Causal Inference Methods and their Challenges: The Case of 311 Data

Farzana Beente Yusuf
fyusu003@fiu.edu
Florida International University
Miami, Florida, USA

Shaoming Cheng
scheng@fiu.edu
Florida International University
Miami, Florida, USA

Sukumar Ganapati
ganapati@fiu.edu
Florida International University
Miami, Florida, USA

Giri Narasimhan
giri@fiu.edu
Florida International University
Miami, Florida, USA

ABSTRACT

The main purpose of this paper is to illustrate the application of causal inference method to administrative data and the challenges of such application. We illustrate by applying Bayesian networks method to 311 data from Miami-Dade County, Florida (USA). The 311 centers provide non-emergency services to residents. The 311 data are large and granular. We aim to explore the equity issues and biases that might exist in this particular type of service requests. As a case study, the relationship between population characteristics (independent variables) and request volume and completion time (dependent variables) is examined to identify the disparities, if any, from the observational data. The empirical analysis shows that there are no biases in services provided to any specific demographic, socioeconomic, or geographical groups. However, the administrative data do have various challenges for inferring causality due to missing or impure data, inadequacy, and latent confounders. The precautions of applying causal techniques to analyzing administrative data like 311 are discussed.

CCS CONCEPTS

• **Computing methodologies** → **Bayesian network models**.

KEYWORDS

311 customer service, Causal networks, Bayesian network

ACM Reference Format:
Farzana Beente Yusuf, Shaoming Cheng, Sukumar Ganapati, and Giri Narasimhan. 2021. Causal Inference Methods and their Challenges: The Case of 311 Data. In *DG.O2021: The 22nd Annual International Conference on Digital Government Research (DG.O'21), June 9–11, 2021, Omaha, NE, USA.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3463677.3463717

DG.O'21, June 9–11, 2021, Omaha, NE, USA

ACM ISBN 978-1-4503-8492-6/21/06…$15.00
https://doi.org/10.1145/3463677.3463717

1 INTRODUCTION

Innovations in e-governance have enabled citizens and government agencies to join forces in improving the quality of services and citizen satisfaction. The 311 contact centers are examples of such innovative systems. The 311 centers are organizations within local governments to field non-emergency service requests. They were enabled by a 1997 Federal Communications Commission policy to reduce the volume of non-emergency calls to 911 centers. The 311 centers have become a hub for local services, fielding both information and service requests from residents. A citizen can report issue, complaints, or requests for services related to local government. Examples of such service requests include: tree trimming request on a blocked sidewalk, broken stoplight, trash pickup requests, and many more depending on the locality. The requests can be made via mobile apps, social media, online chats, emails and text messages. The service requests are routed to the relevant department through a 311 Customer Relationship Management (311/CRM) system. The data recorded through the 311 CRM system are quite large and granular, which include details about each service request call's characteristics (e.g. location, time, etc.). As the 311 centers gained popularity over the years, more than 100 cities have implemented them ([42]). With open data movement, many cities have made their 311 data public (Open 311) in an effort toward greater transparency. The 311 data are thus big administrative data sets from local governments. Analysis of the 311 data is useful for policy makers and analysts to gain insights into the nature and demand for services in the local governments.

The purpose of this paper is to apply causal inference methods to the 311 administrative data to gain further policy insights into how local governments respond to service requests, how engaging a particular community is, and to examine if government service provision is uniform across different demographic, socio-economic, and geographical characteristics. With the ever-growing publicly available administrative data like the 311, it is now feasible to seek answers from the data rather than relying on citizen surveys. The availability of public data has also enabled application of sophisticated models to better understand the causal factors for designing better policies that are effective for a community. In order to make new policy decisions, or figure out the effectiveness of existing policies, it is important to understand the underlying causal community characteristics. Extracting the key characteristics from administrative data provides useful information to policy makers.

The ability to infer causal relationships between variables in a system is an important scientific endeavor. Whereas traditional regression models establish correlations, recent advances in machine learning approaches have enabled us to develop causal explanatory models. The *causal Bayesian network* (BN) is one such powerful model, which is easy to understand and reasonably interpretable. BNs are probabilistic frameworks, which are useful for establishing generic causality and captures more complex, and often, more insightful relationships between variables than a traditional model. The most important characteristic of BNs is that they provide a way to distinguish between direct and indirect dependencies from the observations [41]. The theory of *Bayesian networks* (BNs) provides the foundation for us to explore such dependencies. Moreover, these models can simultaneously represent statistically significant knowledge (learned from data) and domain expertise, therefore being the intuitive choice for our causal analysis instead of regression analysis. BNs have been applied successfully in many different fields, such as, gene expressions analysis [15, 36, 49], medical services [4, 29], risk assessment and safety systems [5, 6, 25], epidemiology [16, 18, 31, 34], social sciences [20, 33], econometrics [26, 43].

We focus on applying the BN to 311 data from one local government: Miami-Dade County in the United States. The County was among the early implementers of a 311 call center in the country. The County's 311 system is operational for all of the unincorporated part of the county (i.e. those areas which have not been incorporated as a city) and most of the incorporated municipalities. For example, the City of Miami, which is located within the County, is also served by the 311 center. Our selection of the County to illustrate the causal model is appropriate since the 311 system is multi-jurisdictional and is among the large well established systems in the country. The 311 center has fielded over 200,000 calls every year in the last five years.

The rest of the paper is structured as follows. Section 2 provides a background of the paper, highlighting the recent related literature. Section 3 outlines the fundamental aspects of Bayesian Networks. Section 4 shows the application of BN to the Miami-Dade 311 center data. Section 5 highlights the challenges of applying the BN framework. The Section 6 concludes with the major lessons from the study.

2 BACKGROUND

Local governments (such as cities, counties, school districts etc.) provide direct public services. They are closest to the residents in the jurisdiction to field any inquiries and request for services. Traditionally, residents could request these services only by contacting the appropriate local government agency or department. For example, a resident would need to call the public works department directly to report a pothole. Often, residents would not know which agency should be contacted for obtaining a service. Consequently, the local government services would be available to residents who have the means and the contacts to local government agencies. Inequities in local government services (such as street maintenance, lighting, trash pickup, etc.) have been well documented. Wealthy areas get better services as they have better political contacts. Citizen engagement and satisfaction measures were lopsided substantially with the inequitable distribution of government services [45]. Studies show that the citizens' perceptions and satisfaction toward their government are correlated with accessibility of public services.

With the advent of 311 call centers, any person can call the local government for a service. The person does not need to know which agency fulfills a particular service, and how to contact the agency. The service request is routed at the back end of the 311 CRM system to the appropriate department or jurisdiction. The 311 centers thus arguably democratize the access to local government services, whereby anybody could request services directly without having the luxury of prior political contact. Residents could make the service requests through multiple methods–over the phone, online (through an app or website), or social media. Although the 311 expands residents' access to local government agencies through a one-stop method, residents would have to be actually aware of the 311 system's existence and make the call. Thus, the empirical question of who calls the 311 center and whether the calls are equitable across different demographic, socio-economic groups, and geographical areas remains an empirical question. On the flip side, another important question is that of how equitably the government agencies provide the services. Even though wealthy and low-income areas may make the service requests, there could be inequities in fulfilling the requests (and their efficiency in doing so) between wealthy and poorer areas for various reasons.

A key underlying assumption in the above empirical questions is that the demographic, socio-economic, and geographical factors drive the service requests and their fulfilment. That is, there is a causal assumption of the factors which drive the service requests and their performance of how efficiently the requests are complied with. The problem with such an assumption is that we do not know the direction of causality. Although not the classical chicken-egg problem of which came first, there are inherent problems with assuming the causal directions. We would also need to be beware of reverse causality. The causal assumptions are especially problematic for policy-making in social science since these assumptions drive how future investments are made and which groups and areas should be prioritized for obtaining public services.

The above discussion leads to two crucial research questions. The first question is: Are the service requests equitably distributed through demographic groups, socio-economic levels, and geographical areas? This is a question from the demand side, looking at who makes the 311 service requests. The second question is: Are the service requests equitably fulfilled across the groups? This is a question from the perspective of the local government agencies that oversee the service requests. Extant studies have investigated facets of these questions using traditional qualitative and quantitative methods [7–9, 12]. Some of the studies performed hypothesis testing using survey data [8, 12]. One of these papers examines these effects on a wide range of responses, from evacuation timing and emotional support to housing and job conditions and plans to return to pre-storm communities, using survey data from over 1200 Hurricane Katrina survivors. The findings show significant ethnic and socioeconomic disparities. In [7], the authors look at service requests made to the City of Boston over the course of a year (2010-2011) and use geospatial analysis and regression analysis to look at potential inequities in service requests based on race, education, and income. The results show that there is no concern that

311 systems (non-emergency call centers) would favor one ethnic group over another. They included race/ethnicity, median income, education, home ownership, and the population as independent variables. A 15-city review of 311 systems (non-emergency service requests made by city residents) found no systemic disparities in how cities respond that would suggest a bias toward minorities and lower-income residents including the independent variables like income, race/ethnicity, and education [9]. The results are not consistent in all of 15 cities, with some showing no bias and others showing it. Based on in-depth interviews with Philadelphia City government officials and managers responsible for creating and operating the 311 center, Nam [30] and Hartmann et all. [19], argue that the program is resulting in a more efficient, effective, transparent, and collaborative city government.

Open311 has allowed for new opportunities to gain insights from the observations quantitatively [27]. Open 311 is a standardized protocol which allows for commensurate measurement of 311 service requests across different cities. Predictive models have been applied to extract useful insights from 311 data [17, 23, 46–48] of different cities, i.e., New York City, Chicago, Philadelphia, etc. These models present analytical frameworks to study overall or particular requests to help better resource allocation and reduce the response time. These studies have shed light on the key features that affect the different types of requests. Different approaches have also been applied to improve service quality in terms of responsiveness at the time of disasters [28, 50]. Elliott et al.'s study of responses after Hurricane Katrina revealed strong biases in providing services by government[12]. Xu et al.'s study of 311 service requests after Hurricane Michael in the City of Tallahassee, Florida, also show similar biases. To date, however, we are not aware of any study that has applied causal learning algorithms to administrative data like 311. The application is important to understand the causal factors that can have substantive policy impacts on how resources should be allocated based on residents' service requests.

It is in the above context that we are taking advantage of recent advances in machine learning to apply causal learning algorithms to the 311 administrative data set. Since this is among the first such studies to apply the causal methods, this paper is exploratory in nature. Our aim is to examine the extent to which the causal methods are applicable to administrative data like 311 for inferring causal relationships in a substantive way. In this process, we also highlight the challenges that arose in the application of causal methods to the administrative data. As we explain in more detail below, we chose the Bayesian networks method to understand the causal relations.

3 CAUSAL METHOD: BAYESIAN NETWORK

Bayesian Networks (BNs), a class of *Probabilistic Graphical Models* (PGMs) [22, 32], can be represented as a *Directed Acyclic Graph* (DAG) along with a conditional probability table. The DAG is denoted by $G = (\mathbf{V}, A)$, where V is a set of nodes (or vertices) and A is a set of directed edges connecting one node to another. Each node in $\mathbf{V}$ represents a random variable from a set $\mathbf{X} = \{X_i, i = 1, \ldots, n\}$. A directed edge A between two nodes reflects a dependency between them. If (v_i, v_j) is a directed edge in G, then v_i is said to be a *parent* of v_j. The conditional probability table describes the *marginal* distribution of a node v_i given the joint distribution of its parents. Each random variable, X_i follows a probability distribution $P(X_i)$, which may be discrete or continuous. The Bayesian network describes the relationships between these distributions. The *joint probability distribution*, $P(\mathbf{X})$ is the product of all the *probability distributions* for each random variable, X_i. The local probability distributions, $P(X_i)$, satisfy the *Markov property* , which states that every random variable X_i is dependent only on its parents (i.e., set of vertices where there exists a directed edge from those to the node) [24]:

$$P(X) = \prod_{i=1}^{n} P(X_i | Parents(X_i)) \tag{1}$$

The above joint probability distribution can be simplified if the BN can be made sparser by eliminating edges. In particular, if two variables are independent or conditionally independent (conditioned on other variables) of each other, then the corresponding edge can be eliminated. Note that if all the variables are independent of each other, then the joint probability distribution is simply the product of the individual distributions, reflecting an empty BN. Also, note that BNs are acyclic since the structure would fail to generate a cause-effect relationship if the cycles are allowed in the network.

The task of fitting a BN is called a "model learning". It involves two steps [38, 39] as follows:

- Structure learning: learning the structure of the network from the data;
- Parameter learning: estimation of the local probability distribution implied by the structure learned.

Given a dataset D, if the parameters of the global distribution is denoted by θ, the model, denoted by M, learning can be defined as follows for the graph G:

$$\underbrace{P(M|D) = P(G, \theta|D)}_{\text{model learning}} = \underbrace{P(G|D)}_{\text{structure learning}} \cdot \underbrace{P(\theta|G, D)}_{\text{parameter learning}} \tag{2}$$

Here our focus is on *Structure learning* only. The objective is to find a network (BN) that will encode all the conditional dependencies from the data. If the edges represent relationships that are believed to be causal, we have a causal BN. Structure learning approaches can be grouped into three broad categories: constraint-based, score-based and hybrid. Constraint-based algorithms utilize the probabilistic relations defined by the Markov property of BN, based on the *Inductive Causation* (IC) [44] algorithm, which provides a theoretical framework to learn the BN using *conditional independence* (CI) test. The algorithm might not resolve all the directional dependencies from the data; hence, may provide a partially directed acyclic graph (PDAG) [21]. A constraint-based IC approach was proposed by Sprites et al. to learn the BN, [40]. This approach is better suited than score-based methods as it generates network with fewer false positives, resulting in a conservative structure in terms of number of edges; therefore being the intuitive choice for our causal analysis. *PC-Stable* is an improved order-independent version of IC algorithm [10, 11].

Next we briefly describe the key features of the PC-Stable algorithm. The first step in learning the structure using PC-Stable is to find the pairs of connected nodes in the network. This step starts from a complete undirected graph and eliminates edges using

conditional independence tests. This results in a skeleton structure. A 3-node DAG is the smallest structure that can be causally inferred. The basic idea is to use the *Conditional Independence* (CI) test to distinguish between the three possible 3-node DAG structures [44]. Also, when the acyclic assumption is relaxed, the structure would fail to generate a cause-effect relationship. The next step is to identify useful substructures, one of which is an important structure called a *v-structure*. Depending on the conditional dependencies among between three random variables X_i, X_j, X_k, the useful substructures can be categorized into three different types, as shown in Figure 1.

- *Causal chain*: This describes variables that affect each other sequentially. In a BN G, for any three nodes v_i, v_j, v_k, representing variables X_i, X_j, X_k, if there exist edges v_i, v_j and v_j, v_k, and the edges are oriented as follows: $v_i \rightarrow v_j \rightarrow v_k$, then it is called a *causal chain*. This represents the case when once X_j is known, X_i and X_k become independent of each other.
- *Common cause*: This represents the case where two variables are impacted by a third variable. The causal structure for this case is as follows: $v_i \leftarrow v_j \rightarrow v_k$. Here, as in the case of the causal chain, once X_j is known X_i and X_k become independent of each other.
- *Common effect*: If the edges are oriented as follows: $v_i \rightarrow v_j \leftarrow v_k$, then it is called a v-structure. The variable X_j is considered to be a "common effect". Here, X_i and X_k are independent of each other, but become dependent when conditioned on X_j. This property makes the v-structure distinguishable from the other two with the help of a simple CI test.

As mentioned above, causal chains and common causes are not distinguishable with CI tests. As a result, we start by identifying the v-structures and then orient the remaining edges of skeleton to make the network acyclic and consistent. Generally, all the edges in the graph are directed, but some algorithms leave some edges as undirected to reflect the uncertainty in determining the direction of the relationship.

Finally, the next step of Bayesian learning is parameter learning, where it does regression over the variables learned from the structure and can be used for prediction. Thus, the approach used in this paper does use regression, but only on specific subsets suggested by the structure learning step. Note that standard regression techniques do not have the ability to determine the subset of variables on which to apply regression. Our primary focus is to infer the causal structure from the observational data and find the subset of variables using Conditional Independence that affects the dependent variables from a set of independent variables.

4 CAUSAL INFERENCE

4.1 Dataset

4.1.1 311 dataset. For the analytics presented here, we downloaded the data from Miami-Dade County's open data portal [2]. The data have been made publicly available to promote quick access and transparency, thus enabling different case studies with the aim of understanding the community better. The dataset includes the service requests from 2013 to present made by the community member to the 311 center. The requests were made in one of many ways i.e., through a phone call, app, or website. The breakdown of the volumes of the different types of service requests, mainly dominated by trash pickup requests, is shown in Figure 2.

Each request record contains key information regarding the type of request, timestamp, and exact GPS location from where the requests were made. This data was combined with the TIGER/Line files and shapefiles from the U.S. Census Bureau for the county, thus associating each request with the the *geographic entity codes* (GEOIDs) at two different geographical granularity levels, i.e., block group and Census tract. The dataset has the Zip code level. We also us the timestamp to calculate the service completion time (number of days taken to complete servicing a request since its initiation). We exclude the observation for which either the request was never closed or the completion time was negative (suggesting a recording error on the timestamp). The services requested were divided into four broad types: requests, complaints, issues, and others, based on a keyword search on the issue type. We focused our analysis on the records labeled as "Requests". Records labeled as "complaints" or "issues" typically have longer completion times. Then, we aggregated the total number of requests and the average completion time for each geographic unit. There were 520 and 1, 595 entries in the dataset for the most recent year. Requests aggregated at the Zip Code level were excluded from further analysis since the number of different zip codes in Miami-Dade was too small for worthwhile analytics (79 entries for the latest year).

4.1.2 Census dataset. The 311 dataset does not contain the identity of the person who requested the services. As a result, we combined the geocoded data with the available demographic information for the geographical unit. The demographic information is collected from the 5-year estimates of the American Community Survey (ACS). We conducted analyses of the data at both the Census tract and the block group level to understand the relationships between the measured variables. For this purpose, we included the percentage of the demographic and socioeconomic condition variables from ACS for the year 2013-2019. The average completion time and total request volume (aggregated at the appropriate geographic unit) were used as the target variables. The independent variables considered for this purpose were housing conditions (owner-occupied vs rentals, single unit vs multiple units), race (black vs white), ethnicity (Hispanic vs non-Hispanic), gender distribution (male vs female), and economic condition (unemployed vs employed and below poverty vs above poverty). The details of the nine independent and dependent variables are described in Table 2.

Note that for each type, only one variable was retained. For example, female population was retained for gender, since the male population would be the remainder. Inclusion of both variables introduced extraneous correlations into the network, making inferencing more error-prone and noisy. As another example, the total population is mostly composed of black and white populations.

4.2 Signed Bayesian Network

To generate the *Signed Causal Bayesian Network* (sBN) [35], we follow a two-step approach. First, we apply the PC-stable algorithm, a constraint-based BN learning approach from the `bnlearn` R package [37]. This generates an inferred causal network. All

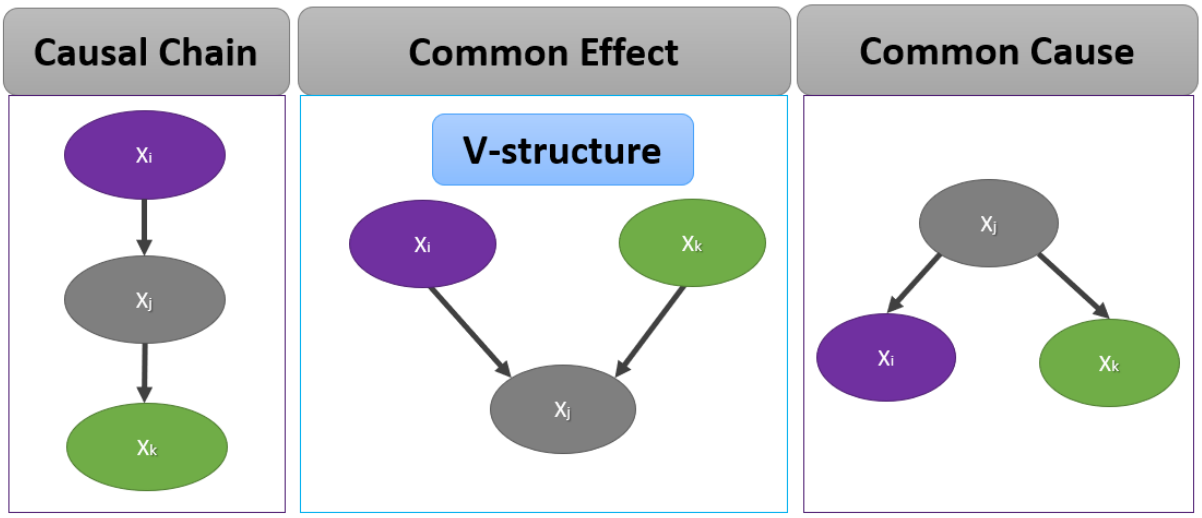


Figure 1: Causal structures among three nodes.

Table 1: Miami-Dade county 311 data description

	No. of records	Avg. Completion Time		
		Block group	**Census tract**	**Zip code**
Total requests	860,254 (41.15%)	6.82	7.10	8.60
Total complaints	133,035 (6.36%)	8.57	8.99	9.43
Total issues	110,023 (5.26%)	14.15	14.78	15.32
Others	986,736 (47.2%)	24.02	25.31	27.60
Total	2,090,177	13.29	18.40	18.95

Table 2: Target and independent variable's statistics

Variable name	**Block group**		**Census tract**		**Zip code**	
	Mean	**SD**	**Mean**	**SD**	**Mean**	**SD**
Volume	114	111	293	309	1627	1853
Completion Time	6.82	5.34	6.55	5.97	8.60	7.69
Owner-occupied units	0.57	0.26	0.59	0.23	0.52	0.20
One unit	0.59	0.35	0.49	0.32	0.47	0.29
Female population	0.51	0.06	0.51	0.04	0.51	0.04
Black population	0.20	0.28	0.19	0.26	0.17	0.21
Hispanic population	0.62	0.28	0.64	0.26	0.56	0.25
Unemployed population	0.42	0.11	0.42	0.09	0.44	0.07
Poor population [3]	0.15	0.12	0.16	0.10	0.14	0.08

the variables are represented as nodes in the network, and each edge is believed to represent a causal dependency. As suggested by Sazal et al. [35], we augment the edges of the network with the help of a *co-occurrence network* (CoNs) [13]. In CoNs, the edges are colored using the Pearson correlation coefficient between the variables. Green (red) colored edge means the correlation between the variable represented by the endpoints is positive (negative, resp.). Finally, the weights of the edges are determined by the bootstrap strength score [14]. This counts the fraction of times the edge appears in the network out of 100 runs. This augmented network is called *Signed Bayesian Network* (sBN).

4.2.1 Block group: The first set of analyses was performed on block level data. The variable representing the volume of requests was categorized as low, medium, or high. The levels were determined using a histogram of values, which suggested a trimodal distribution, as shown in Figure 3. Finally, we use the one-hot encoding for

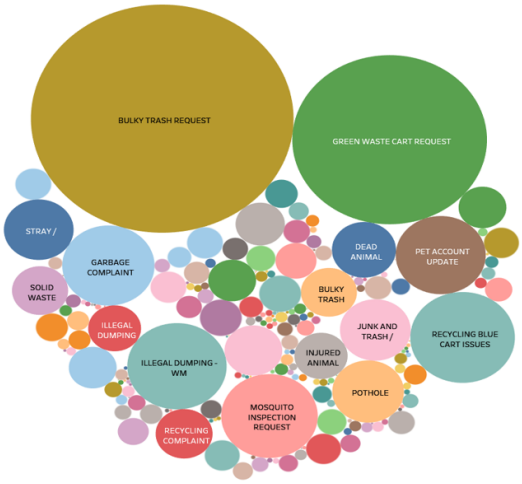


Figure 2: Distribution of types of service requests for Miami-Dade county.

the three different categories. No transformation was performed on the other target variable (completion time) since the variable exhibited a unimodal distribution after the outliers were excluded.

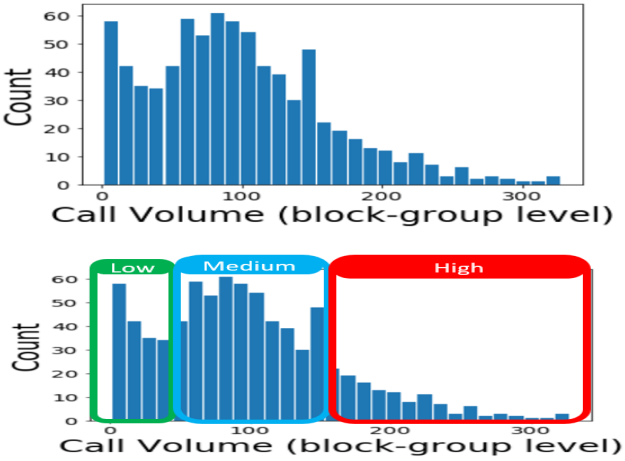
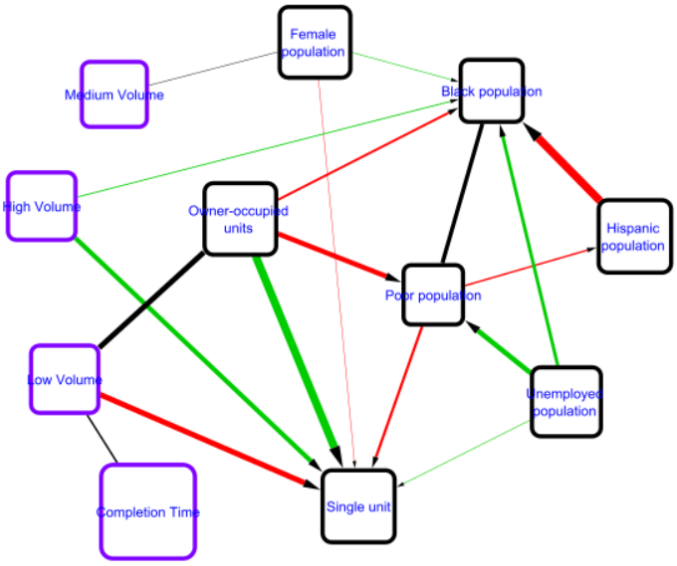
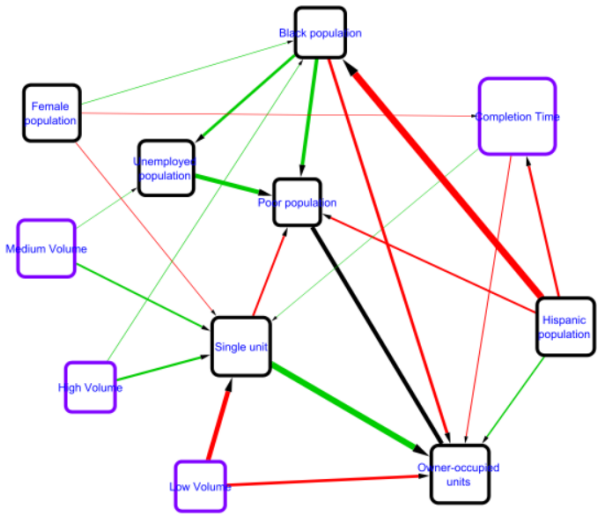


Figure 3: Request Volume histogram (Block group level); Left: Numerical, Right: Categorical.

There were a total of 15 directed edges in the network. The edges were augmented in terms of their color and thickness as described earlier. The target variables had no incoming edges suggesting that none of the variables influenced the volume or completion time. The edge between Completion time and Low volume is undirected. There were two undirected edges between the target and independent variables, one between the Medium Volume and Female population node, and another between Low volume and Owner-occupied units. There were a total of 3 directed edges from the target variables to independent variables. These edges do not support our intuition since, in general, we do not expect an edge from



(a) Using all requests



(b) Using the largest request type only

Figure 4: Signed Bayesian Networks for Block group data.

the target variables to the independent variables. These spurious relationships may be caused by the presence of latent variables (confounders) as described in Section 5. A confounder is a variable that is either not measured or not analyzed, but connects two variables that are connected by an edge. Knowledge of confounders can help correct the dependencies between two variables. These spurious edges include the following edges: Low Volume → Single Unit; High Volume → Single Unit; and High Volume → Black Population.

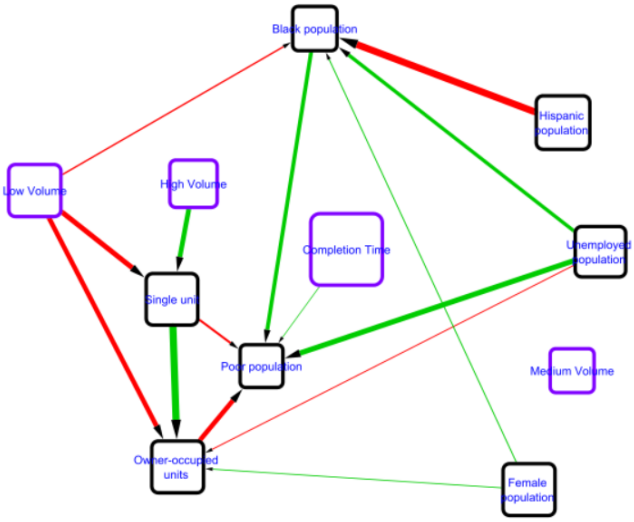
Next, we narrowed down our analysis to only include the largest request type (i.e., Bulky trash request), since this category of request is dominant (501,972 out of 860,254, 58% of the total requests) in the data. As before, the nodes representing call volumes have no incoming edges suggesting that none of the independent variables affect the volume. Completion time has two incoming edges, one from Female population and another from Hispanic population. Both edges have negative Pearson correlation coefficients. Inspecting the inferred regression formula at the nodes suggests that the weights of these two edges are relatively low compared to the others in the network. There is one undirected edge connecting two independent variables, i.e., Poor population and Owner-occupied units, and which cannot be supported by intuition. There are a total of 8 directed edges from the target variables to independent variables that are also likely to be spurious. These edges include the following: Low Volume → Single Unit; Low Volume → Owner-occupied units; High Volume → Single Unit; High Volume → Black population; Medium Volume → Single Unit; Medium Volume → Unemployed population. Completion time → Single unit; and Completion time → Owner-occupied units.

Some edges are intuitive in the network, i.e., Unemployed population → Poor population, Single unit → Owner-occupied units, Single unit → Poor population. It is a known fact that unemployment contributes to the increase in poverty. The other two edges suggest that the single unit is a common cause in this network, influencing both owner-occupied units and poverty. Usually, the single units are owner-occupied, and it is less likely the owner will suffer from poverty. Also, the edge from Female to Black population is supported by the literature [1].

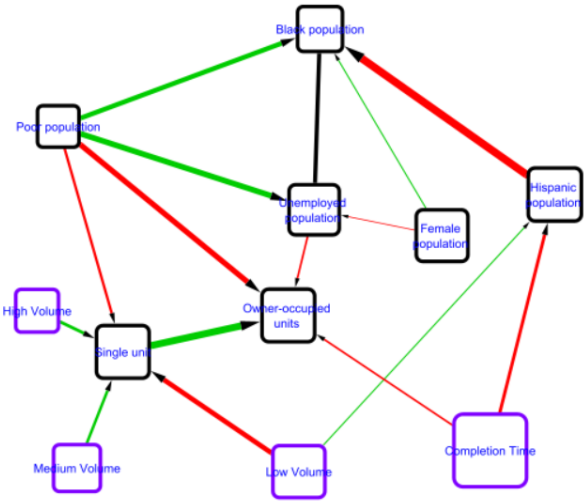
4.2.2 Census tract: Next, we generated the sBNs for the Census tract level. We combined the census data with the aggregated completion time and categorized requested volume on the Census tract level for this task. First, we analyzed the dataset of all requests. The network generated from this set had a comparatively fewer number of edges compared to the network from the block level. As we aggregate larger geographic regions, the data exhibits less diversity in terms of community characteristics. This may explain why we find fewer interactions among the variables.

The network has no undirected edge. The target nodes have no incoming edges. There are a total of 5 directed edges from the target variables to independent variables that are also likely to be spurious. These edges include the following: Low Volume → Single Unit; Low Volume → Owner-occupied units; Low Volume → Black population; High Volume → Single Unit; and Completion time → Poor population.

As before, we also generated the sBNs for the largest request group only. This network has one undirected edge between the Black and Unemployed population. The target nodes have no incoming edges. There are a total of 6 directed edges from the target variables to independent variables. These include: Low Volume → Single Unit; Low Volume → Hispanic population; Completion time → Hispanic population; Completion time → Owner-occupied units; High Volume → Single Unit; and Medium Volume → Single Unit.



(a) Using all requests



(b) Using the largest request type only

Figure 5: Signed Bayesian Networks for Census tract data.

4.3 Discussion

From the resulting networks, we find that the edges Low Volume → Single Unit and High Volume → Single Unit are present in all the structures. The directions are also consistent. It is unusual for target variables to have such outgoing edges to other variables in the network. The algorithms generating these networks are based on heuristics and assumptions, which could lead to misleading inferences. One of the issues with algorithms based on the CI test is that an edge might appear between two variables if a variable (confounder) that affects both these variables is not included in the analysis. Therefore, there is a possibility that confounding variables (common causes) exist for this dataset. Moreover, we cannot draw

any conclusions from this structure, as we know that causal chain and common cause structures are not distinguishable with the CI test, and the direction can be misleading.

4.3.1 v-structures. If the dataset has no hidden confounders, then we can be most confident about the directions of the edges in v-structures in the network. The directions of the rest of the edges are not uniquely determined by the CI tests as explained in Section 3. Although any two edges incoming into a node may appear to be a v-structure, they are labeled as v-structures only after they can be confirmed using the CI test. Our analyses identified three different v-structures, some of which appeared in more than one of these networks.

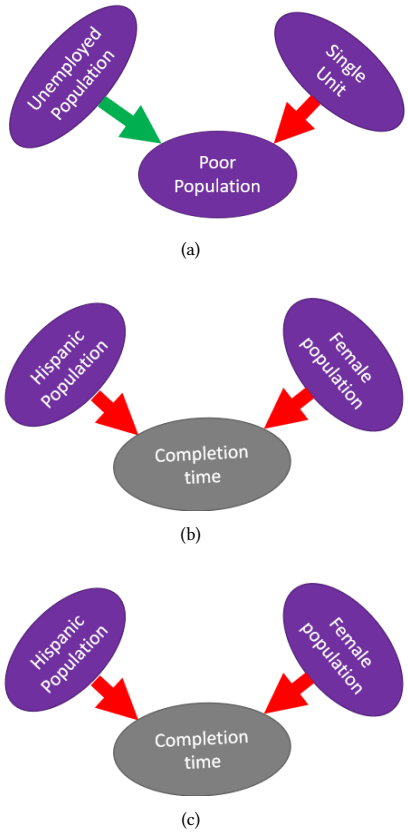


Figure 6: v-structures in the sBNs; (a) and (b) appear in the networks using the largest request type only (Bulky Trash), while (c) appears in the network using all requests.

(a) The v-structure, Unemployed population → Poor population ← Single unit, suggests that the size of the unemployed population results in a rise in the indigent population, which makes intuitive sense. In contrast, the percentage of the Single units in the community affects the Poor population percentage. It is reasonable that the correlation is negative, however the direction of the edge may be spurious. This v-structure is only inferred from the block group data with bulky trash pickup requests.

(b) The v-structure, Hispanic population → Completion Time ← Female population, is an excellent example of a target variable identified as the "common" effect of two independent variables. Both the Hispanic and Female population is inferred to affect Completion time. This structure is only present in the block group data with bulky trash pickup requests. The correlation is negative for both the edges which indicates that the Completion Time decreases with an increase in the Hispanic and Female populations. Neither demographic group can be considered as "minority" in the context of Miami-Dade county (Table 2).

(c) The v-structure, Hispanic population → Black population ← Female population, suggests that a rise in the Hispanic population causes the size of the Black population to decrease. This is consistent with the data that the Hispanic population in Miami-Dade County is predominantly white; The other edge suggests that a rise in the female population results in an increase in the percentage of the Black population. This is consistent with published results from the literature that suggest that *female-Headed* black families has seen an increase in the USA over the years [1]. This structure was inferred in all structures, using both block and Census tract data.

4.3.2 Complaints only: We also examined another type of request, namely complaints. "Complaints" usually take longer (8.5 days on average for block group level) than "Requests". The results from Figure 7 indicate that there is no effect of demographic or socioeconomic status on completion time for this dataset.

This network also has the v-structure, Hispanic population → Black population ← Female population we identified in 6. The nodes representing the target variables are disconnected from all the others except the Low volume. Low volume is affected by both the Owner-occupied units and Female population. One of the findings is that completion time is not affected by any other variables indicating no bias induced by community characteristics.

5 CHALLENGES WITH CAUSAL BAYESIAN NETWORKS

The main challenges in inferring causal relationships from observational data arise from the fact that there are several assumptions made in applying causal inferencing. Real-world observations may not always follow those assumptions; hence, it introduces challenges in applying the method successfully.

5.1 Missing and impure data

The process of data acquisition process often results in incomplete administrative datasets. Administrative datasets may contain missing data points, and may have recording errors. For example, in the 311 data, we found records whose request status has never been closed. Other records show the closing date recorded to be prior to the open date. Such inaccurate records (16.28% of the total) must either be manipulated or ignored, thus reducing the number of accurate observations available for analysis. When we exclude the inaccurate data points, the number of total observations decreases.

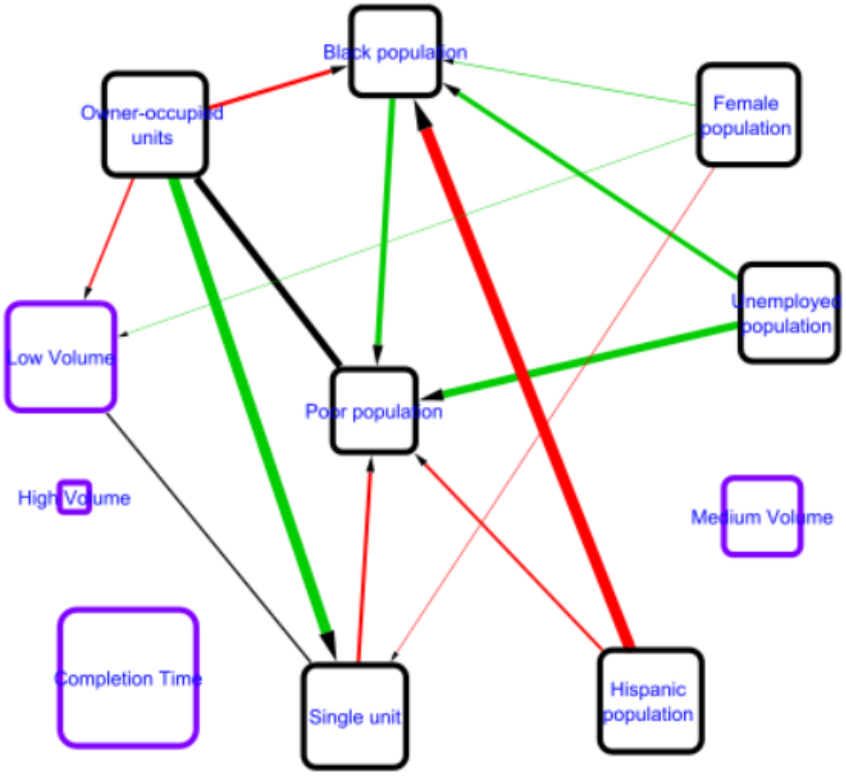


Figure 7: Signed Bayesian Network: using only "Complaints" (Block group).

5.2 Inadequate data

The accuracy of machine learned models also depend on the quantity of available data. Large amounts of good data can help develop better models as they can better capture the underlying relationships under investigation. This requirement becomes higher for high-dimensional data.

Even though the total number of observations may be large for some 311 datasets, a single record is not associated with the features of the individual who requested the service; this is done to ensure anonymity in public datasets. As a result, we consider the community characteristics rather than metadata associated with individuals, obtained by aggregating the data on an appropriate geographic unit, such as a block group or a census tract. When we aggregate the data on a geographical unit, the resulting number of observations is reduced while the geographical location data gets coarser. For example, when we aggregate the Zip code level observations, the number of records becomes considerably smaller (5-fold reduction from the Census tract level). Inadequate data tends to produce a misleading model. Barring such fine-grained information from the 311 data, there is a greater possibility for biases in the data (e.g., by a small set of individuals making most of the requests). An alternative data collection approach that collects demographic data of each individual requester could add valuable richness to the analysis proposed here.

5.3 Latent confounders

Finally, there exist variables that are missing in the dataset, either because they were not measured, or because there was no known way to measure them, but affect the target variables by creating incomplete models, and leading to potentially incorrect inferences. These variables are called latent "confounders". Structure learning models are based on the assumption that all the independent variables affecting the target variables are present in the observation, which may not be true. In such cases, the model cannot discover the real cause-effect relationships accurately. It is often impossible to avoid the possibility of having unobserved variables in the real world because they are often unknown. When the data fails to capture the key factors of interest, the model will also be inadequate in explaining the findings. Our analysis discovered that Low Volume → Single Unit and High Volume → Single Unit are recurring edges in many of the inferred networks. Intuitively, the requested volume should not cause the housing conditions to change. We can explain this edge with the help of confounding variables. We assume that a latent variable, not included in the network, affects both the requested volume and housing condition resulting in a directed edge between them. We used a limited number of demographic and socioeconomic information. Our analysis did not include any information regarding the department (a specific unit

that handles particular types of requests), infrastructure, and resources. Excluding that information may result in an incomplete model since the available resources may impact the efficiency while also being correlated to the types of homes in that block.

Also, since the structure learning algorithms are based on heuristics, more than one structurally equivalent BN can be obtained from the same observations. Once the v-structures are identified, the structure learning algorithm's last step assigns the remaining directions based on some predefined rules. These rules may still leave some edges to be undirected. Also, the directions inferred may be inconclusive in some cases due to ambiguity in determining the causal chain and common cause structures. We try to overcome this limitation with bootstrapping by generating several models, and assigning weights to each edge.

6 CONCLUSION

The 311 administrative dataset paves the way to study the government's responsiveness in providing non-emergency services to local residents. We analyzed data from Miami-Dade county to measure the effectiveness in terms of completion time and call request volumes. We applied causal inference resulting in causal Bayesian Network models, which helps to determine the demographic and socioeconomic factors that have a causal impact on the target variables such as completion time and call request volumes. We concluded that the results do not support demographic and socioeconomic bias in providing non-emergency services to the residents of Miami-Dade county only for the considered independent variables. The case study using data from just one city cannot ensure that data from other cities or municipalities will result in the same conclusion. However, the findings are consistent with extant research on 311 data from other cities. More importantly, this paper aims to provide a framework to apply causal inference on 311 datasets, which can be readily extended to data from other cities or regions. Finally, we provide a discussion on the challenges in applying the causal approach to this type of dataset.

ACKNOWLEDGMENTS

This work was partially supported by an NSF grant OAC-1924154 (SC, SG, GN)).

REFERENCES

[1] [n.d.]. Female-headed black family statistics. https://www.statista.com/statistics/205106/number-of-black-families-with-a-female-householder-in-the-us.
[2] [n.d.]. Miami-Dade County Open Data Hub. https://gis-mdc.opendata.arcgis.com.
[3] [n.d.]. Poverty definition by Census. https://www.census.gov/quickfacts/fact/note/US/IPE120219.
[4] Silvia Acid, Luis M de Campos, Juan M Fernández-Luna, Susana Rodrıguez, José Marıa Rodrıguez, and José Luis Salcedo. 2004. A comparison of learning algorithms for Bayesian networks: a case study based on data from an emergency medical service. *Artificial intelligence in medicine* 30, 3 (2004), 215–232.
[5] Yahya Y Bayraktarli, Jens-Peder Ulfkjaer, Ufuk Yazgan, and Michael H Faber. 2005. On the application of Bayesian probabilistic networks for earthquake risk management. In *9th International Conference on Structural Safety and Reliability (ICOSSAR 05)*. 20–23.
[6] Baoping Cai, Yonghong Liu, Zengkai Liu, Xiaojie Tian, Yanzhen Zhang, and Renjie Ji. 2013. Application of Bayesian networks in quantitative risk assessment of subsea blowout preventer operations. *Risk Analysis* 33, 7 (2013), 1293–1311.
[7] Benjamin Y Clark et al. 2014. Do 311 Systems Shape Citizen Satisfaction with Local Government? *Available at SSRN 2491034* (2014).
[8] Benjamin Y Clark, Jeffrey L Brudney, and Sung-Gheel Jang. 2013. Coproduction of government services and the new information technology: Investigating the distributional biases. *Public Administration Review* 73, 5 (2013), 687–701.
[9] Benjamin Y Clark, Jeffrey L Brudney, Sung-Gheel Jang, and Bradford Davy. 2020. Do Advanced Information Technologies Produce Equitable Government Responses in Coproduction: An Examination of 311 Systems in 15 US Cities. *The American Review of Public Administration* 50, 3 (2020), 315–327.
[10] Diego Colombo and Marloes H Maathuis. 2014. Order-independent constraint-based causal structure learning. *The Journal of Machine Learning Research* 15, 1 (2014), 3741–3782.
[11] Diego Colombo, Marloes H Maathuis, Markus Kalisch, and Thomas S Richardson. 2012. Learning high-dimensional directed acyclic graphs with latent and selection variables. *The Annals of Statistics* (2012), 294–321.
[12] James R Elliott and Jeremy Pais. 2006. Race, class, and Hurricane Katrina: Social differences in human responses to disaster. *Social science research* 35, 2 (2006), 295–321.
[13] Mitch Fernandez, Juan D Riveros, Michael Campos, Kalai Mathee, and Giri Narasimhan. 2015. Microbial" social networks". *BMC genomics* 16, 11 (2015), S6.
[14] Nir Friedman, Moises Goldszmidt, and Abraham Wyner. 2013. Data analysis with Bayesian networks: A bootstrap approach. *arXiv preprint arXiv:1301.6695* (2013).
[15] Nir Friedman, Michal Linial, Iftach Nachman, and Dana Pe'er. 2000. Using Bayesian networks to analyze expression data. *Journal of computational biology* 7, 3-4 (2000), 601–620.
[16] Pilar Fuster-Parra, P Tauler, M Bennasar-Veny, A Ligęza, AA Lopez-Gonzalez, and A Aguilo. 2016. Bayesian network modeling: A case study of an epidemiologic system analysis of cardiovascular risk. *Computer methods and programs in biomedicine* 126 (2016), 128–142.
[17] Xian Gao. 2018. Learning within the 311 service policy community: Conceptual framework and case study of Kansas City 311 program. (2018).
[18] Nicholas J Harding. 2011. *Application of Bayesian networks to problems within obesity epidemiology.* Ph.D. Dissertation. The University of Manchester (United Kingdom).
[19] Sarah Hartmann, Agnes Mainka, and Wolfgang G Stock. 2017. Citizen relationship management in local governments: The potential of 311 for public service delivery. In *Beyond Bureaucracy*. Springer, 337–353.
[20] Linwood D Hudson, Bryan S Ware, Kathryn B Laskey, and Suzanne M Mahoney. 2005. *An application of Bayesian networks to antiterrorism risk management for military planners.* Technical Report. George Mason University.
[21] Markus Kalisch, Martin Mächler, Diego Colombo, Marloes H Maathuis, and Peter Bühlmann. 2012. Causal inference using graphical models with the R package pcalg. *Journal of Statistical Software* 47, 11 (2012), 1–26.
[22] Daphne Koller and Nir Friedman. 2009. *Probabilistic graphical models: principles and techniques.* MIT press.
[23] Constantine Kontokosta, Boyeong Hong, and Kristi Korsberg. 2017. Equity in 311 reporting: Understanding socio-spatial differentials in the propensity to complain. *arXiv preprint arXiv:1710.02452* (2017).
[24] Kevin B Korb and Ann E Nicholson. 2010. *Bayesian artificial intelligence.* CRC press.
[25] Chang-Ju Lee and Kun Jai Lee. 2006. Application of Bayesian network to the probabilistic risk assessment of nuclear waste disposal. *Reliability Engineering & System Safety* 91, 5 (2006), 515–532.
[26] Chee Kian Leong. 2016. Credit risk scoring with bayesian network models. *Computational Economics* 47, 3 (2016), 423–446.
[27] Yuchen Li, Ayaz Hyder, Lauren T Southerland, Gretchen Hammond, Adam Porr, and Harvey J Miller. 2020. 311 service requests as indicators of neighborhood distress and opioid use disorder. *Scientific reports* 10, 1 (2020), 1–11.
[28] Nader Madkour. 2020. *Predicting Non-Emergency 311 Requests for an Efficient Resource Allocation After a Disaster in Houston, Tx.* Ph.D. Dissertation. Lamar University-Beaumont.
[29] Dimitris Margaritis. 2003. *Learning Bayesian network model structure from data.* Technical Report. Carnegie-Mellon Univ Pittsburgh Pa School of Computer Science.
[30] Taewoo Nam and Theresa A Pardo. 2013. Identifying success factors and challenges of 311-driven service integration: a comparative case study of NYC311 and Philly311. In *Proceedings of the 46th Hawaii international conference on system sciences.*
[31] Arnold Ojugo and Oghenevwede Debby Otakore. 2020. Forging an optimized bayesian network model with selected parameters for detection of the coronavirus in Delta State of Nigeria. *Journal of Applied Science, Engineering, Technology, and Education* 3, 1 (2020), 37–45.
[32] Judea Pearl, F Bacchus, P Spirtes, C Glymour, and R Scheines. 1995. Probabilistic reasoning in intelligent systems: Networks of plausible inference. (1995).
[33] Anna Rigosi, Paul Hanson, David P Hamilton, Matthew Hipsey, James A Rusak, Julie Bois, Karin Sparber, Ingrid Chorus, Andrew J Watkinson, Boqiang Qin, et al. 2015. Determining the probability of cyanobacterial blooms: the application of Bayesian networks in multiple lake systems. *Ecological Applications* 25, 1 (2015), 186–199.

[34] Karen Sachs, David Gifford, Tommi Jaakkola, Peter Sorger, and Douglas A Lauffenburger. 2002. Bayesian network approach to cell signaling pathway modeling. *Science's STKE* 2002, 148 (2002), pe38–pe38.
[35] Musfiquer Sazal, Kalai Mathee, Daniel Ruiz-Perez, Trevor Cickovski, and Giri Narasimhan. 2020. Inferring directional relationships in microbial communities using signed Bayesian networks. (2020).
[36] Musfiqur Rahman Sazal, Daniel Ruiz-Perez, Trevor Cickovski, and Giri Narasimhan. 2018. Inferring relationships in microbiomes from signed bayesian networks. In *2018 IEEE 8th International Conference on Computational Advances in Bio and Medical Sciences (ICCABS)*. IEEE, 1–1.
[37] Marco Scutari. 2009. Learning Bayesian networks with the bnlearn R package. *arXiv preprint arXiv:0908.3817* (2009).
[38] Marco Scutari. 2014. Bayesian network constraint-based structure learning algorithms: Parallel and optimised implementations in the bnlearn r package. *arXiv preprint arXiv:1406.7648* (2014).
[39] Marco Scutari and Jean-Baptiste Denis. 2014. *Bayesian networks: with examples in R*. CRC press.
[40] P Spirtes, C Glymour, and R Scheines. 1993. Causation, prediction and search (MIT Press, Cambridge). (1993).
[41] Mark Steyvers, Joshua B Tenenbaum, Eric-Jan Wagenmakers, and Ben Blum. 2003. Inferring causal networks from observations and interventions. *Cognitive science* 27, 3 (2003), 453–489.
[42] John Clayton Thomas. 2013. Citizen, customer, partner: Rethinking the place of the public in public management. *Public Administration Review* 73, 6 (2013), 786–796.
[43] Michail Tsagris. 2020. A New Scalable Bayesian Network Learning Algorithm with Applications to Economics. *Computational Economics* (2020), 1–27.
[44] Thomas Verma and Judea Pearl. 1992. An algorithm for deciding if a set of observed independencies has a causal explanation. In *Uncertainty in artificial intelligence*. Elsevier, 323–330.
[45] Eric W Welch, Charles C Hinnant, and M Jae Moon. 2005. Linking citizen satisfaction with e-government and trust in government. *Journal of public administration research and theory* 15, 3 (2005), 371–391.
[46] Corey Kewei Xu and Tian Tang. 2020. Closing the Gap or Widening the Divide: The Impacts of Technology-Enabled Coproduction on Equity in Public Service Delivery. *Public Administration Review* (2020).
[47] Li Xu, Mei-Po Kwan, Sara McLafferty, and Shaowen Wang. 2017. Predicting demand for 311 non-emergency municipal services: An adaptive space-time kernel approach. *Applied geography* 89 (2017), 133–141.
[48] Yilong Zha and Manuela Veloso. 2014. Profiling and prediction of non-emergency calls in new york city. In *Proceedings of the Workshop on Semantic Cities: Beyond Open Data to Models, Standards and Reasoning, AAAI*.
[49] Xiao-Fei Zhang, Le Ou-Yang, and Hong Yan. 2017. Incorporating prior information into differential network analysis using non-paranormal graphical models. *Bioinformatics* 33, 16 (2017), 2436–2445.
[50] Christopher W Zobel, Milad Baghersad, and Yang Zhang. 2017. Calling 311: evaluating the performance of municipal services after disasters.. In *ISCRAM*.