



Analyzing network-wide patterns of rail transit delays using Bayesian network learning

Mehmet Baran Ulak^{c,*}, Anil Yazici^a, Yun Zhang^b

^a Department of Civil Engineering, Stony Brook University, Stony Brook, NY 11794, United States

^b Mobileware Inc., Stony Brook, NY 11790, United States

^c Department of Civil Engineering, University of Twente, Enschede 7522 NB, Netherlands



ARTICLE INFO

Keywords:

Transit delays
Crowdsourced data
Delay pattern identification
Bayesian network learning
Network dependency metrics

ABSTRACT

Rail transit delays are generally discussed in terms of on-time performance or problems at individual stops. Such stop-scale approaches ignore the fact that delays are also caused and perpetuated by network-wide factors (e.g., bottlenecks caused by shared tracks by multiple transit lines). The objective of this paper is to develop a network model and metrics that can quantify the delay dependencies between transit network stops, and identify local sources of network-wide issues. For this purpose, Bayesian network learning (at the intersection of machine learning and network science) was utilized. Based on the calculated Bayesian networks (BNs), network metrics (*inducer* and *susceptible*) were formulated to quantify the network-wide impacts of the delays experienced at the stops. To implement the proposed framework, the delays at Long Island Rail Road (LIRR) were gathered through a crowdsourced real-time transit information app called onTime. The developed BN model was tested through cross-validation, yielded promising accuracy results, successfully identified the problematic stops based on LIRR reports, and provided further insights on network impacts. The BN model and the developed metrics were further tested using a natural experiment, i.e., a before and after study focusing on a recently completed track expansion project at LIRR. The findings imply that BN learning can successfully identify the network dependencies and indicate the rail links/corridors that are the best candidate for subsequent improvement investments. Overall, the developed metrics can quantify the delay dependencies between stops and they can be used by policy makers and practitioners for investment and improvement decisions.

1. Introduction

Public transportation is an important component of a healthy transportation system. In order to maintain an efficient public transportation system, it is required to exhibit a high level of reliability, which is defined as the certainty of being served by the provided services within a given schedule (van Oort, 2014). Reliability is affected by transit delays that can occur due to various causes: irregular passenger movements or high volume of boarding/alighting passengers; insufficient capacity of rail tracks; train connections, bottlenecks (i.e., bifurcations) on the rail network; environmental conditions (i.e., weather); equipment failures and

* Corresponding author at: Transportation Engineering and Management Group, Department of Civil Engineering, University of Twente, Enschede 7522 NB, Netherlands. Tel.: +1 929 262 9870.

E-mail addresses: m.b.ulak@utwente.nl (M.B. Ulak), Anil.Yazici@stonybrook.edu (A. Yazici), yun.zhang@mobilewareinc.com (Y. Zhang).

<https://doi.org/10.1016/j.trc.2020.102749>

Received 20 December 2019; Received in revised form 17 June 2020; Accepted 8 August 2020

Available online 20 August 2020

0968-090X/© 2020 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

breakdowns; and crashes (Flier et al., 2009; Harris et al., 2013; Lee et al., 2016). These causes can be classified as primary or secondary depending on the nature of the event causing delay.

Primary delays can be caused by the events such as the equipment failure or crashes. Secondary delays (as known as the *knock-on* delays) are fallout of the primary delays on other trains. Furthermore, drawing from the existing terminology, delays at rail transit systems can be categorized as recurrent and non-recurrent. Non-systematic events such as equipment failures, breakdowns, or crashes would lead to nonrecurrent delays. Recurrent delays, on the other hand, are caused by daily operations (e.g., passenger alighting and boarding), systematic dependencies between different routes (e.g., prioritizing or holding trains approaching the same stop/track), and operational constraints such as track capacity imposed by the infrastructure. Moreover, delays can propagate due to the connectedness of the railways and utilization of common routes by other trains (Flier et al., 2009), and can eventually cause a domino effect on the entire system (Goverde, 2010). Such domino effects indicate the importance of the network-wide thinking, i.e., dependencies between railway stops in terms of amplifying or negating the existing delays.

A network-wide analysis requires delay data for the whole rail network in order to identify the delay patterns and dependencies. Fortunately, the wider utilization of GPS enabled tracking of trains by transit agencies and crowdsourced applications that collect train arrival information can provide the necessary data. The developments in machine learning and network science fields in parallel to the increasing availability of large datasets enable researchers to undertake the task of network delay analysis.

Accordingly, this study delves into the transit delay data to achieve following objectives: 1) developing a network model that can identify the relationship between delay patterns at transit stops, i.e., which stops induce, propagate, and/or negate the delays throughout the rail network, and 2) quantifying the dependencies between the transit stops through network metrics in order to help policy makers and practitioner prioritize the investments and/or projects on the rail network. Thereby, this study investigates delays from a network-wide perspective as stop-delays in the complete transit network were analyzed, rather than station-level factors (e.g., passenger movements, passenger volume, etc.) affecting the delays. For this purpose, Bayesian network learning was utilized. Based on the results of developed Bayesian networks (BNs), network metrics were formulated to quantify the network-wide impacts of the delays experienced at individual stops. To implement the proposed framework, delays (in minutes) experienced at stops of the Long Island Rail Road (LIRR) were gathered through a crowdsourced real-time transit information app called onTime.

This study contributes to the body of research on the railway network delays as follows: 1) Bayesian network learning approach was used to identify delay patterns of complex and intricate transit network for the first time; 2) Bayesian network-based metrics were developed in order to provide a tangible index that can be used for decision-making and performance evaluation. In the following sections, first, the literature review on delays at transit networks and Bayesian network learning are provided, followed by the description of the study area, and the data. Second, the Bayesian network learning approach is discussed, data processing is explained, and the network metrics are formulated. Then the developed BN model is tested for accuracy, validated through natural experiment of before & after the double-track project at LIRR Ronkonkoma Branch. Last, the findings are summarized in the conclusion section along with caveats and future research directions.

2. Literature review

Several studies have investigated delay prediction, delay propagation forecasting, delay pattern identification in rail networks, and dependency of transit components, previously (see Ghofrani et al., 2018 for a review of these studies). It is widely acknowledged that detecting the delay dependencies among different elements (e.g., trains, stations) in the rail network is critical in order to maintain reliable and efficient railway transport. However, detecting such delay dependencies is complicated and requires extensive knowledge about the infrastructure (Conte, 2007; Conte and Schobel, 2007). Nonetheless, Conte and Schobel (2007) identified the delay dependencies without such comprehensive information about the system by using the so called Tri-Graph approach from a family of graphical methods (Wille and Bühlmann, 2004). The approach basically identifies the full conditional model for events (e.g., departure or arrival of trains) via modelling subgroups of events at a time to find conditional dependence among these events. However, such subgrouping requires scrutinizing the timetable of trains and hence is difficult to implement to complex transit networks. Therefore, authors implemented the model to a long distance railway network comprising 8 stations with 928 events and it is not clear whether the approach could perform for a complex network and big data. Nonetheless, authors showed that graph-based models perform well for identification of delay dependencies in a railway system. Flier et al. (2009) developed a set of algorithms to identify timetable-caused dependencies among trains operating in the Swiss rail network. Their study showed that it is necessary to discover the underlying relationship between train trips in order to accurately identify delays at a large rail network. However, their approach was developed to detect dependencies at a single stop rather than identifying network-wide delay dependencies.

Delay propagation in large networks is also a concern for maintaining efficient operation (Büker and Seybold, 2012; Sørensen et al., 2017; Wallander and Mäkitalo, 2012). Focusing on this issue, Büker and Seybold (2012) developed a stochastic approach that computes the potential delay propagation in large networks. This framework is argued to have the potential to assess the reliability of the railway timetables. With the same motivation, Wallander and Mäkitalo (2012) adopted a “sequence analysis” based on association analysis (i.e., co-occurrence of events) to detect train delays that occur sequentially implying a propagating pattern. Authors sought to reveal relationships between delays of long distance trains of the Finnish railways and to identify causes of these delays. However, the study focused on delays of trains arriving to and departing from a single station and did not consider a complete rail network. A similar association analysis was also implemented by Yabuki et al. (2015), in order to evaluate the effectiveness of countermeasures that are adopted to reduce delays and delay propagations.

Several approaches such as support vector machines (Marković et al., 2015), neural networks (NN) (Milinković et al., 2013; Yaghini et al., 2013), deep and shallow learning (Oneto et al., 2017), data driven approaches (Wang and Work, 2015), random forest models

(Nabian et al., 2019), Markov models (Gaurav and Srivastava, 2018), econometric models (Agbelie and Libnao, 2018; Gorman, 2009), hybrid models (Oneto et al., 2020), and Bayesian networks (Corman and Kecman, 2018; Huang et al., 2020; Lessan and Fu, 2019) were implemented in the literature. These approaches help predict the railway delays and compare alternative options in the context of delays. Neural network approach was reported to outperform decision tree and multinomial logistic regressions in predicting the delays (Yaghini et al., 2013). Compared to neural networks, the support vector machine modeling was found to be better in capturing the relationship between delays and the railway system features (Marković et al., 2015).

Bayesian network learning was considered in the context of delay analysis by Conte (2007) for the first time, however, only recently utilized in delay prediction by Lessan and Fu (2019), Corman and Kecman (2018), and Huang et al. (2020). All recent studies used BN learning to construct relationships between nodes and developed delay prediction models. Corman and Kecman (2018) proposed a stochastic delay prediction model in which delay predictions evolve using the real-time data feed. Lessan and Fu (2019), on the other hand, tested the performance of three different BN learning approaches, namely hill-climbing, primitive linear, and hybrid BN for delay prediction. Study found that hybrid BN approach performs better than the other two particularly in terms of variation of results; however, their results show that there is not much difference between the approaches in terms of average prediction error. In a recent study, Huang et al. (2020) developed a hybrid BN model to predict delays, number of affected trains, and total delay through constructing relationships among these three variables. Nevertheless, none of these studies examined the delay patterns and delay dependencies among components (i.e., stops) of complex transit networks by using Bayesian network learning.

Identifying delay patterns and sources of delays (e.g., network bottlenecks) is very important to improve the reliability of transit systems (Cule et al., 2011). However, discovering the true origin of the delays is generally difficult due to the network complexity. Nonetheless, there are studies focused on this problem by identifying delay causing train sequences (Cule et al., 2011); developing algorithms to analyze delay propagation in single tracks (Sørensen et al., 2017); or optimizing the train timetables (Andersson et al., 2015). In order to understand the systematic delay issues of Belgian railways and reveal potential underlying causes of delays, Cule et al. utilized frequent pattern mining techniques. Researchers did not use the duration of delays but considered the trains as either delayed or not, in order to reveal sequences of trains that cause delays. Although this approach is useful to identify delay sequences of trains in a systematic manner, the magnitude of the network-wide dependencies is not fully revealed. The majority of the studies in the literature focus on delays of the long distance trains that operate on relatively less complex rail networks. However, very few studies attempt to scrutinize the delay patterns in commuter networks generally plagued by high volume, low track capacity, and complex infrastructure. This study aims to address this gap by developing a Bayesian network model and metrics that can elucidate the frequent delay patterns in transit systems, quantify the delay dependencies between transit network stops, and identify local sources of network-



Fig. 1. Complete Long Island Rail Road network (MTA, 2019b) and the highest ridership portion.

wide issues.

3. Case study area and data

3.1. Study area

The study area is the Long Island Rail Road (LIRR) system which connects New York City to Long Island, New York via multiple lines (Fig. 1). The LIRR system has a total of 11 branches (i.e., rail lines) covering 620 miles of tracks along the Long Island with 124 stops. There are two terminal stops located at New York City, namely Penn Station and Atlantic Terminal. LIRR carried 89.2 million riders in 2017 on its 735 daily trains. These statistics make LIRR the busiest commuter railroad in North America with approximately \$710 million in revenues (Metropolitan Transportation Authority, 2017). In 2017, LIRR reported 91.4% on-time performance (i.e., trips with less than 5 min and 59 s of delay at stops).

Table 1 shows the daily number of trains at each LIRR stops within the data span of 32 days between 12/23/2017 and 01/23/2018. Numbers on Table 1 are estimated based on the scheduled trips within the study period (MTA, 2019a). Jamaica Station has the highest number of trains with a total of 543 trains per day on average, which is followed by Penn Station with 470 trains on average per day.

3.2. Crowdsourced delay data – onTime user logs

onTime is a real-time transit information app which has a crowdsourcing feature where users can check-in (opt-in) to track the status of a train; thereby the users let the app use their GPS coordinates during their trip. The users' coordinates are collected anonymously and used to calculate the distance to the next stop and the train speed, and accordingly, the estimated time of arrival (ETA) for each station is calculated. The train delays for each stop are calculated by comparing these ETAs with the scheduled arrival times, i.e., published timetables of the Metropolitan Transportation Authority (MTA, 2019a; Transitfeeds, 2019). These delays are communicated back to all commuters who are using the app. Note that the app users are not actively involved (by entering information) in the calculation of estimated time of arrival and delays. It is possible (yet very rare) that a train may arrive at a stop before scheduled arrival time. In such cases, there is no delay recorded for the analysis. At such early arrival cases, the train does not depart until its scheduled time. Hence, from the delay propagation and modeling perspective, early arrival and no delay are practically equivalent, and the early arrivals do not affect the model results.

The crowdsourced log data include 372,502 time-stamped logs of *onTime* users within the LIRR system between 12/23/2017 and 01/23/2018 (32 days). Among these logs, there are records with very large delays which were deemed to be either erroneous or as a result of an extreme situation. In order to avoid biased and inaccurate estimates due to those outlier records, delays longer than two hours were discarded. Note that, these records only correspond to 0.2% of the data (756 out of 372,502 logs); hence, the exclusion of those outlier records will not bias the data (especially within probabilistic models such as Bayesian Networks). The final dataset contains 371,746 delay estimates along the LIRR network. Fig. 2 shows a sample of these delays throughout the LIRR tracks, illustrating that the delay magnitude is increasing towards New York City (west).

Table 1

Average number of trains stopping at stops per day within the data span.

Branch	Stop	Volume	Branch	Stop	Volume
New York	Penn Station	470	Ronkonkoma	Bethpage	72
City Terminal	Woodside	263		Farmingdale	72
Zone	Hunterspoint Ave	14		Wyandanch	67
	Forest Hills	92		Deer Park	68
	Kew Gardens	91		Brentwood	68
	Jamaica	543		Central Islip	69
	St. Albans	32	Babylon	Rosedale	57
	Atlantic Terminal	121		Laurelton	56
	Nostrand Avenue	103		Valley Stream	113
	East New York	104		Lynbrook	109
	Locust Manor	56		Rockville Centre	105
Port Jefferson (Main Line)	New Hyde Park	72		Baldwin	106
	Merillon Avenue	73		Freeport	123
	Mineola	181		Merrick	106
	Carle Place	66		Bellmore	113
	Westbury	89		Wantagh	107
	Hicksville	192		Seaford	106
Port Jefferson	Cold Spring Harbor	113		Massapequa	114
	Huntington	138		Babylon	168
				Bay Shore	43

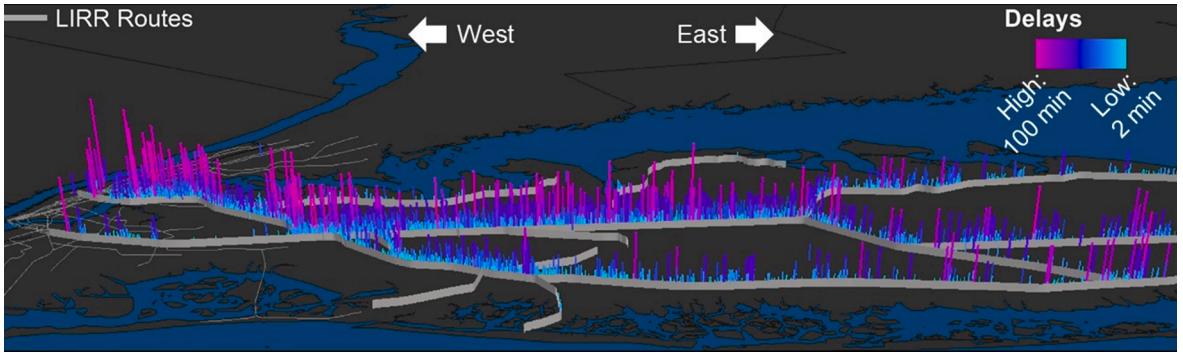


Fig. 2. A sample of delays experienced in the LIRR system.

4. Methodology

4.1. Bayesian network learning and inference

The relationships between stops in terms of experienced delays were investigated using a Bayesian network (BN) learning approach. A Bayesian network is a statistical model in which conditional dependencies of random variables are represented by a directed acyclic (i.e., one-way interaction) graph (DAG) (Buntine, 1996; Charniak, 1991). That is, DAG is a type of graph whose edges (i.e. paths) are directed and not cyclic (i.e. one-way relationship). Bayesian networks are useful to represent the statistically significant relationships between events that possibly have causal dependencies with each other as parent node and child node. Therefore, Bayesian networks are useful to predict the probability of a set of events to occur given that other sets of events were observed. A popular example in this context is that the symptoms manifested by a patient can be used to estimate the probabilities of diseases which patient may have (Buntine, 1996). BN approach has been recently adopted in the rail transportation and delay prediction (Conte, 2007; Corman and Kecman, 2018; Huang et al., 2020; Lessan and Fu, 2019; Zilko et al., 2016). These studies showed that BN learning is very efficient in analyzing and predicting delays at components of railway networks. Given the promising results, BN was adopted in the current study as well. However, note that, the prediction accuracy of the model is also dependent on the data, even though the prediction capabilities of the BN learning were shown in the literature. For the case of this study, spatio-temporal coverage of the crowdsourced data is an important factor on the accuracy of the model, i.e., the more the number of user delay records for each stop at different times of the day, the better the prediction accuracy. Overall, the onTime records provide sufficient spatio-temporal coverage for good prediction accuracy that was shown in subsequent sections through cross-validation.

A Bayesian network structure can be represented by a DAG that has n nodes($1, 2, \dots, n$) where each node j is associated with a random variable $X_j \in (X_1, X_2, \dots, X_n)$. Given this DAG, conditional dependencies of n random variable X can be represented as a Bayesian network (Buntine, 1996; Charniak, 1991):

$$P(X_1, X_2, \dots, X_n) = \prod_{j=1}^n P_j | \Pi_{X_j} \quad (1)$$

where Π_{X_j} is the parent nodes of node j . Fig. 3 illustrates a simple Bayesian network that has seven nodes and seven edges. In this graph, nodes A and B are parent nodes of node D, while nodes F and G are the child nodes of node D. This network indicates that node D is dependent on the nodes A and B. Similarly, node F is dependent on the node D, whereas it is conditionally independent of A and B. Please refer to Buntine (1996) and Scutari (2010) for further details on the theory and application of Bayesian networks.

In this study, Bayesian network learning was implemented using the R-project and “bnlearn” package. Maximum likelihood estimation (MLE) was utilized to estimate the relationship between nodes (i.e. parent-child) and a hybrid structure learning algorithm (Lessan and Fu, 2019), namely max-min hill-climbing (mmhc) method, was implemented to learn the graph structures heuristically. The final graphs are Bayesian networks consisting of statistically significant edges that are identified based on p values of the

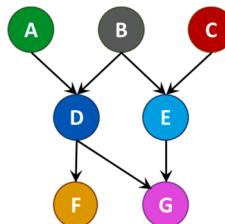


Fig. 3. An example Bayesian network.

conditional independence tests with a threshold of 0.05 (i.e., 95% confidence interval) (Scutari, 2010; Scutari and Ness, 2019).

4.2. Implementation of Bayesian network approach

4.2.1. Data processing

The crowdsourced data consist of logs with estimated delays. These logs were further processed to identify the most accurate estimates of the delays. That is, generally more than one commuter use the onTime app to check the delays at the same time; thus there are several records that provide delay estimates for the same “trip-destination-time” set. This is illustrated in Fig. 4 that shows seven logs for a hypothetical example *Train 1*. The most accurate delay for this *Train 1* arriving to a hypothetical *Stop B* is provided by the last log (Log #7) since it is the closest to *Stop B*. In the dataset with total of 371,746 logs, 45,621 such records (e.g., Log #7 in the figure) were included in the analysis. In other words, 326,125 user logs out of 371,746 provide less accurate delay estimations than the last logs, such as logs #1 to #6 compared to Log #7 in Fig. 4. Those earlier records were not considered, and the last (hence most accurate) logs were used to estimate the train delay.

In order to develop Bayesian network of stops, the stop delays were used to develop a network delay data matrix (Table 2). For this purpose, each record was classified according to time of day (e.g., 5:00–6:00 PM), trip direction (East-West), and arrival stop. Then, the delays were aggregated based on these stop-time intervals, and the total delay at the stop for a given time interval was identified. This process is exemplified in Fig. 4 where three trains arrive at an imaginary stop B between 10:00 and 11:00 AM. The total delay at the stop B for this time interval was calculated by summing three delay estimates obtained from Log #7, Log #6, and Log #4 for imaginary *Train 1*, *Train 2*, and *Train 3*, respectively. As a result, a delay matrix with 768 time rows was obtained by dividing total data period (32 days) into one hour intervals (32 days × 24 h a day = 768 intervals in total). Finally, the delay matrix of the stop-time interval pairs (see Table 2 for a snapshot) was fed into the learning process in order to identify the delay interactions. The rows of the delay matrix were treated as individual observations which indicate the delay status of the network at the given time interval.

The analysis was conducted for two peak travel periods based on LIRR peak hour pricing; AM peak (trains arriving at New York City terminals between 6:00 AM and 10:00 AM), PM peak (trains departing from New York City terminals between 4:00 PM and 8:00 PM). For the AM and PM peak analysis, the time-interval rows that correspond to weekday AM and PM peak travel were extracted. Finally, the aggregated delays were divided by the number of trains (within the focus period, e.g., PM peak) passing through each stop, in order to calculate the delay per train. Note that, due to reliance on users, the crowdsourced data do not include delay records for all periods and stops. For the missing period-stop delay values (e.g., 2117 time-stop slots of 4410 for Eastbound PM peak), we implemented the “*structural.em*” tool in “*bnlearn*” package to estimate the missing values (Scutari, 2010; Scutari and Ness, 2019). In order to predict the missing values, it is necessary to know the Bayesian network structure beforehand. However, the network structure is also needed to be estimated using the missing data; which creates the chicken or the egg problem.

The approach to learn a Bayesian network structure using incomplete data was first developed by Friedman (1997) using an extension of Expectation-Maximization algorithm to search large numbers of candidate network structures and identify the best-fitting one. Friedman (1997) also showed an example in which the proposed approach can successfully learn network structures when there

Time Interval ► 10:00 – 11:00

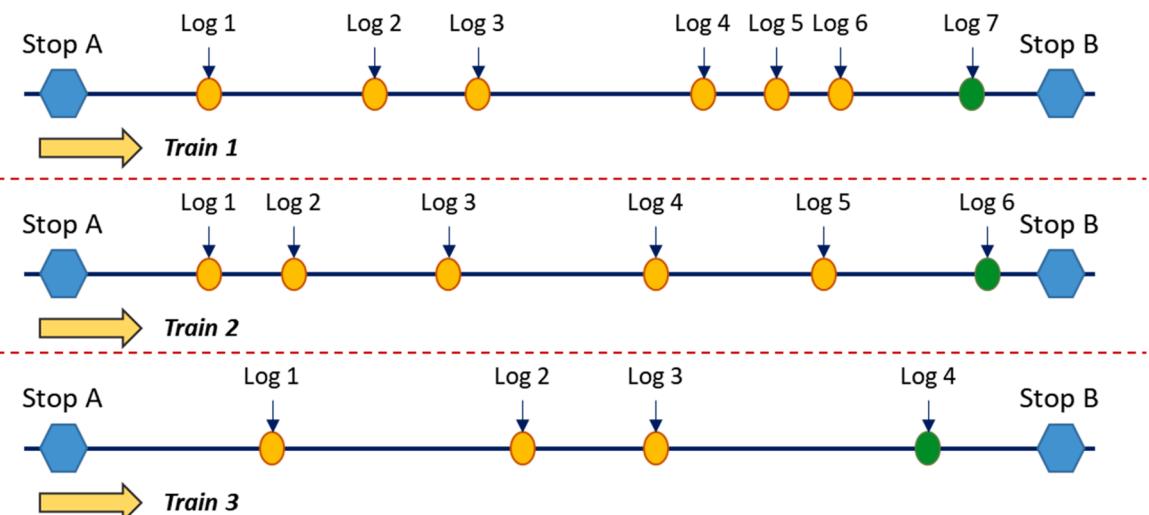


Fig. 4. An example dispersion of logs recorded by the app for the same stop-time interval.

Table 2

A snapshot of total delay data matrix of stops for trips towards West at AM peak.

Total Delays (minutes)		Stops							
Time Interval	Penn Station	.	Forest Hills	Jamaica	.	.	Free-port	Bay Shore	.
1	311	.	348	305	.	.	168	300	.
2	116	.	137	39	.	.	284	387	.
3	11	.	341	192	.	.	178	349	.
4	98	.	247	61	.	.	0	223	.
5	233	.	313	160	.	.	116	331	.
.
127	405	.	376	409	.	.	193	356	.

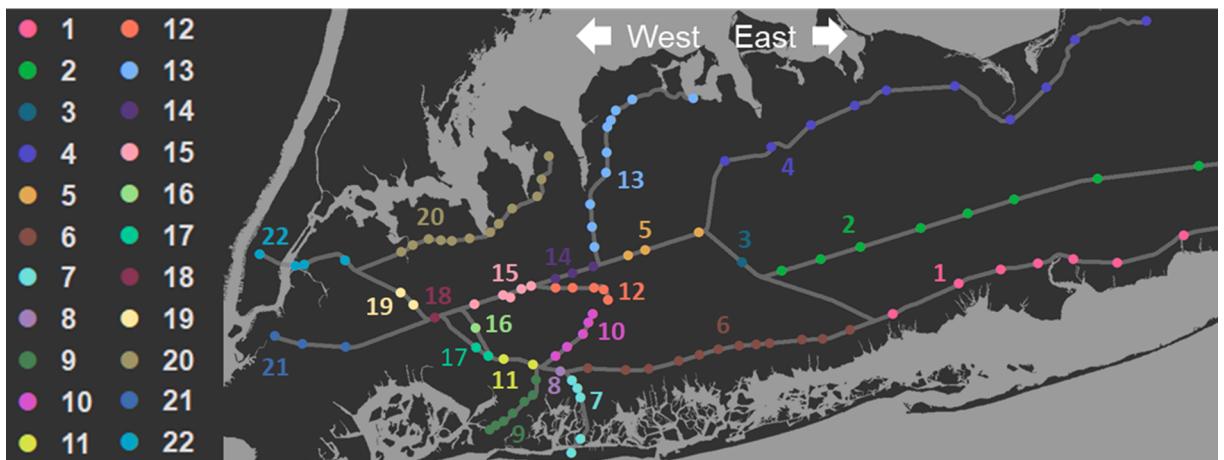
are several variables with a total of 30% missing values. Please refer to [Friedman \(1997\)](#) study for the theoretical background and the implementation of the methodology. The “structural.em” tool in “bnlearn” package ([Scutari and Ness, 2019](#)) is based on the approach proposed by [Friedman \(1997\)](#). Estimation of the missing values and the Bayesian networks repeated for 100 times (i.e., bootstrapping process) for each analysis group. Only the frequently identified (more than 80% of the repeated analysis) dependencies between nodes were included in the final graphs.

4.2.2. Network topology

The structure of the Bayesian network (BN) impacts the validity and interpretation of outcomes; thus it is one of the important components of Bayesian network learning. There are two main approaches to define the structure of a BN: 1) building the structure prior to feeding data and identifying the interaction among nodes, and 2) learning the network structure directly from the data without any prior constraints. Essentially, the first approach is possible if there is expert knowledge regarding the structure of the network; whereas the second approach is more common (and less biased) since it is generally difficult to have prior knowledge on the network interactions.

The second approach was adopted in this study to identify network structures without any prior construction. However, due to the physical connections in the rail network, delay dependencies between stops are dictated by the physical network and the scheduled trips. For instance, there are no trains from Penn Station ([Fig. 5 – Group 22](#)) to the Atlantic Terminal ([Fig. 5 – Group 21](#)) and they are connected through Jamaica Station ([Fig. 5 – Group 18](#)). Thus, delay dependencies among Penn Station and Atlantic Terminal can only emerge through Jamaica Station. In order to reflect those dependencies, “blacklisting” constraints were introduced, i.e., enforcing the BN learning to ignore certain edges (i.e., interactions) ([Scutari, 2010](#)). Blacklisting helps avoid assigning node (transit system stops) interactions that are not physically possible. Accordingly, the topology of the physical rail network is included in the BN analysis in terms of the route-based adjacency. Arguably, the adopted approach might be described as a “hybrid approach” which predominantly learns the network structure from the data.

The route-based adjacency was developed by utilizing the flow scheme of LIRR trains. For this purpose, the stops were grouped based on the “segments” separated by the bifurcations of the rail network. Note that, there are a total of 13 bifurcations bringing about 12 stop groups (one group between two bifurcations) and 11 terminal branches bringing about 10 stop groups; which sum up to 22 stop groups ([Fig. 5](#)). Two stop groups were considered adjacent if a train can travel to/from one of these stops from/to the other one. For instance, there are scheduled transit trips from *Group 4* stops to *Group 22* stops ([MTA, 2019a](#)), whereas there is no such trip between

**Fig. 5.** Assigned stop groups.

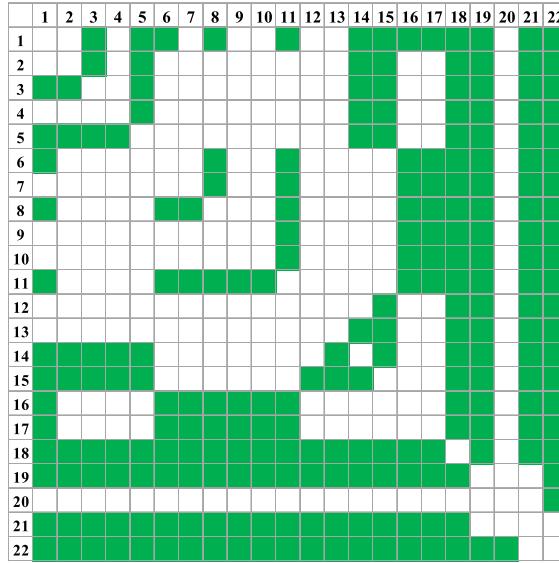


Fig. 6. Stop adjacency matrix created based on the assigned group numbers. A green colored box indicates the presence of adjacency between two groups. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Group 4 stops and *Group 20* stops. The adjacency matrix based on group numbers is shown in Fig. 6. Adjacency structure ensures that the delays of two nonadjacent stops can only be in relation through a stop which is adjacent to both these stops.

4.3. Network performance metrics

The Bayesian networks (BNs) extract the dependencies between the stops in terms of experienced delays; however, the Bayesian networks are not necessarily simple to understand due to the high number of nodes and links in the rail network. Within the delay interaction and delay propagation context, it is possible to distinguish the roles played by the stops into two categories based on their impacts within the network: 1) A stop can be the source of the delay that propagate in the network, or 2) it might be influenced by the delays propagating within the network. This can also be allegorically described as the “prey-predator relationship”. Such “source of delay” vs. “influenced by delay” scheme can also be considered as the concept of primary delay vs. secondary delay in the rail operation literature (Huang et al., 2020; Lee et al., 2016). We proposed that these two aspects of network-wide impacts are different, i.e., the first one is the inducer of delays, whereas the latter is the susceptible one. Therefore, two metrics namely “Inducer” and “Susceptible” were developed to quantify the delay dependencies in the rail network. In other words, the metrics quantify the impacts of the stops on network-wide delays as well as their vulnerabilities against delays created by other stops within the rail network. From this perspective, the inducer metric identifies and quantifies the impacts of stops on the overall delays experienced in the network. The susceptibility metric identifies and quantifies vulnerabilities of the stops against the delay impacts induced by other stops.

For the delay-*inducer* score, the selected stop is assigned its median delay (observed in the data) and all other stops are given zero delay. Then, using the Bayesian network, the delays at all other stops are estimated one by one with the given initial conditions (i.e., median delay at the inducer stop and zero delays at all other stops). Then, the estimated delays are multiplied with the daily number of trips at the affected stops (based on the posted LIRR trip schedule) and the delay-*inducer* score is calculated by averaging all values (Equation (2)).

For delay-*susceptible* score, all stops other than the selected stop are assigned their respective median delays (observed in the data) one by one. Then, the delay at the susceptible stop is estimated using the developed Bayesian network given those initial conditions. Finally, calculated average delay is multiplied with the daily trip number of at the susceptible stop (Equation (3)). The resultant scores are delay minutes per stop, i.e., the average delay induced at (inducer) or being influenced by (susceptible) each stop in the identified Bayesian network.

$$\text{Inducer}_s = \sum_{j \neq s}^n \{E(X_j | X_s = x_s, Y = 0)^* V_j\} / n \quad \text{where } \forall X_{i \neq s} \in Y \quad i = 1, 2, \dots, n \quad (2)$$

$$\text{Susceptible}_s = \sum_{j \neq s}^n \{E(X_s | X_j = x_j, Y = 0)^* V_s\} / n \quad \text{where } \forall X_{i \neq j} \in Y \quad i = 1, 2, \dots, n \quad (3)$$

where X_i is the delay of stop i , V_i is the daily number of trips to/from stop i , n is the total number of stops, x is the median delay value of

variable x . $E(X|Y = y)$ is the conditional expectation of X given Y is equal to y , and $E(X|Y = y) = \int_{-\infty}^{\infty} x * f_{X|Y}(x|y) dx$ where $f_{X|Y}(x|y)$ is the joint probability density function found by using the developed Bayesian network.

In order to estimate delays used in the metrics, the *predict* function provided in the “bnlearn” package was used along with *bayes-lw* method which use the complete Bayesian network (not only parent nodes) to calculate the expected values by averaging the likelihood weighting simulations (Scutari, 2010; Scutari and Ness, 2019). Since this is a stochastic process, analysis was repeated a multiple times (100 repetition) and the mean conditional expectations of delays were found. As an example, Table 3 shows both how the delay at Jamaica Station affects the delays at other stops, and how the delays at other stops affect Jamaica Station. For instance, “Delay by Jamaica” rows show the delays induced by Jamaica station at other stops when the median delay at Jamaica (34.89 min) is fed into the BN. “Delay at Jamaica” rows illustrate the delay susceptibility of Jamaica due to delays at other LIRR stations.

Note that formulated metrics are not directly connected with the BN modeling. The formulated metrics utilize the delay interactions between the stations and the delay estimations. These necessary inputs can be provided by models/approaches other than BNs. Nevertheless, the BN modeling is the most appropriate approach to obtain the delay interactions in the rail network; hence it is anticipated that the formulated metrics are good companions to the adopted approach, i.e., BN learning.

5. Results and discussions

Fig. 7 shows the statistically significant Bayesian networks for peak hour travel periods at LIRR (westbound AM peak and eastbound PM). The graphs were plotted to reflect the topology of the actual rail network by positioning the stops according to their geographical locations in the rail network. The distances between stops in the graphs were scaled to represent the actual distances in the physical rail network. The widths of the links represent the strength of the interactions between the railway stops. In Fig. 7, the sizes of the nodes (circles) indicate the number of trips at given direction and time. For instance, the biggest node in the eastbound PM peak graph is Jamaica Station because Jamaica has the highest number of trips. Jamaica Station also serves the highest number lines and experiences the highest number of transfers (Abt SRBI, 2016). The colors of the nodes demonstrate the stop groups that were shown in Fig. 5. Note that Bayesian networks do not specifically highlight the stops that experience large delays unless these delays have dependencies with the delays at other stops.

The goodness of fit of the identified BNs was investigated through residuals of the estimated delays at stops. For this purpose, a stop and time interval was randomly chosen and the delay at that stop was estimated based on the BNs. This process repeated 10,000 times and residuals were calculated and plotted (Fig. 8). The residual error histogram shows that the majority of the residuals are clustered around 0 while there are very few estimates with an error larger than 10 min. Hence, the distribution of residual errors imply that the BNs accurately predict transit delays at stops.

5.1. Network topology and delay propagation

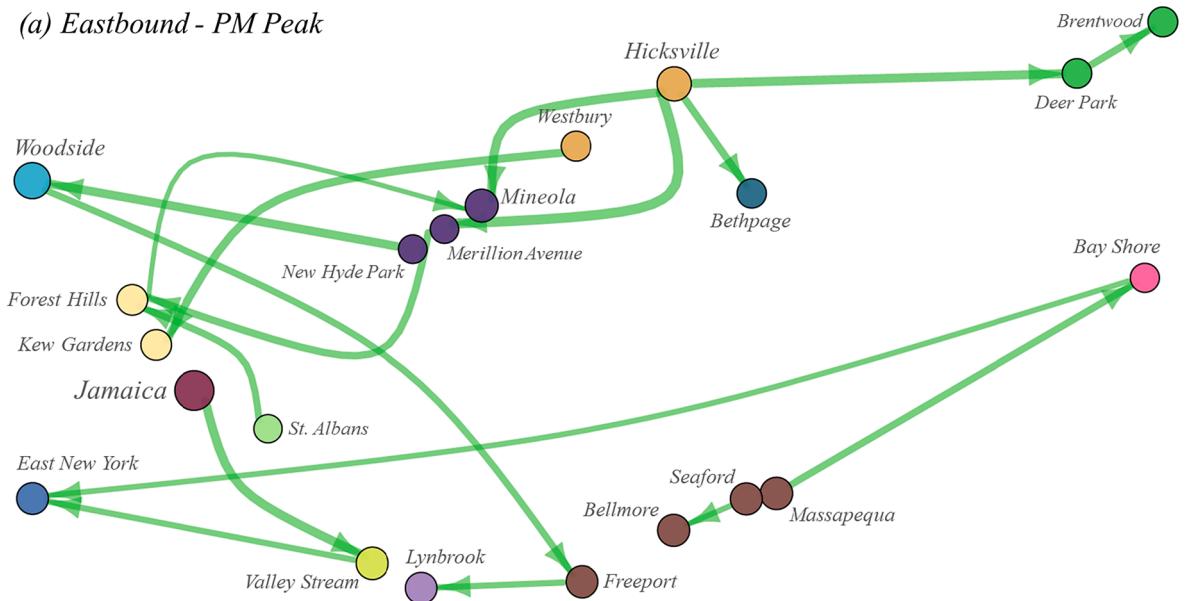
One can expect that a delay in a certain stop will cause delays further along the stops at the direction of travel, i.e., forward propagation. Such thinking can be valid for a single line with no intersection with other rail service lines. However, for most transit networks, multiple rail lines utilize the same tracks and/or same stops; hence the delay at a certain stop may prevent another train from a merging line to make its scheduled stop. Fig. 7 reflects these network considerations, and shows that delay can propagate in the network both backwards and forwards for both travel directions. For instance, for eastbound PM peak trips, Hicksville affects two stops on the East (Bethpage and Deer Park). Deer Park further affects the Brentwood stop, implying a forward propagating pattern. On the other hand, Hicksville also affects a stop on the West (Mineola) implying a backward propagation with respect to the travel direction. Similarly for AM peak (i.e., westbound travel), Jamaica affects the stops both on the West (Woodside) and the East (Mineola and Hicksville). Penn Station (westernmost terminal) affects Jamaica, indicating that the delays propagate in the reverse direction, i.e., a shockwave pattern.

Table 3

Delays (minutes) induced by Jamaica (median delay: 34.89 min) at other stops and delays which Jamaica is susceptible due to delays at other stops for eastbound PM peak trips.

Stop	Kew Gardens	St. Albans	East New York	New Hyde Park	Merillon Avenue	Carle Place
Median Delay per Train	7.60	9.26	8.44	16.92	11.44	8.05
Delay by Jamaica	4.84	6.29	2.62	10.97	1.31	4.45
Delay at Jamaica	0	0	0	0.4	0	0
Stop	Hicksville	Huntington	Bethpage	Brentwood	Valley Stream	Rockville Centre
Median Delay per Train	31.01	17.18	16.34	5.31	9.12	11.60
Delay by Jamaica	0	14.55	11.78	5.20	7.54	4.93
Delay at Jamaica	0	4.55	0	6.96	0	8.04
Stop	Lynbrook	Freeport	Bellmore	Seaford	Massapequa	Babylon
Median Delay per Train	10.35	8.20	4.69	10.24	9.77	14.48
Delay by Jamaica	4.00	4.28	1.83	9.87	7.86	16.16
Delay at Jamaica	0	0	0	0	0	0

(a) Eastbound - PM Peak



(b) Westbound - AM Peak

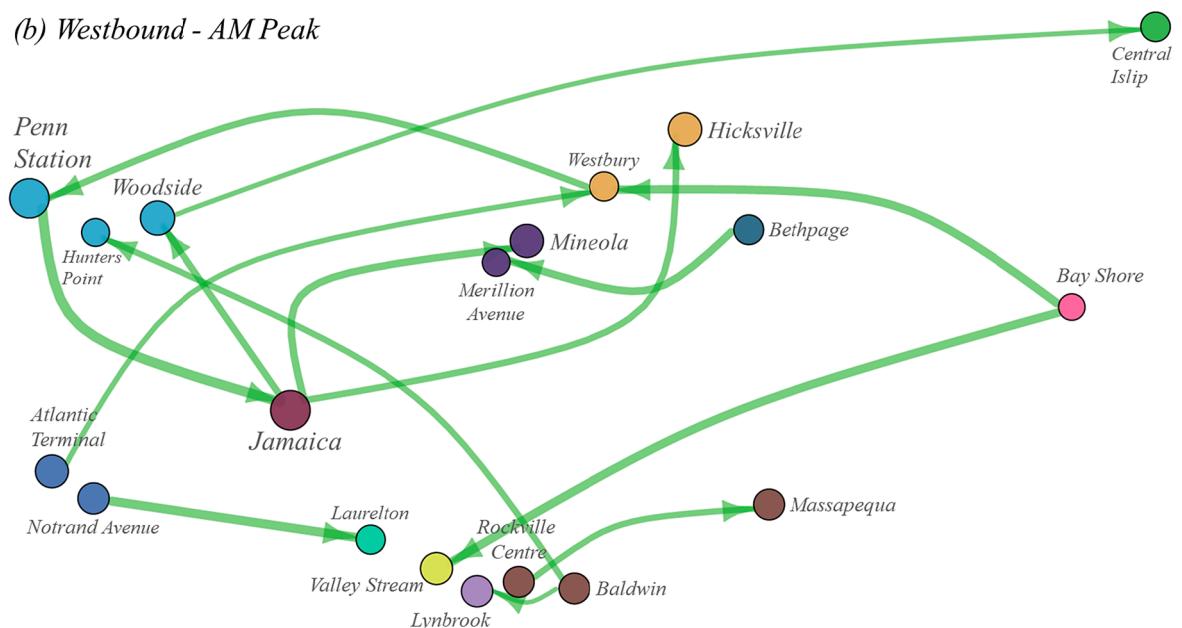


Fig. 7. Bayesian network structures for stop delays: a) Eastbound PM peak trips; b) Westbound AM peak trips. Figure illustrates statistically significant interactions between the railway stops and the widths of the links represent the strength of these interactions.

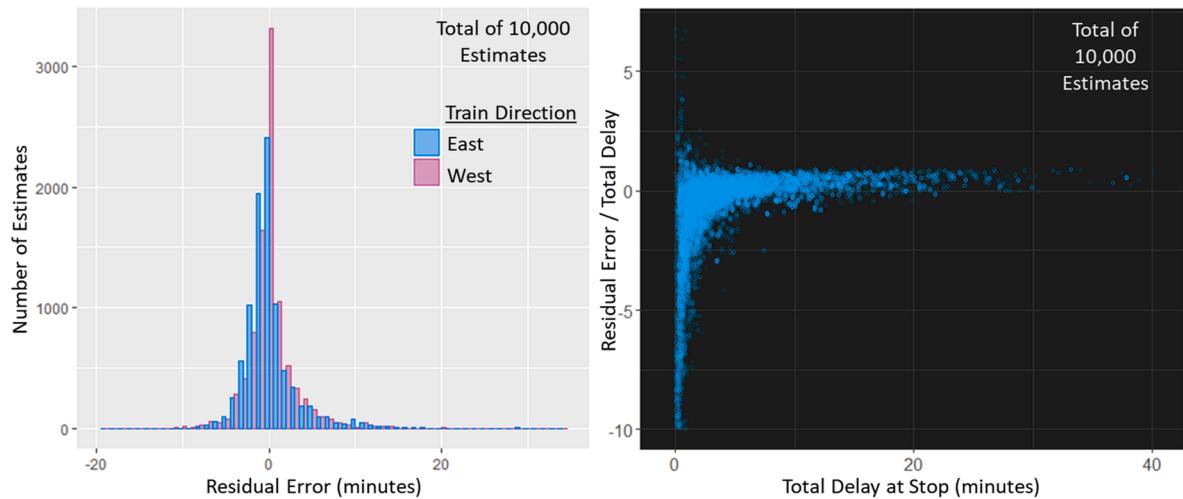


Fig. 8. Residuals histogram and error rate plot with respect to stop delays.

These varying dependencies arise due to the topology of the network. For instance, AM peak trains congest Penn Station, causing a bottleneck that prevents trains to move forward towards Jamaica (the junction) stop. Then the delay propagation continues over Jamaica to other stops on the West (Merillon Avenue and Hicksville). Similarly for PM peak, Hicksville creates a bottleneck preventing other trains to travel eastward and causing delays at the stops at the west of Hicksville (e.g., Mineola, Merillon Avenue). This pattern can be observed for other dependency couples such as Westbury – Kew Gardens, Merillon Avenue – Forest Hills, and New Hyde Park – Woodside. BN graphs consistently include stops that are located at the bifurcations of the physical rail network, such as Jamaica, Mineola, Bethpage, and Hicksville. These stops are also the ones that experience very high volumes of passenger transfers between lines ([Abt SRBI, 2016](#)).

Table 4
Metric Scores.

Metric Scores		East - PM Peak		West – AM Peak	
LIRR Branch	Stop	Ind.	Sus.	Ind.	Sus.
City Terminal Zone	Penn Station	~	~	6.5	2.1
	Woodside	5.0	0.1	5.9	5.2
	Hunterspoint Ave.	~	~	5.9	7.0
	Forest Hills	5.2	0.1	~	~
	Kew Gardens	4.9	4.3	~	~
	Jamaica	5.1	2.4	7.0	0.2
	St. Albans	4.5	8.0	~	~
	Atlantic Terminal	~	~	5.2	22.9
	Nostrand Avenue	~	~	5.9	2.8
Port Jefferson	East New York	5.1	3.2	~	~
	New Hyde Park	4.7	10.1	~	~
	Merillon Avenue	5.0	3.4	6.1	1.3
	Mineola	4.9	3.3	5.8	10.1
	Westbury	5.0	0.2	6.3	0.0
Ronkonkoma	Hicksville	5.6	0.7	5.4	15.3
	Bethpage	4.5	9.7	6.0	5.2
	Deer Park	5.0	0.4	~	~
	Brentwood	4.7	5.0	~	~
Babylon	Central Islip	~	~	6.2	2.2
	Laurelton	~	~	6.3	3.8
	Valley Stream	5.2	4.0	6.0	3.9
	Lynbrook	4.8	3.0	5.5	13.3
	Rockville Centre	~	~	6.2	2.6
	Baldwin	~	~	5.9	4.8
	Freeport	4.8	5.0	~	~
	Bellmore	5.0	3.2	~	~
	Seaford	4.5	8.9	~	~
	Massapequa	4.6	7.6	6.2	3.0
	Bay Shore	5.0	1.1	5.8	3.9

The above findings are arguably intuitive and expected based on literature, i.e., the stops at bifurcations have multiple merging and branching tracks, and they are more likely to be the bottlenecks which influence (or are influenced by) other stops. The agreement with the literature serves as the initial “reality check” for the proposed methodological approach. Nevertheless, the BNs allow drawing further conclusions and assigning magnitudes for the interactions. For instance, in order to represent these results in a tangible way, the interactions between stops can be assigned categories that reflect the nature of the dependencies, e.g., propagating/causing delay on other stops vs. affected by other stops. For instance, Fig. 7a shows that Hicksville stop affects Deer Park which affects Brentwood. Therefore, the Brentwood stop is also indirectly affected by Hicksville as well.

The graphs demonstrate that some stops such as “Jamaica” consistently affect others, while others such as Lynbrook are mostly affected by other stops (Fig. 7). These dependencies make “Jamaica” an *inducer* stop since it has influence on many stops. Lynbrook, on the other hand, is a *susceptible* stop, i.e., the delays at other stops elicit delays at Lynbrook. This means that a susceptible stop might experience delays even though there is no delay-causing issue associated with the stop itself. As presented in the next section, these dependencies are quantified using the proposed inducer and susceptible metrics, in order to provide a tangible index that can be used for decision-making and performance evaluation.

Note that, the Bayesian networks concern with the systematic and consistent issues and delay propagation in the network, rather than the effects of dynamic changes in the system. From this perspective, they are not instantaneous delay interactions (or states); but rather the “steady state” of delay interactions observed in the rail network.

5.2. Metric results on LIRR network

The calculated “Inducer” and “Susceptible” scores for LIRR are presented in Table 4. The scores are indicators of network-wide delays in terms of “minutes per stop” in the identified Bayesian network. The metric scores for peak travel directions are also illustrated in Fig. 9. The inducer and susceptible scores are labeled with red and yellow respectively, and the size of the circle represents the magnitude of the score. Note that, for the sake of clarity, Fig. 9 does not include all the scores in Table 4 but shows the stops whose scores are more than median score for each peak period.

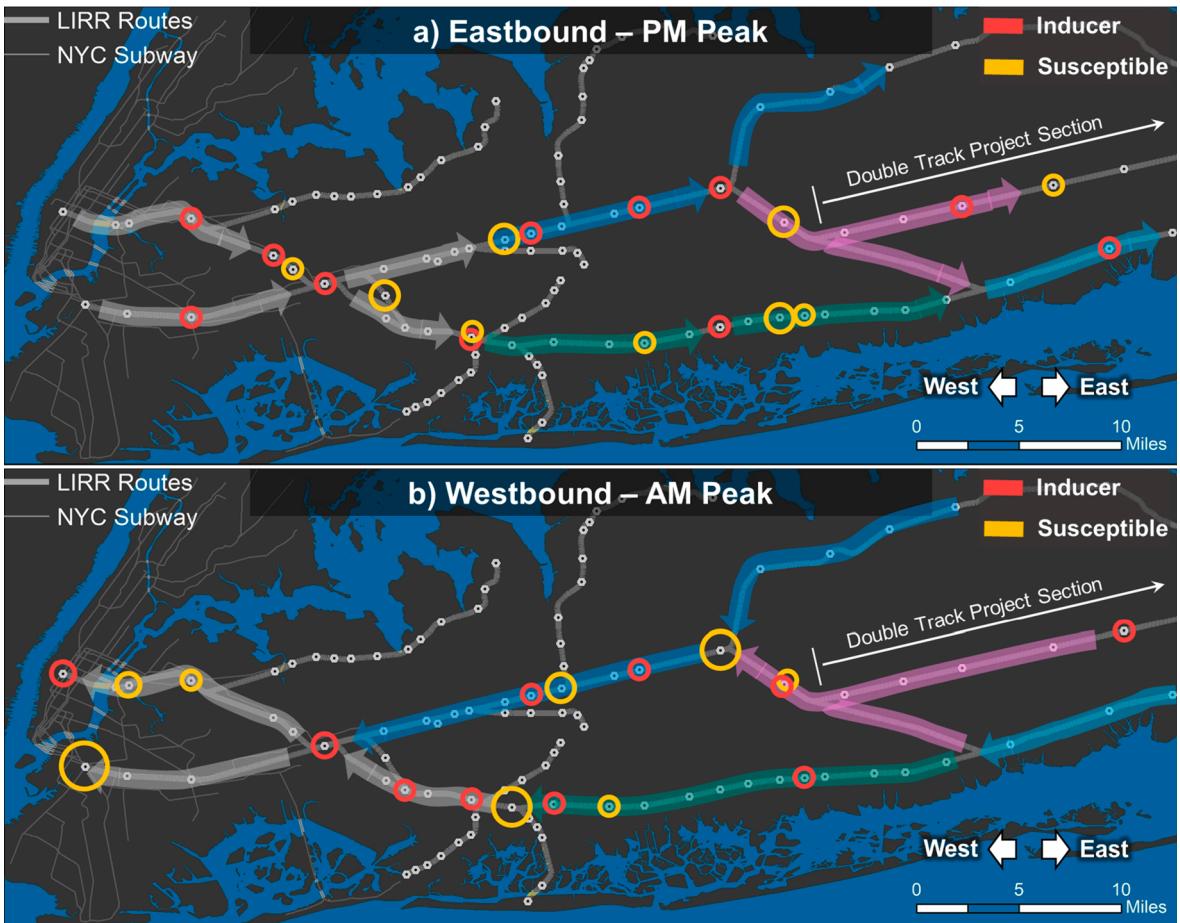


Fig. 9. Inducer and Susceptible stops based on Bayesian networks: a) Eastbound PM peak, and b) westbound AM peak.

Overall, the metrics quantify the network-wide impacts of stops which are represented graphically by the BNs, thus helping differentiate their impacts. For instance, the effect of Jamaica (the main junction node at LIRR) on delays during AM peak is larger than other stops (7 min per stop), followed by Penn Station (6.5 min per stop). Note that the stops close to Penn Station and Jamaica (Atlantic Terminal, Woodside, and Hunterspoint Avenue) are highly susceptible stops during AM peak. Bayesian network (Fig. 7) shows that Penn Station affects Jamaica which affects both Woodside and Hunterspoint Avenue. These dependencies can be attributed to the high congestion as the trains enter Penn Station, which creates a ripple effect of delays at upstream. These identified issues are already recognized by Metropolitan Transit Agency (MTA). As one of the mega projects at the NYC metropolitan area, “East Side Access” project’s objective is to provide new access from Woodside to Grand Central Terminal (at Manhattan, New York) which will be used as a new LIRR terminal, and relieve the congestion at Penn Station. In other words, the findings are in line with the already identified issues. Thus, the metrics can be used to prioritize the improvements/investments at the stops and/or lines other than the ones that are already underway.

For PM peak, Hicksville has the highest inducer score followed by Valley Stream. Both stops are at critical locations for PM peak trains, i.e., they connect main lines coming from City Terminal Zones to the LIRR branches (Port Jefferson, Ronkonkoma, and Babylon) passing through highly populated residential zones. In terms of delay susceptibility, New Hyde Park (Port Jefferson Branch) has the highest susceptible score during PM peak, followed by Bethpage (Ronkonkoma Branch). However, although the Babylon Branch does not include a stop with high susceptible score, a branch-wide susceptibility is observed. As a matter of fact, the delay susceptibility at Ronkonkoma Branch has already been addressed. In September 2018 (about 8 months after the data period that was used in this paper), a double track project stretching 13 miles between Farmingdale and Ronkonkoma stops at the Ronkonkoma branch of the LIRR was completed (A Modern LI, 2019a). Completion of the double-track project also enables us to use this case as a “natural experiment” and perform a before-after study to illustrate the use of the developed metrics.

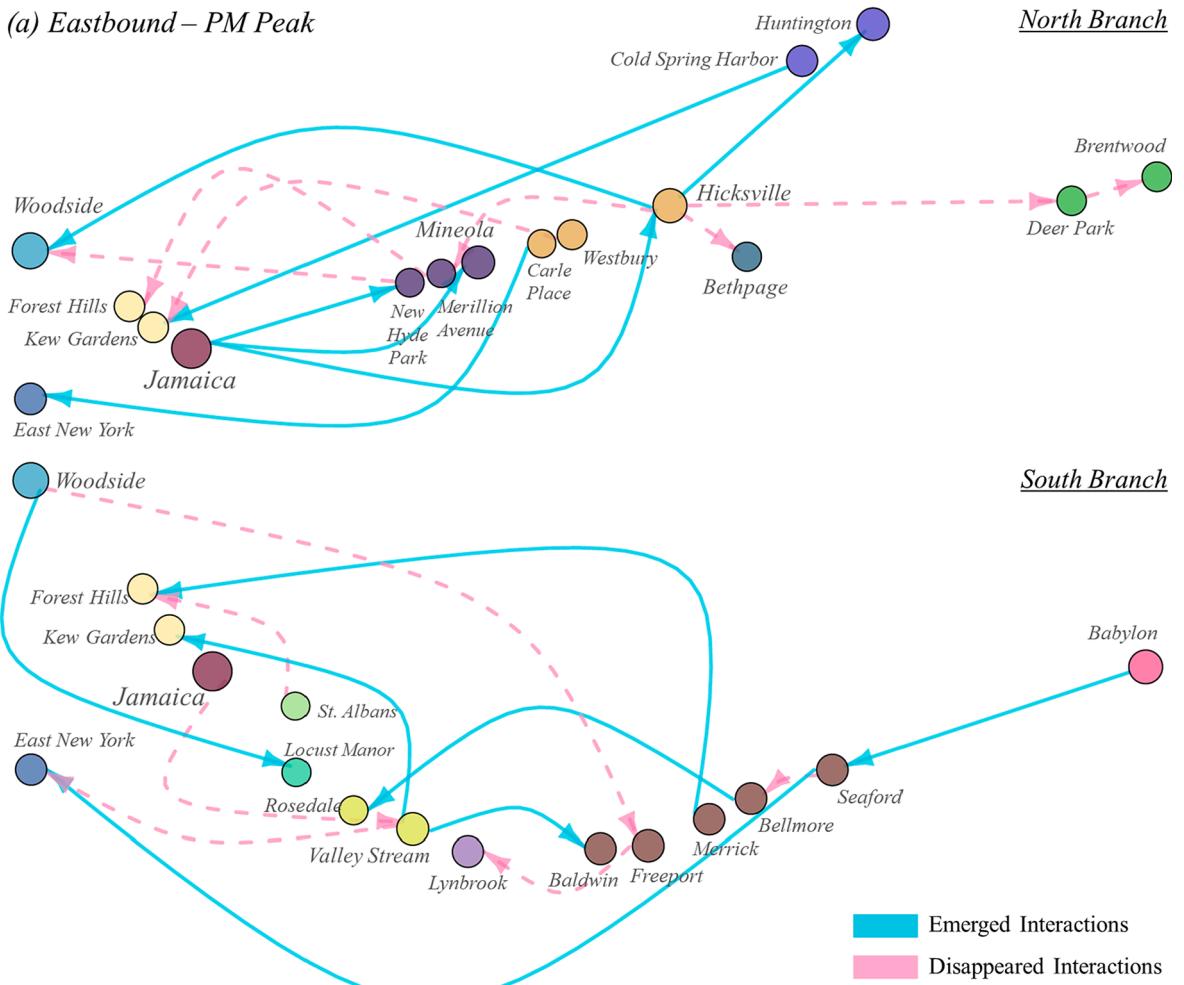


Fig. 10. Emerged and dissolved dependencies identified by comparing “post-project” and “pre-project” datasets a) Eastbound PM peak, and b) westbound AM peak.

(b) Westbound – AM Peak

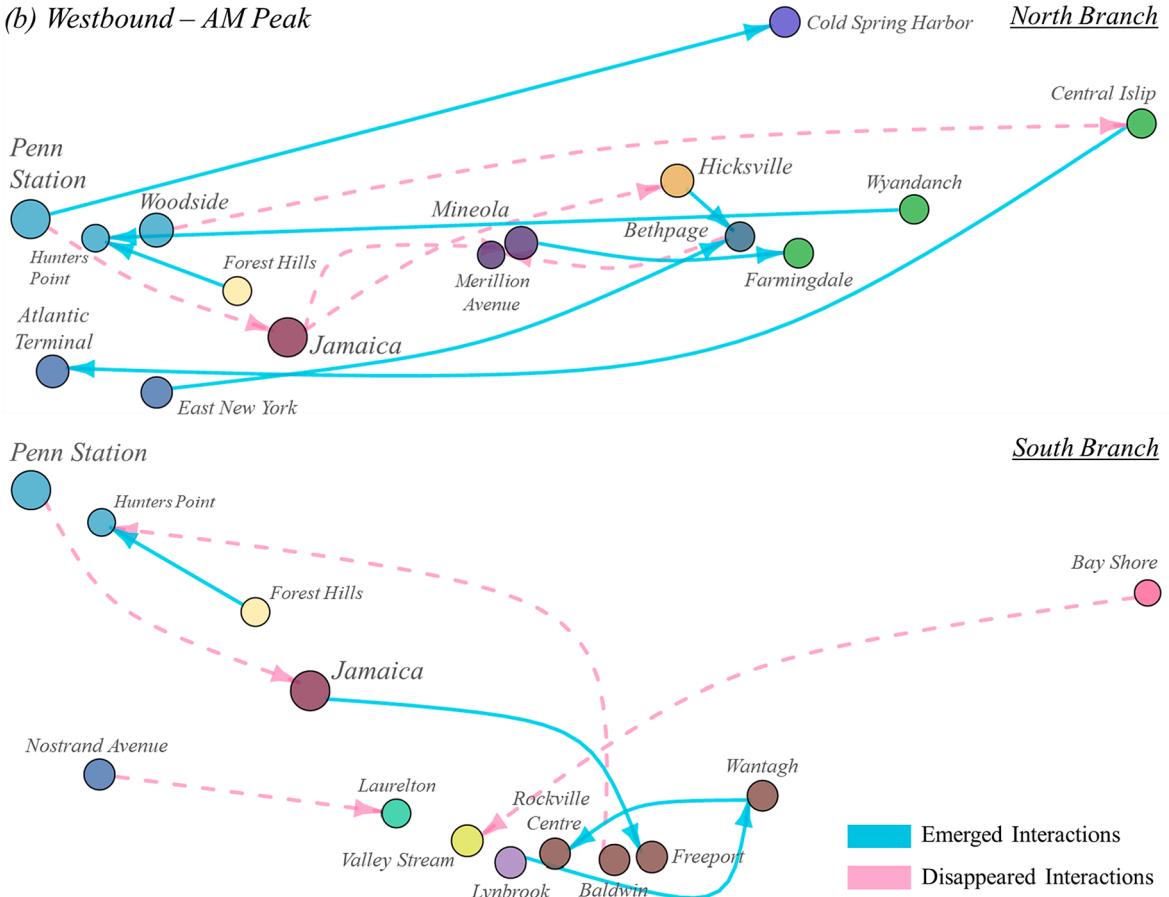


Fig. 10. (continued).

5.3. Case study: before and after double-track project at Ronkonkoma branch

The double-track project aimed to improve the reliability and on-time performance of the Ronkonkoma branch in particular, and LIRR network in general. In order to identify the impacts of the LIRR double-track project, a more recent dataset covering between 1st of March and 1st of July 2019 was collected through *onTime* app. The Bayesian networks and metrics were calculated using the new dataset, and compared with the pre-project findings. To observe the differences between “post-project” and “pre-project” Bayesian networks, “emerged” and “disappeared” dependencies were identified as shown in Fig. 10. In the figure, cyan colored solid lines show newly emerged dependencies after the project (i.e., post-project) whereas pink colored dashed lines show disappeared dependencies based on pre-project findings. Note that, if a dependency was identified in both dataset (i.e., no variation observed), then that dependency was not shown in graphs.

As shown in Fig. 10, the impacts of the double-track project are captured by the Bayesian network analysis. For instance, the dependencies between Hicksville Deer Park and Brentwood disappeared at eastbound PM peak trips (Fig. 10a – North Branch). The dependencies between Hicksville-Mineola and Hicksville-Bethpage also disappeared, possibly due to congestion relief at the Hicksville stop after track capacity increased on the Ronkonkoma branch. The metric scores indicate decreased susceptibility to delays at Ronkonkoma Branch (Table 5). However, the reduction in delay susceptibility at Ronkonkoma Branch also introduced new dependencies within the network (not confined to Ronkonkoma Branch), such as Woodside's emerging dependency on Hicksville.

As an example of how the BN approach and metrics can be utilized, PM peak graph (Fig. 10a) illustrates two major issues: 1) dependencies at Port Jefferson branch (Cold Spring Harbor and Huntington) and 2) dependencies between Jamaica stop and stops at the main line (New Hyde Park, Mineola, and Hicksville) connecting City Terminal Zone to Ronkonkoma and Port Jefferson branches. In pre-project analysis, these dependencies were not as significant as the disappeared Ronkonkoma dependencies, therefore they did not appear in the significant network. After Ronkonkoma branch dependencies disappeared (due to track improvement), BN results indicate that dependencies of aforementioned groups gained importance in the LIRR network. In other words, the metrics revealed the potential candidate stops/branches for improvement for the next investment cycle. Case in point, there is an ongoing third track expansion project stretching 9.8 miles between Floral Park and Hicksville (A Modern LI, 2019b). This stretch coincides exactly with the emerged dependencies after the double-track project, i.e., Jamaica to New Hyde Park, Mineola and Hicksville.

Table 5

Change in metric scores after project implementation.

Score Change		East - PM Peak		West – AM Peak	
LIRR Branch	Stop	Ind.	Sus.	Ind.	Sus.
City Terminal Zone	Penn Station	~	~	-1.33 ▼	-0.86 ▼
	Woodside	0.41 ▲	0.69 ▲	-0.53 ▼	-2.22 ▼
	Hunterspoint Ave.	~	~	-0.93 ▼	-3.35 ▼
	Forest Hills	-0.02 ▼	5.89 ▲	~	~
	Kew Gardens	0.55 ▲	-0.17 ▼	~	~
	Jamaica	1.37 ▲	2.85 ▲	-2.29 ▼	22.62 ▲
	Atlantic Terminal	~	~	-0.24 ▼	-16.33 ▼
	Nostrand Avenue	~	~	-1.23 ▼	3.97 ▲
Port Jefferson	East New York	-0.07 ▼	3.82 ▲	~	~
	New Hyde Park	0.65 ▲	0.81 ▲	~	~
	Mineola	1.37 ▲	0.35 ▲	-0.83 ▼	-1.05 ▼
	Westbury	0.06 ▲	8.81 ▲	~	~
	Hicksville	1.36 ▲	-0.30 ▼	-0.97 ▼	-1.02 ▼
	Bethpage	0.99 ▲	-5.27 ▼	-0.72 ▼	-5.15 ▼
	Central Islip	~	~	-1.11 ▼	-1.45 ▼
	Valley Stream	0.27 ▲	-2.80 ▼	~	~
Babylon	Lynbrook	~	~	-0.59 ▼	-9.15 ▼
	Rockville Centre	~	~	-1.33 ▼	0.41 ▲
	Bellmore	0.83 ▲	-2.92 ▼	~	~
	Seaford	1.12 ▲	-8.77 ▼	~	~
	Massapequa	~	~	-1.68 ▼	5.10 ▲
	Bay Shore	0.45 ▲	1.33 ▲	~	~

Notes: blue color, minus sign “-”, and ▼ indicate a decrease in the metric score, whereas red color and ▲ indicate an increase in the metric score. “~” indicate that there is no statistically significant change identified in the metric score.

The delay impacts of improvements or changes (in infrastructure and/or operation) on the transit network can be diffusive and unprecedented rather than isolated and deterministic. These examples demonstrate the applicability of the Bayesian network learning approach and metrics that are formulated to identify network-wide problems of transit systems. These capabilities can help guide strategic decision making and policy decisions regarding transit network investments.

6. Conclusions

This paper provides a network model that can identify the relationship between delay patterns at transit stops, and suggest metrics to quantify the delay dependencies between the transit network stops. For this purpose, a Bayesian network learning approach was utilized and two novel network metrics, namely *Inducer* and *Susceptible* were developed. To implement the proposed approaches in a case study, the delays at Long Island Rail Road (LIRR) commuter rail system in the State of New York were utilized. These delays were gathered through crowdsourced transit app (*onTime*) that provides information about real-time train arrivals.

In order to illustrate the use of the BN model and the metrics, before and after a track expansion project at Ronkonkoma Branch at LIRR network were used as a natural experiment case study. The before and after study of adjusted physical networks provided important observations on the applicability of BN Learning approach and validity of the proposed network metrics. The results indicate that the proposed Bayesian network learning methodology and formulated metrics are capable of identifying network-wide problems of transit systems, delay ramifications, and sections where there are backward and forward delay propagating patterns. Results also illustrate that the proposed metrics do not only identify the problem but also quantify the magnitude of the problem, which is an important aspect from operational and improvement perspectives.

Overall, the contributions of this paper are twofold: 1) Bayesian network learning approach was utilized to identify the delay interactions among the stops as well as to reveal delay propagation patterns in a complex commuter rail network; and 2) two network delay metrics were developed in order to quantify and rank railroad stops which can eventually assist policy makers and practitioners in encompassing investment plans and identifying target stops for improvement. The novelty of the approach is that the BN learning is used to identify the delay interactions and propagation patterns of a complex and intricate rail network for the first time in the literature. Previous studies utilizing BN learning in the rail delay context (Conte, 2007; Corman and Kecman, 2018; Huang et al., 2020; Lessan and Fu, 2019) generally focused on the delay prediction in single-line long-distance railroads. Our study reveals the delay interactions and local sources of network-wide issues, rather than solely predicting delays on consecutive railway stops. Furthermore,

the proposed metrics are a novel contribution to the rail network delay literature, as there is a lack of studies that focus on quantifying the role of individual stops in network delay patterns. Identifying the sources of network-wide delays in commuter rail networks is as critical as predicting delays. Therefore, the use of BN learning for rail network delay and the proposed metrics contribute to the literature in this critical issue.

6.1. Limitations and future research

One caveat is that the accuracy of the model is dependent on the spatio-temporal coverage of the crowdsourced data, i.e., the more the number of user delay records for each stop at different times of the day, the better the prediction accuracy. In this study, the delays are based on the crowdsourced data that were gathered from users of *onTime* app. Therefore, delay records are not complete for the whole LIRR network, but concentrated more at the busy stops where most of the transit riders board or alight the trains. This resulted in the missing values in the delay matrix, which were estimated. Nevertheless, complete delay data are generally available at most modern transit systems; thus data availability does not present a major obstacle for future implementations of the proposed framework. On the contrary, given the illustrated good performance with missing data, BNs should provide even more accurate dependency graphs with complete data. Furthermore, identifying the delay propagation patterns particularly during events such as major crashes and breakdowns, is also very valuable. Therefore, an excellent future research direction is to understand the network-wide delays of the transit system under extreme conditions.

Another limitation is that the proposed BN learning approach does not identify the factors contributing to the delays at stops. That is, these factors are multifaceted including infrastructure features, passenger movements, volume of boarding/alighting passengers that are recurrent, and mechanical or weather conditions and structural breakdowns which are non-recurrent. For instance, a heavy snow or down tree or a locomotive malfunction can block the rail tracks. From a network point of view, such events would act as if the railway links are practically removed. These delays are different from recurrent events which do not alter the underlying rail network, but mainly affect the scheduling operations. In these respects, it would be overreaching to state that a BN model based on recurrent delays can predict external disturbances that are not accounted for within the input data. An extended model could delve into the problematic stops pointed out by the proposed approach in order to identify such stop-level issues. This kind of extension would be possible if additional data sources providing information such as passenger volumes can be acquired.

The developed inducer and susceptible metrics were calculated based on median delays experienced in the system. A potential future improvement can be expanding the metrics to be more stochastic by using delay distributions rather than median delays. Moreover, further delay dependencies can be investigated and modeled using the Markov blankets identified by the Bayesian network (Pellet and Elisseeff, 2008; Scutari, 2010).

Note that the premise and the scope of this paper is the use of Bayesian Networks for analysis of network delay patterns in rail networks. Therefore, extensive experiments were conducted to verify that the model and the findings are credible. The good prediction accuracy of the developed BN model was shown through cross-validation, and the validity and applicability of the model were established through using natural experiments as case studies. Nevertheless, the comparison of different models with the adopted approach would be helpful to benchmark these different methodologies, which would be an excellent follow-up future research.

CRediT authorship contribution statement

Mehmet Baran Ulak: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing - original draft, Visualization. **Anil Yazici:** Conceptualization, Validation, Investigation, Writing - original draft, Supervision, Project administration, Funding acquisition. **Yun Zhang:** Data curation, Writing - review & editing.

Appendix A. Supplementary material

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.trc.2020.102749>.

References

- A Modern LI, 2019a. LIRR Double Track Project Farmingdale to Ronkonkoma [WWW Document]. Complet. Proj. URL <http://www.amodernli.com/project/doubletrack/>.
- A Modern LI, 2019b. LIRR Expansion Project Floral Park to Hicksville [WWW Document]. Curr. Proj. URL <http://www.amodernli.com/project/thirdtrack/>.
- Abt SRBI, 2016. 2012-2014 LIRR Origin and Destination Report Volume I: Travel Behavior Among All LIRR Passengers. Metropolitan Transportation Authority.
- Agbelie, Bismark, Libnao, Kathleen, 2018. Unobserved heterogeneity analysis of rail transit incident delays. *Transp. Res. Part A: Policy Pract.* 117, 39–43.
- Andersson, Emma V., Peterson, Anders, Törnquist Krasemann, Johanna, 2015. Reduced railway traffic delays using a MILP approach to increase Robustness in Critical Points. *J. Rail Transp. Plann. Manage.* 5 (3), 110–127.
- Büker, Thorsten, Seybold, Bernhard, 2012. Stochastic modelling of delay propagation in large networks. *J. Rail Transp. Plann. Manage.* 2 (1-2), 34–50.
- Buntine, W., 1996. A guide to the literature on learning probabilistic networks from data. *IEEE Trans. Knowl. Data Eng.* 8, 195–210.
- Charniak, E., 1991. Bayesian networks without tears. *AI Mag.* 12, 50–63. <https://doi.org/10.1145/988164.988169>.
- Conte, C., 2007. Identifying Dependencies Among Delays. *Georg-August-Universität zu Göttingen*.
- Conte, C., Schobel, A., 2007. Identifying dependencies among delays. *Proc. IAROR*, 2007.
- Corman, Francesco, Kecman, Pavle, 2018. Stochastic prediction of train delays in real-time using Bayesian networks. *Transp. Res. Part C: Emerg. Technol.* 95, 599–615.

- Cule, B., Goethals, B., Tassenoy, S., Verboven, S., 2011. Mining train delays. *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)* 7014 LNCS, 113–124. https://doi.org/10.1007/978-3-642-24800-9_13.
- Flier, H., Gelashvili, R., Graffagnino, T., Nunkesser, M., 2009. Mining railway delay dependencies in large-scale real-world delay data. *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)* 5868 LNCS, 354–368. https://doi.org/10.1007/978-3-642-05465-5_15.
- Friedman, N., 1997. Learning belief networks in the presence of missing values and hidden variables. In: *Proceedings of the 14th International Conference on Machine Learning*, pp. 125–133.
- Gaurav, R., Srivastava, B., 2018. Estimating train delays in a large rail network using a zero shot Markov model. In: *Proceedings of 21st IEEE Conference on Intelligent Transportation Systems*, pp. 1221–1226. <https://doi.org/10.1109/ITSC.2018.8570014>.
- Ghofrani, F., He, Q., Goverde, R.M.P., Liu, X., 2018. Recent applications of big data analytics in railway transportation systems: A survey ☆. *Transp. Res. Part C 90*, 226–246. <https://doi.org/10.1016/j.trc.2018.03.010>.
- Gorman, Michael F., 2009. Statistical estimation of railroad congestion delay. *Transp. Res. Part E: Logist. Transp. Rev.* 45 (3), 446–456.
- Goverde, Rob M.P., 2010. A delay propagation algorithm for large-scale railway traffic networks. *Transp. Res. Part C: Emerg. Technol.* 18 (3), 269–287.
- Harris, Nigel G., Mjøsund, Christian S., Haugland, Hans, 2013. Improving railway performance in Norway. *J. Rail Transp. Plan. Manage.* 3 (4), 172–180.
- Huang, Ping, Lessan, Javad, Wen, Chao, Peng, Qiyuan, Fu, Liping, Li, Li, Xu, Xinyue, 2020. A Bayesian network model to predict the effects of interruptions on train operations. *Transp. Res. Part C: Emerg. Technol.* 114, 338–358.
- Lee, Wei-Hsun, Yen, Li-Hsien, Chou, Chien-Ming, 2016. A delay root cause discovery and timetable adjustment model for enhancing the punctuality of railway services. *Transp. Res. Part C: Emerg. Technol.* 73, 49–64.
- Lessan, J., Fu, L., 2019. A hybrid Bayesian network model for predicting delays in train operations. *Comput. Ind. Eng.* 127, 1214–1222. <https://doi.org/10.1016/j.cie.2018.03.017>.
- Marković, Nikola, Milinković, Sanjin, Tikhonov, Konstantin S., Schonfeld, Paul, 2015. Analyzing passenger train arrival delays with support vector regression. *Transp. Res. Part C: Emerg. Technol.* 56, 251–262.
- Metropolitan Transportation Authority, 2017. 2017 Annual Report. New York, New York.
- Milinković, Sanjin, Marković, Milan, Veskovčić, Slavko, Ivić, Miloš, Pavlović, Norbert, 2013. A fuzzy Petri net model to estimate train delays. *Simul. Model. Pract. Theory* 33, 144–157.
- MTA, 2019a. Metropolitan Transit Agency LIRR GTFS Schedule [WWW Document]. Metrop. Transit Agency. URL <https://datamine.mta.info/>.
- MTA, 2019b. Long Island Rail Road Map [WWW Document]. Metrop. Transit Agency. URL <http://web.mta.info/lIRR/Timetables/LIRRweb-Feb17.pdf>.
- Nabian, Mohammad Amin, Alemazkoor, Negin, Meidani, Hadi, 2019. Predicting near-term train schedule performance and delay using bi-level random forests. *Transp. Res. Rec.* 2673 (5), 564–573.
- Oneto, L., Buselli, I., Lulli, A., Canepa, R., Petralli, S., Anguita, D., 2020. A dynamic, interpretable, and robust hybrid data analytics system for train movements in large-scale railway networks. *Int. J. Data Sci. Anal.* <https://doi.org/10.1007/s41060-018-00171-z>.
- Oneto, Luca, Fumeo, Emanuele, Clerico, Giorgio, Canepa, Renzo, Papa, Federico, Dambra, Carlo, Mazzino, Nadia, Anguita, Davide, 2017. Dynamic delay predictions for large-scale railway networks: deep and shallow extreme learning machines tuned via thresholdout. *IEEE Trans. Syst. Man Cybern. Syst.* 47 (10), 2754–2767.
- van Oort, Niels, 2014. Incorporating service reliability in public transport design and performance requirements: International survey results and recommendations. *Res. Transp. Econ.* 48, 92–100.
- Pellet, J.-P., Elisseeff, A., 2008. Using Markov blankets for causal structure learning. *J. Mach. Learn. Res.* 9, 1295–1342.
- Scutari, M., 2010. Learning Bayesian networks with the bnlearn R package. *J. Stat. Softw.* 35, 1–22.
- Scutari, M., Ness, R., 2019. Package “bnlearn”: Bayesian Network Structure Learning, Parameter Learning and Inference. R Packag.
- Sørensen, Anette Østbø, Landmark, Andreas Dypvik, Olsson, Nils O.E., Seim, Andreas Amdahl, 2017. Method of analysis for delay propagation in a single-track network. *J. Rail Transp. Plann. Manage.* 7 (1-2), 77–97.
- Transitfeeds, 2019. OpenMobilityData [WWW Document]. OpenMobilityData. URL <https://transitfeeds.com/news/open-mobility-data>.
- Wallander, Jouni, Mäkitalo, Miika, 2012. Data mining in rail transport delay chain analysis. *IJSTL* 4 (3), 269. <https://doi.org/10.1504/IJSTL.2012.047492>.
- Wang, R., Work, D.B., 2015. Data driven approaches for passenger train delay estimation. In: *IEEE Conf. Intell. Transp. Syst. Proceedings, ITSC 2015-Octob*, pp. 535–540. doi:10.1109/ITSC.2015.94.
- Wille, A., Bühlmann, P., 2004. Tri-graph: a novel graphical model with application to genetic regulatory networks. Zurich.
- Yabuki, H., Ageishi, T., Tomii, N., 2015. Mining the cause of delays in urban railways based on association rules. In: *CASPT 2015*, pp. 1–16.
- Yaghini, M., Khoshrafter, M.M., Seyedabadi, M., 2013. Railway passenger train delay prediction via neural network model. *J. Adv. Transp.* 47, 355–368. <https://doi.org/10.1002/atr>.
- Zilko, Aurelius A., Kurowicka, Dorota, Goverde, Rob M.P., 2016. Modeling railway disruption lengths with Copula Bayesian Networks. *Transp. Res. Part C: Emerg. Technol.* 68, 350–368.