



Conference Paper

Data Processing on Large Interdependent Networks: An Application for Infrastructure Preparedness, and Restoration

Aniela Garay-Sianca¹, Sarah G. Nurre¹, Carlos L. Castro Salas², and Humberto R. Alvarez A.³

¹University of Arkansas, Fayetteville, Arkansas, The United States

²Intel, Santa Clara, California, The United States

³Technological University of Panama, Panama City, Panama

Abstract

This paper presents a method for validating and transforming data for use in interdependent infrastructure network analysis. Critical infrastructure are interdependent on each other for delivery of services and execution of restoration activities. These interdependencies make infrastructure systems vulnerable to extreme events and highlights the needs for preparedness and response plans. Optimization models have been used to create effective plans using interdependent infrastructure networks. These models require accurate input data. However, many data sources have inconsistencies or errors which inhibit the ability to use such optimization models. This work identifies common errors in input network data and provides a method for processing and correcting these errors. We demonstrate the effectiveness of this method on data representing the transportation network in Juan Diaz town, in Panama.

Keywords: Data processing, network, infrastructure, interdependence.

Resumen

Este artículo presenta un método para validación y transformación de datos para su uso en análisis de redes de infraestructura interdependientes. Infraestructuras críticas presentan interdependencia entre ellas en las entregas de servicios y ejecución de actividades de restauración. Estas interdependencias hacen a los sistemas de infraestructuras vulnerables a eventos extremos, y resalta la importancia de los planes efectivos de preparación y respuesta. Los modelos de optimización son utilizados para crear planes efectivos usando las redes de infraestructuras interdependientes. Estos modelos requieren datos de entrada precisos. Sin embargo, muchas fuentes de datos tienen inconsistencias o errores los cuales dificultan su uso en dichos modelos de optimización. Este trabajo identifica problemas comunes en los datos de entrada de

Corresponding Author:
Aniela Garay Sianca
aniela.garay@gmail.com

Received: 15 November 2017
Accepted: 5 January 2018
Published: 4 February 2018

Publishing services provided
by Knowledge E

© Aniela Garay-Sianca
et al. This article is distributed
under the terms of the
[Creative Commons Attribution
License](#), which permits
unrestricted use and
redistribution provided that
the original author and source
are credited.

Selection and Peer-review
under the responsibility of the
ESTEC Conference Committee.



la red y provee un metodo para procesamiento y correction de errores. Nosotros demostramos la efectividad de este metodo en datos que representan la red de transporte en el pueblo de Juan Diaz en Panama.

Palabras claves: Procesamiento de datos, redes, infraestructuras, interdependencia.

1. Introduction

The frequency and severity of extreme events, either natural or manmade, have increased over the past years. Between 1974 to 2003, a total of 6,367 disasters occurred impacting 5.1 million people in the world. The worldwide economic damaged reported for these natural disasters is 1.38 trillion USD (Guha-Sapir et al., 2004). In Panama, 42 events have occurred since 2000 which is an increase in frequency. In the last 10 years (2007-2016), 172,097 people in Panama have been impacted by a disaster causing 210,000,000 USD of damage. This is a dramatic increase compared to the preceding 10 years (1995-2006) in which 82,514 people were impacted resulting in 15,850,000 USD of damage (EM-DAT, 2017). Thus, we need to focus on and be prepared for the increase in frequency and severity of disastrous extreme events.

Critical infrastructure systems are vulnerable to damages and loss of services as a consequence of extreme events. Infrastructure managers need effective disaster preparedness and response plans to increase the resilience of the systems. Resilience can be increased by (i) reducing the impact incurred as a result of a disaster and (ii) more rapidly restoring disrupted services. In addition, infrastructure systems are highly interdependent, thereby making it difficult to determine the best system-wide interdependent disaster preparedness and restoration plans. Interdependent infrastructures also need to coordinate their efforts to enable the rapid restoration of infrastructure services. Network optimization models are a promising approach for creating effective interdependent network preparedness and restoration plans. However, the effectiveness of these types of models rely on accurate data.

These models need data that represent infrastructures as interdependent networks with nodes, arcs, supply, demands, and other critical pieces of information. However, this is often not appropriate for immediate input into interdependent network models. Oftentimes, the data was collected for an alternative purpose (e.g., visualization), or the data has human or systematic errors. These types of errors hinder the ability of the optimization models to output effective and accurate restoration and preparedness

plans. Thus, inaccurate or incomplete data will limit the scope and power of interdependent optimization models.

In this paper, a methodology to create a validated and verified infrastructure network from inaccurate input data is proposed. Specifically, we use the transportation network in Juan Diaz, Panama as a case study and apply the data processing method to remove issues, such as connectivity. As a result of our process, we produce a validated transportation network which is appropriate for future use in interdependent network models. This paper proceeds as follows. In Section 2, we review the related background literature. In Section 3, we outline the proposed data processing method. Lastly, in Section 4, we summarize our conclusions and avenues for future work.

2. Literature review

Critical Infrastructure systems are vital to the function of society. The U.S. department of homeland security has identified 16 critical infrastructures including energy, communications, and transportation (U.S. Department of Homeland Security, 2013). These infrastructure systems are interdependent because the services of one system is required for the operation and/or restoration of another system. Rinaldi et al., 2001 provide a commonly used classification of infrastructure interdependence. This classification includes four main types of interdependence: physical, cyber, geographical, and logical. Another classification is presented by Chang et al. 2007 denoted infrastructure failures interdependencies (IFIs). IFIs occur during disasters when a failure of one infrastructure causes the failure of a second infrastructure system. According to Chang et al. 2007, there are five types of IFIs: cascading, escalating, restoration, compound damage propagation, and substitutive. In 2016, Sharkey et al. expanded upon these initial classifications to include the new type of restoration interdependence. A restoration interdependency occurs when the restoration tasks of one infrastructure contribute to or affects the restoration efforts of other infrastructure.

Many critical infrastructure systems can be modeled using a network. A network $G = (N, A)$ has a set of nodes N , and a set of arcs A . Nodes are points i, j on the network and arcs are line segments (i, j) linking nodes (Deo, 1974). When modeling infrastructures, a node could represent a roadway intersection and an arc could represent a road segment connecting intersections. The degree of a node is the number of incident arcs. The out-degree $d^+(i)$ and in-degree $d^-(i)$ are the number of arcs leaving and entering node i , respectively. A network is classified as directed or undirected. Directed networks include directed arcs wherein an arc (i, j) can only be traversed in one direction from

i to j . Undirected networks are comprised of only undirected arcs wherein arc (i, j) may be traversed from i to j or j to i . Often, infrastructure networks are modeled using directed networks. In contrast to graphs, networks have numerical values associated with nodes and arcs wherein the values could represent costs, capacities, and/or supplies and demands. Flow representing a particular commodity can traverse through a network originating at a supply node, passing through transshipment nodes, and ending at a demand node. A Network $G = (N, A)$ is connected when any node i can be reached by any other node j or equivalently between every i and j in N , there is a least one path (Deo, 1974). A path is a sequence of nodes and arcs without repeated nodes (Ahuja et al., 1993).

Problems on a network can be identified by applying existing algorithms. A Search algorithm can help to recognize the structural information of a network (Cormen et al., 2009). One way to execute a search algorithm is via breath-first. The Breath-First-Search (BFS) algorithm starts from a source node and iteratively adds nodes that can be reached from this source node to a list. The output of this algorithm is a set of nodes which are reachable from the starting source node. We will explain further how we adapted this algorithm for use in data-processing of infrastructure networks.

Researchers have combined these two areas of interdependent infrastructures and network optimization to model preparedness and restoration models for pre- and post-disaster analysis. (Lee et al., 2007, Nagurney and Qiang, 2007, Matisziw et al., 2008, Cavdaroglu et al., 2013, Sharkey et al., 2015). An underlying assumption of all of these models is that the input network data representing critical infrastructures is accurate and capable for performing network analysis. Oftentimes, the input data is insufficient to satisfy this assumption. Thus, we describe the methods utilized to analyze and correct input data so that it satisfies the criteria necessary for use in optimization models. We demonstrate these methods on a transportation network in Panama. While the methods worked for our input data set, additional analysis might be necessary for other types of input data. We present this method to summarize how existing network algorithms can be leveraged to correct errors in existing infrastructure network data.

3. Data Processing Method

The transportation network from the town Juan Diaz in Panama will be analyzed through the development of this section. In this section, we describe the method used to remove inconsistencies and prepare the input data for suitable use in interdependent network analysis. Figure 1 shows the raw data obtained from the area of

study: Juan Diaz, Panama (Contraloría General de la Republica de Panamá, 2010). The transportation system has a set of nodes and a set of arcs, where each node has a corresponding pair of coordinate points. Spatial data is used for the representation of this network. Working with spatial data brings some benefits, such as real location on earth, and performance of spatially analysis (ESRI, 2017).

In order to work with this spatial data for interdependent network models, we execute the following preliminary steps:

1. Assign a unique ID to each node. Working with spatial data, nodes have coordinates points, but using a unique ID for each node will help manage the data.
2. Remove any self-loops. A self-loop is an arc (i, i) for any node i .
3. Remove the origin point. In some datasets, the origin point (all 0's coordinate point) is used as reference point on the data and must be deleted. (ESRI, 2017).
4. Assign and verify the node type. The node type can be achieved by calculating the in-degree and out-degree of each node. Supply nodes must have a non-zero out-degree, demand nodes must have a non-zero in-degree, and transshipment nodes must have a non-zero out-degree and in-degree.

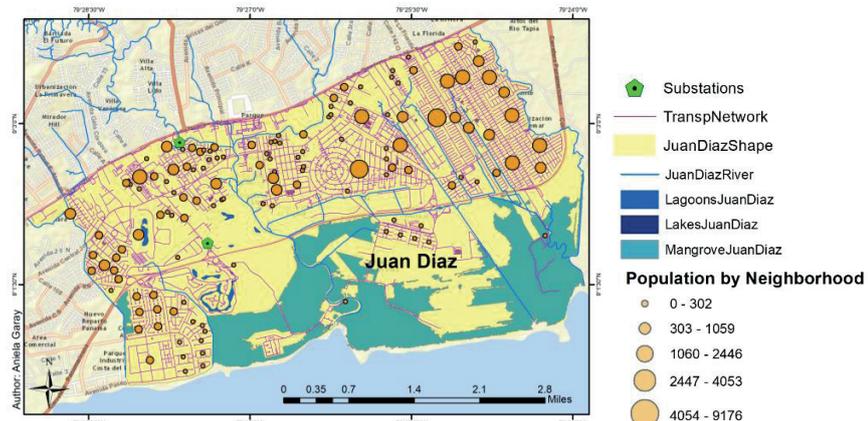


Figura 1: Map of the transportation network and census data for Juan Diaz, Panama City, Panama.

In the next subsection, we describe the detailed data processing conducted by using and adapting network algorithms for this spatial dataset.

3.1. Check for connectivity

In order to implement an interdependent network model, the input data must represent a connected network. Often the spatial data can represent unconnected networks.

In this section, we identify a set of problems found and then discuss the algorithms adapted to fix these problems.

Problem 1: Isolated sub-networks due to lack of existing edges. In some cases, a sub-region of a specific geographic area is required (e.g., towns, counties, or states). Performing a cut of the selected area may cause undesired disconnections in the network. These can be seen in Figure 2, where the red circle on Figure 2(a) highlights a bad cut in the area of study. The purple lines are arcs on the transportation network. As can be seen, a large arc is missing inside of the red circle. This situation generates several disconnections. The pink color in Figure 2(b) represents the cut area. The blue lines are not continuous due to the cut. This discontinuity isolates sections of the network, generating subgraphs or isolated arcs. Connectivity analysis could help to detect if such issues are present in the network.

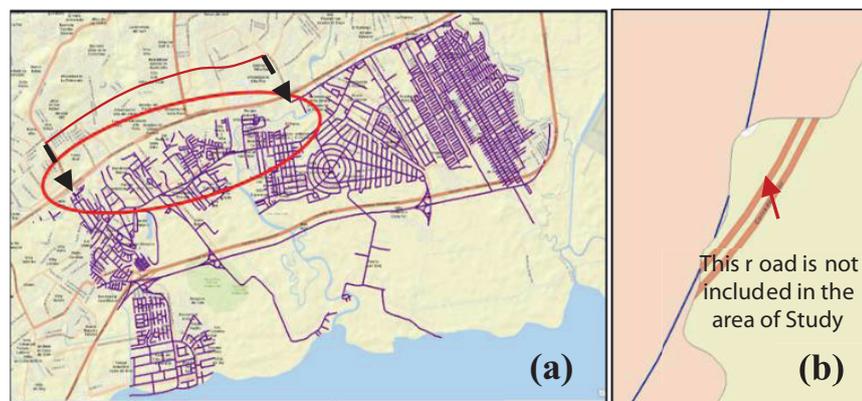


Figure 2: Section of the transportation network of Juan Diaz area: (a) Example of missing arc due to a bad cut on the network; (b) Close up on a section of the network that shows discontinuity.

Problem 2: Arcs not connected due to missing nodes. Spatial data collections are often created by using geographic information systems (GIS), such as Arcgis and AutoCad map 3D. For these cases, the data could be generated by drawing nodes and arcs on the maps within the GIS. The drawing process is often manual and therefore, is prone to human error. One such error is not precisely connecting arcs and nodes. Figure 3(a) shows a section of the network that appears connected. However, upon closer inspection of the same section, we identify a disconnection (Figure 3b).

Problem 3: Erroneous nodes and arcs on the network. We detected erroneous nodes and arcs that did not belong to any road or intersection. In Figure 3(c), a small node was found underneath an arc. This node was completely disconnected from the network which is only visually apparent after the arc on top is removed. We found these nodes

and arcs using algorithms that search for arcs and nodes within a customizable distance (i.e., below 4 meters for our study).

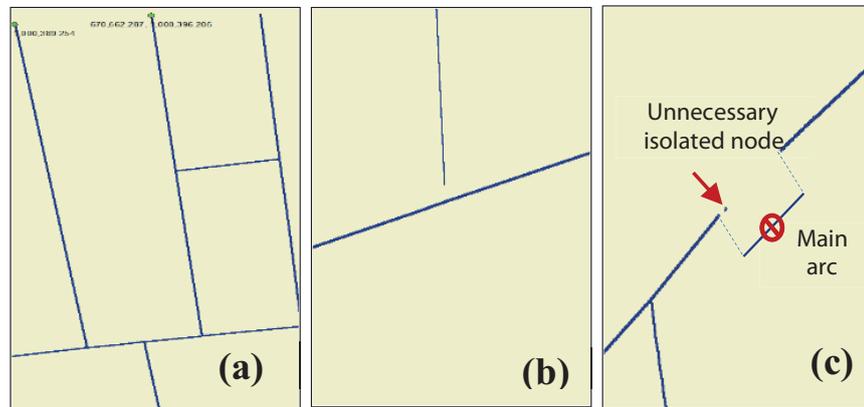


Figure 3: Network errors: (a) Network that appears connected; (b) Close up of a disconnected intersection; (c) Example of unnecessary node underneath an arc.

Problem 4: Overlapping lines with no intersections. In Figure 4(a), a main street (blue line) is drawn on top of several arcs without any intersections along the line. For visualization purposes, this type of drawing is acceptable, but the resulting data set is not appropriate for network modeling as many connections are lost. Overlapping arcs with no connection between them should be fixed if there is no bridge or tunnel present.

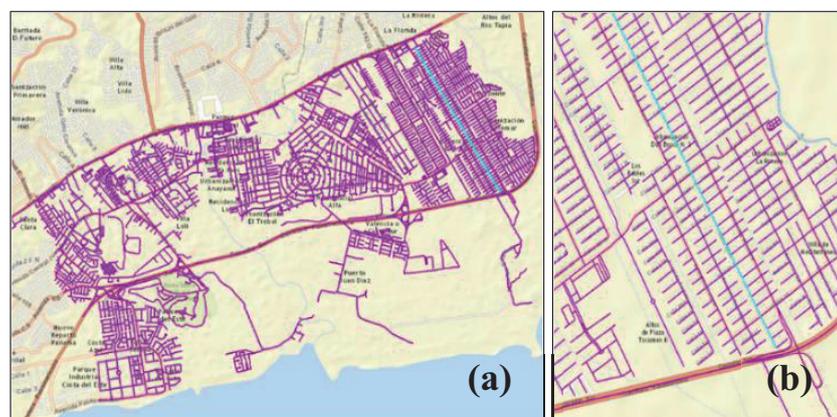


Figure 4: (a) Transportation Network in Juan Diaz, Panama; (b) Close-up on a section of the network where the blue line represents a single arc without any proper intersections.

3.2. Proposed Error detection methods

As many of these errors are not detectable in the visualization, we proceed by presenting algorithms to detect and fix the errors. These algorithms are important for ensuring the accuracy of input data for modeling.

Algorithm 1 (Solving Problems 1 and 4). To assess the connectivity of the transportation network and to address problems 1 and 4, we implement a modified breath first search (BFS) algorithm. The BFS was modified to iteratively identify the number of connected subnetworks. The pseudocode of this algorithm is shown in Figure 5. We used this algorithm on the Juan Diaz transportation network and identified 43 subnetworks. We reviewed each subnetwork and fixed the disconnections.

Algorithm 2 (Solving Problems 2 and 3). Even after a network is connected, the network may exhibit problems 2 and 3. To detect these problems, we calculate the distance between each node and arc (Anton, 2010). Using this distance, we identify the set of nodes within a distance d from each arc. All nodes within each set are possible intersections with the corresponding arc. Based on trial and error and known disconnections, we used $d = 4$. The output of this algorithm is a list of candidate node-arc connections sorted based on distance from smallest to largest. We obtained 250 candidates that required further evaluation. Using visual inspection, around 80% were identified as problem 2 and 3 and were fixed. The randomness and difficulty in visually detecting these errors demonstrate the importance of running this analysis.

4. Conclusion and future works

In this paper, we presented a methodology used to process data representing an infrastructure network. Using the transportation network from Juan Diaz, Panama, we identified lack of connectivity and other errors. We modified the directed breadth-first-search algorithm to identify and fix the main connectivity errors. Overall, the transportation network started as 43 different subnetworks and was fixed to represent one connected subnetwork. Next, we used a distance calculation to identify extraneous nodes and arcs in the network. We detected and fixed around 250 such errors. We verified these errors using visual inspection. Thus, we demonstrate that we can make small modifications to existing network algorithms to detect and fix errors within infrastructure network in order to make the data suitable for interdependent preparedness and response network analysis. The methods outlined can serve as the

```
// Part 1
FUNCTION BSF_Directed (Node list N, Source Node S, Num_Network NN, Arc list A)
  STORE S in Nodes_to_Search and SET S visited flag to 1
  WHILE Nodes_to_Search is not empty
    SET Node_to_Review to the first node in Nodes_to_Search and remove it from list
    SET Connected_Nodes to all nodes connected to Node_to_Review using "In-degree" d and "Out-degree" d* from Arc list A

    FOR each Node in the Connected_Nodes list
      IF Node visited flag is 0 THEN
        STORE Node as last element in Nodes_to_Search list
        SET Node visited flag to Num_Network NN
      END IF
    END FOR
  END WHILE
END FUNCTION BSF_Directed

//Part 2
FUNCTION Connectivity (Node List N, Arc list A)
  SET Num_Network NN to 1
  FOR each Node in the Node List N
    IF Node visited flag is 0 THEN
      SET Source Node S equal to Node
      // Breath First Search will mark all nodes connected to S with the Num_Network NN
      CALL BSF_Directed with (Node list N, Source node S, Num_Network NN, Arc list A)
    END IF
    INCREASE Num_Network NN by 1
  END FOR
  RETURN Num_Network NN reduced by 1 //If there is only one Graph and no subgraph it will
return 1
END FUNCTION Connectivity
```

Figura 5: Pseudocode of Algorithm 1: Modified Breadth-First-Search.

starting point for future verification of other interdependent networks, such as power flow networks.

5. Acknowledgment

The authors thank the collaboration and support from Panama Municipality, and the Electrical, Mechanical and Industrial Research and Innovation Center(CINEMI) of the University Technology of Panama.

Referencias

- [1] Ahuja, R. K., Magnanti, T. L., & Orlin, J. B. (1993). *Network flows: theory, algorithms, and applications*.
- [2] Anton, H. (2010). *Elementary linear algebra*. John Wiley & Sons.
- [3] Cavdaroglu, B., Hammel, E., Mitchell, J. E., Sharkey, T. C., & Wallace, W. A. (2013). Integrating restoration and scheduling decisions for disrupted interdependent infrastructure systems. *Annals of Operations Research*, 203(1): 279-294.
- [4] Chang, S. E., McDaniels, T. L., Mikawoz, J., & Peterson, K. (2007). Infrastructure failure interdependencies in extreme events: Power outage consequences in the 1998 Ice

- Storm. *Natural Hazards*, 41(2), 337-358.
- [5] Contraloría General de la Republica de Panamá (2010). "Instituto Nacional de Estadística y Censo.", XI Censo Nacional de Población y VII de Vivienda 2010: Lugares Poblados de la República: 2010. <http://www.contraloria.gob.pa/INEC/Publicaciones/Default.aspx>. Accessed February 15, 2017.
- [6] Cormen, T. H., Leiserson, C. E., Rivest, R. L., & Stein, C. (2009). *Introduction to Algorithms*. Third Edition Cambridge: MIT press.
- [7] Deo, N. (1974). *Graph theory with applications to engineering and computer science*. Englewood Cliffs, N.J: Prentice-Hall.
- [8] Environmental Systems Research Institute (ESRI). *Coordinate systems, projections, and transformations*, ArcGIS Pro Online help. Retrieved from <http://pro.arcgis.com/en/pro-app/help/mapping/properties/-coordinate-systems-and-projections.htm>. Accessed June 10, 2017
- [9] EM-DAT: The Emergency Events Database - Université catholique de Louvain (UCL) - Centre for Research on the Epidemiology of Disasters (CRED), D. Guha-Sapir, www.emdat.be, Brussels, Belgium. Accessed June 14, 2017.
- [10] Guha-Sapir, D., Hargitt, D., & Hoyois, P. (2004). *Thirty years of natural disasters 1974-2003: The numbers*. Centre for Research on the Epidemiology of Disasters: Presses Universitaires de Louvain.
- [11] Lee, E. E., Mitchell, J. E., & Wallace, W. A. (2007). Restoration of services in interdependent infrastructure systems: A network flows approach. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 37(6), 1303-1317.
- [12] Matisziw, T. C., Murray, A. T., & Grubestic, T. H. (2009). Exploring the vulnerability of network infrastructure to disruption. *The Annals of Regional Science*, 43(2), 307-321.
- [13] Nagurney, A., & Qiang, Q. (2008). A network efficiency measure with application to critical infrastructure networks. *Journal of Global Optimization*, 40(1), 261-275.
- [14] Rinaldi, S. M., Peerenboom, J. P., & Kelly, T. K. (2001). Identifying, understanding, and analyzing critical infrastructure interdependencies. *IEEE Control Systems*, 21(6), 11-25.
- [15] Sharkey, T. C., Cavdaroglu, B., Nguyen, H., Holman, J., Mitchell, J. E., & Wallace, W. A. (2015). Interdependent Network Restoration: Modeling Restoration Interdependencies and Evaluating the Value of Information-Sharing. *European Journal of Operational Research*, 244(1), pages 309-321.
- [16] Sharkey, T. C., Nurre, S. G., Nguyen, H., Chow, J. H., Mitchell, J. E., & Wallace, W. A. (2016). Identification and classification of restoration interdependencies in the wake of Hurricane Sandy. *Journal of Infrastructure Systems*, 22(1), 04015007.

- [17] U.S. Department of Homeland Security. (2013). National Infrastructure Protection Plan 2013: Partnering for Critical Infrastructure Security and Resilience. Retrieved from <https://www.dhs.gov/sites/default/files-/publications/national-infrastructure-protection-plan-2013-508.pdf>

Authorization and Disclaimer

Authors authorize ESTEC to publish the paper in the conference proceedings. Neither ESTEC nor the editors are responsible either for the content or for the implications of what is expressed in the paper.