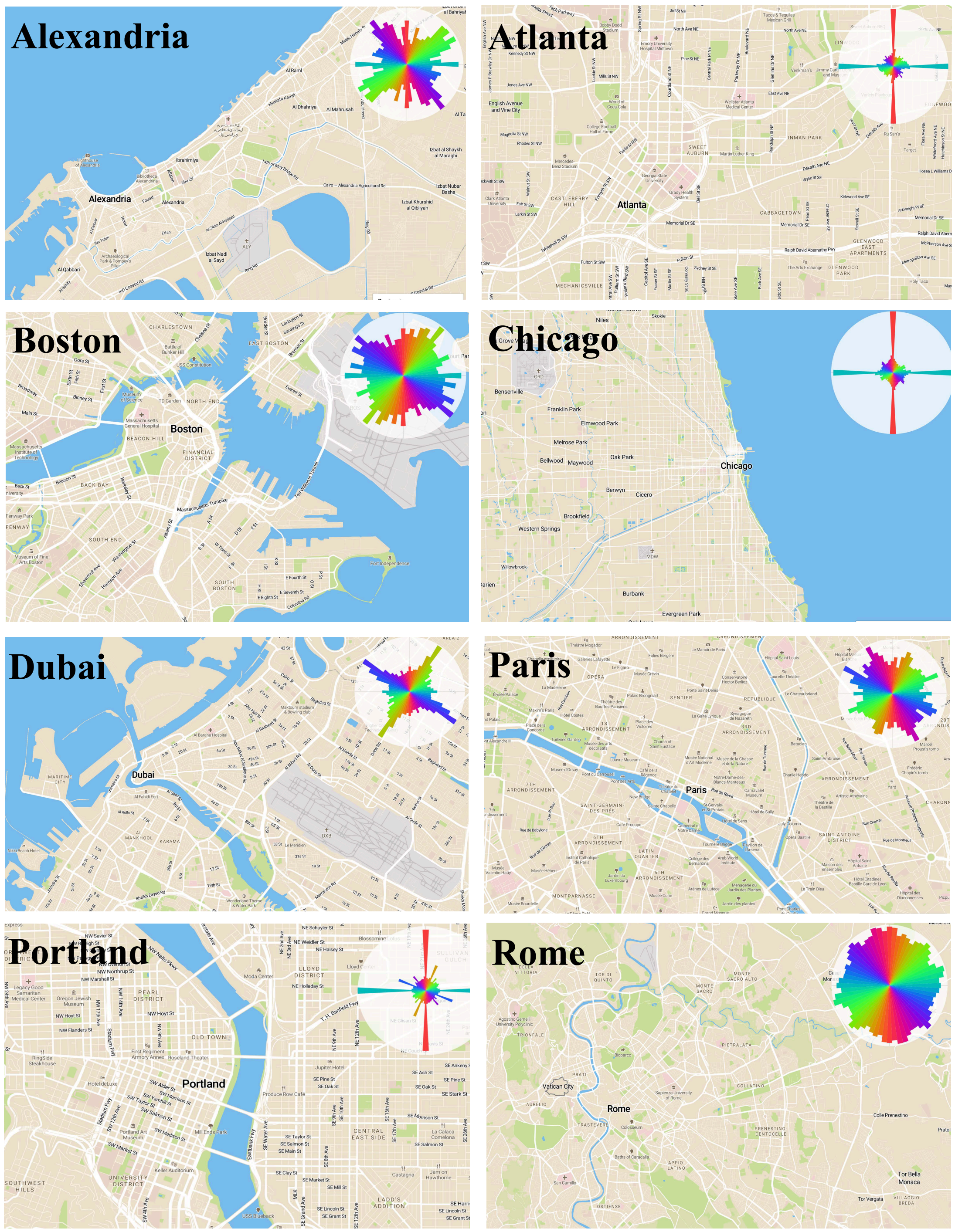


FIG. 2: Maps of the studied cities, with inlaid polar histograms representing the proportions of streets in different directions for each city. Portland is the most grid-like (polar histograms showing the majority of roads are along four directions). Rome is the most circularly developed city (polar historgram showing roads in about equal proportions lie in every direction).

(a)

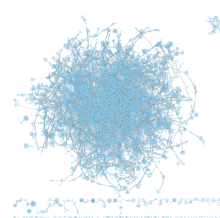
(b)

(c)

(d)

	A	B	C	D
1	Pipes	Roads		
2	24150088	139703		
3	142380856	173305		
4	125781188	176645		
5	121911598	218225		
6	183799181	112493		
7	30112270	112500		
8	30112294	112501		
9	30112316	112502		
10	28324326	112503		
11	129310460	112506		

(e)



(f)

FIG. 3: (a) A Boston water network sample (b) Boston road network (c) Overlay of water and road Networks (d) Physical representation of interface (i.e. co-located pipes and roads) (e) The edgelist of pipes and roads that is imported into cytoscape (f) Interface network representation in cytoscape, giant connected component can be seen in the center (no longer spatially embedded), and much smaller disconnected components are also shown

FIG. 4: Pipe and road intersections: one sample map of each city's intersection of pipes and roads. The interface networks are constructed from these intersections, where each line or point on the map represents a connection between a pipe and road.

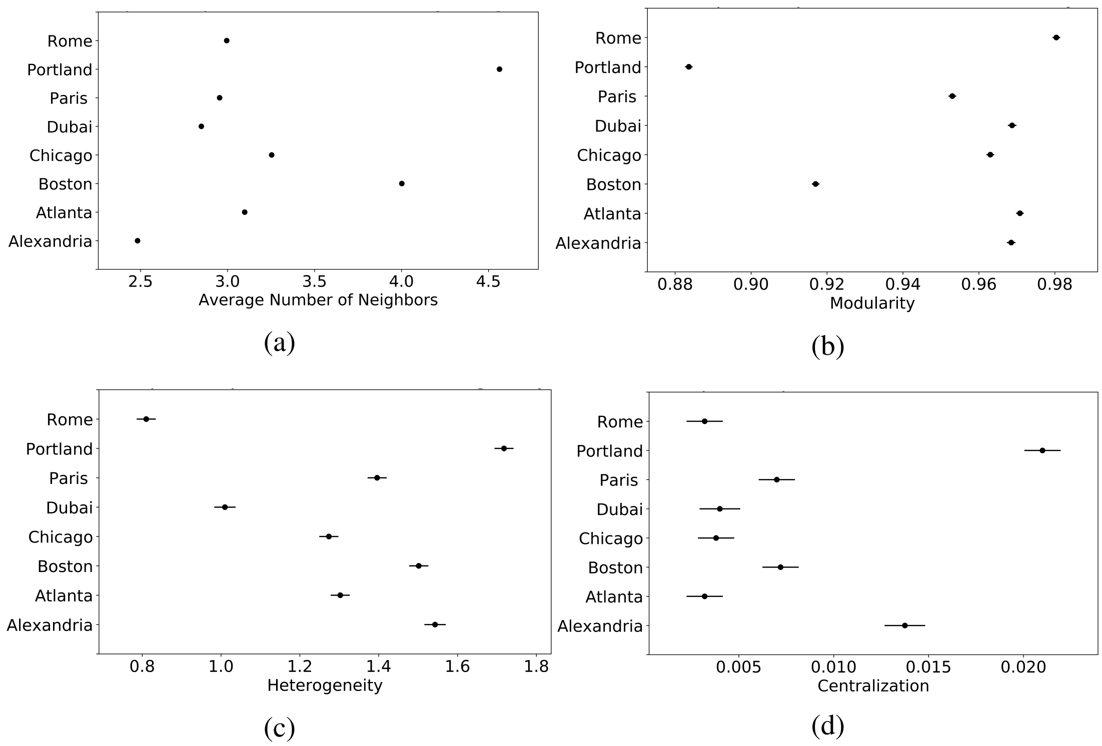
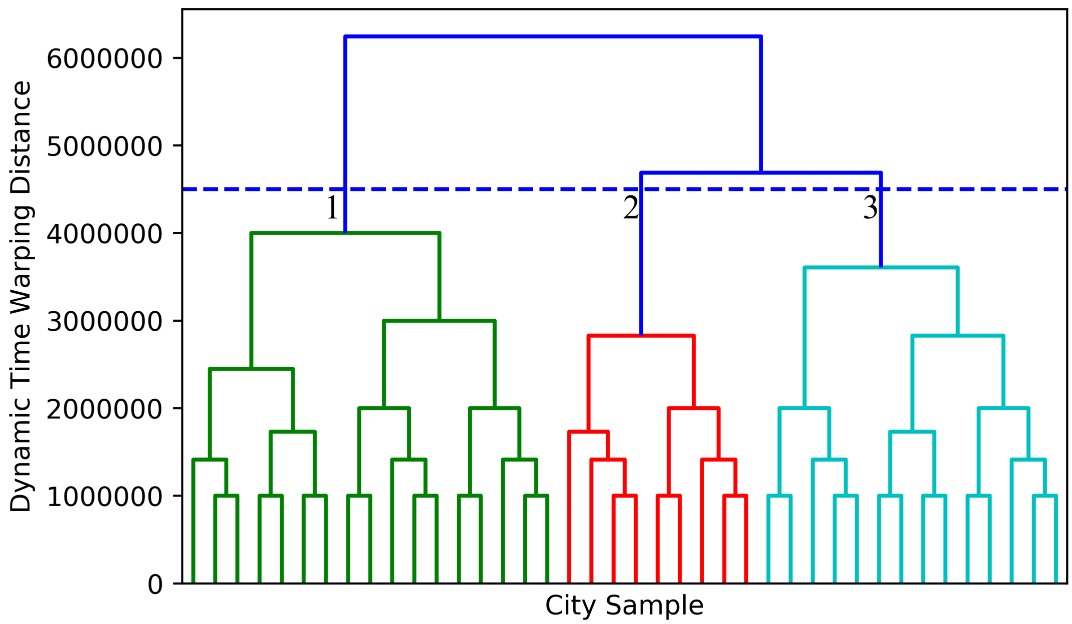


FIG. 5: Plots of the multiple comparison of means for the network-wide connectivity metrics: the x-axes are values of the different connectivity metrics (a) average number of neighbors (b) modularity (c) heterogeneity (d) centralization; the y-axes contain the different cities; each horizontal bar is summarizing the confidence intervals for the 5 samples of a given city; the plots show results from the multiple comparison of means from the post-hoc analysis for each connectivity measure; if the horizontal bars (of different city samples) in a plot overlap, the cities' interface network samples are not statistically different, but if they are not overlapping at all, then they are significantly different with p-value ≤ 0.05.



1	Chicago, Dubai, Rome, Alexandria
2	Alexandria, Atlanta, Paris
3	Paris, Boston, Portland

FIG. 6: Dendrogram of interface network connectivity: Along the x-axis are different city samples; The y-axis represents the distance between samples and their clusters, the lower the distance of the horizontal line connecting two samples or clusters, the more similar they are to each other; the further the distance at which two clusters are connected with a horizontal line, the more dissimilar they are. The dashed line intersecting the dendrogram is the location of the "cut" based on the two criteria for identifying clusters. As a result, three clusters are identified (1, 2, and 3), highlighted in different color.

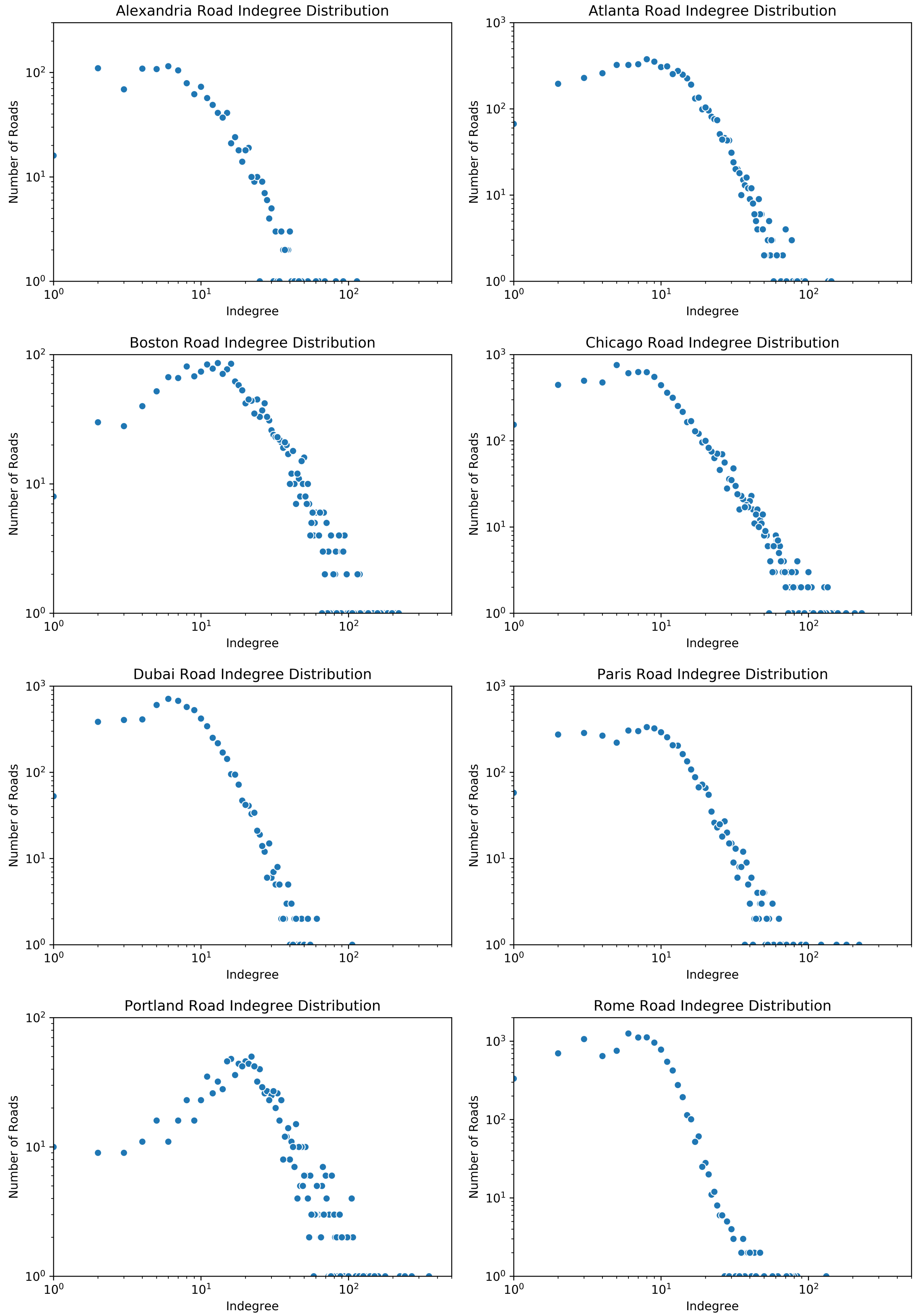


FIG. 7: Indegree distribution of roads in the interface networks for the studied cities. Axes are plotted on a log-log scale. The x-axis of these plots represents the number of pipes that intersect a road, this is the "indegree" of a given road. The y-axis represents the frequency, or number of roads, for which a given indegree occurs. Significant power-law tail is evident for all distributions at the same indegree value of 10, because of the linearity on the log-log scale.

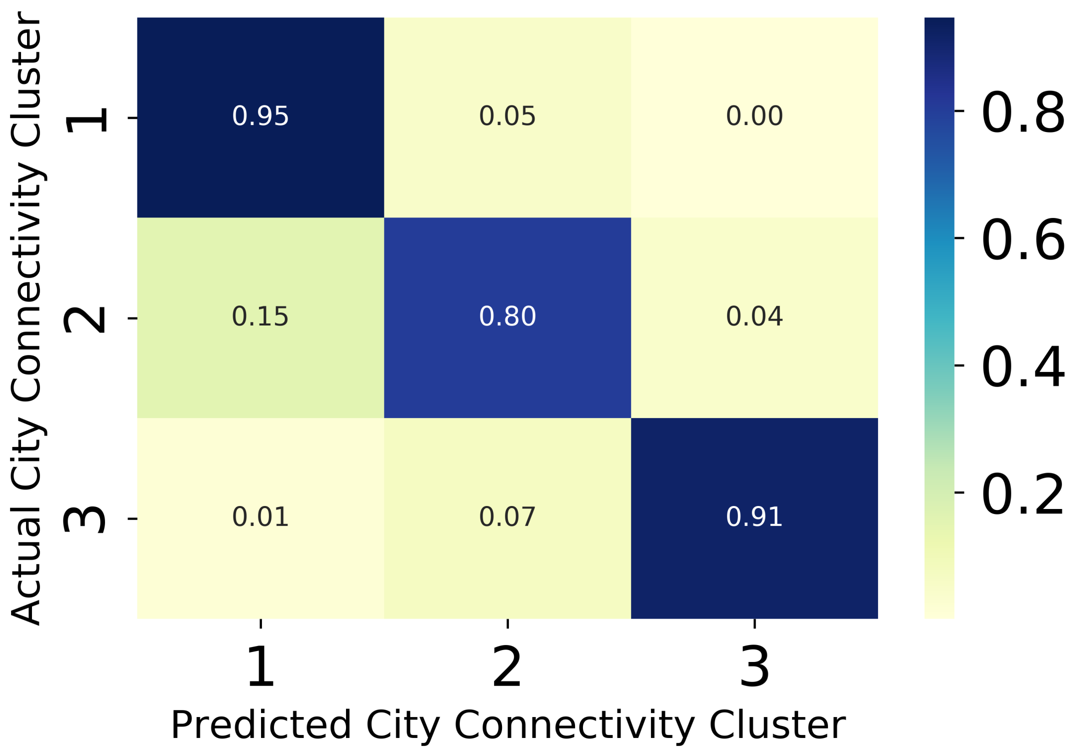


FIG. 8: Random forest model confusion matrix: The x-axis contains the city connectivity clusters predicted by the model; the y-axis contains the actual city connectivity clusters. The matrix is read by row, where values are the normalized number of samples predicted for each cluster. The values along the diagonal represent accuracies because the actual and predicted clusters are the same.

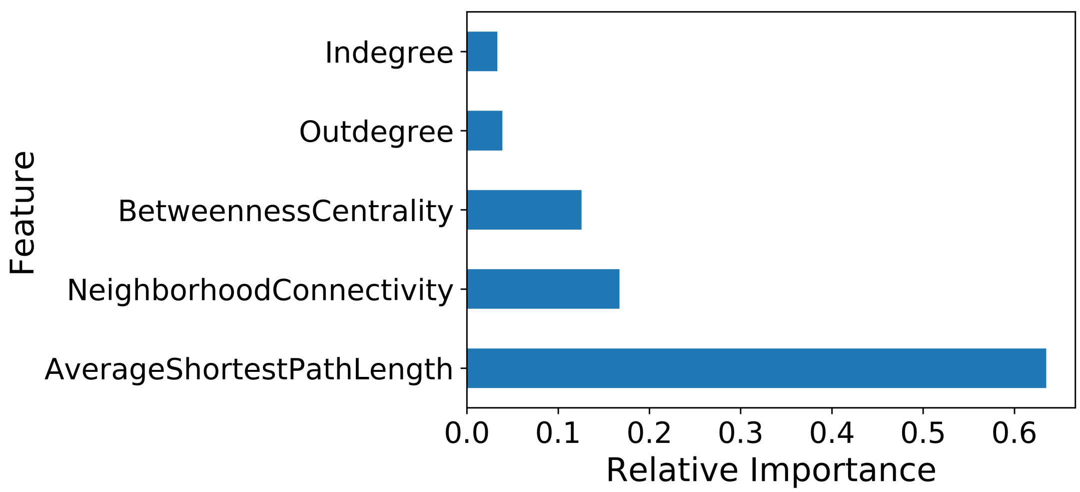


FIG. 9: Relative importance of nodal attributes to the prediction of the network-wide connectivity cluster by the random forest model: Average shortest path length is the most important nodal-attribute for the model in distinguishing between clusters.