Analysis and Prediction of Disruptions in Metro Networks

Yap, Menno; Cats, Oded

DOI
10.1109/MTITS.2019.8883320

Publication date
2019

Document Version
Final published version

Published in
2019 6th International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS)

Analysis and Prediction of Disruptions in Metro Networks

Menno Yap
Department of Transport and Planning
Delft University of Technology
Delft, the Netherlands
M.D.Yap@TUDelft.nl

Oded Cats
Department of Transport and Planning
Delft University of Technology
Delft, the Netherlands
O.Cats@TUDelft.nl

***Abstract*—Public transport disruptions can result in major impacts for passengers and operator. Our study objective is to predict disruption exposure at different stations, incorporating their location-specific characteristics. Based on a 13-month incident database for the Washington metro network, we successfully develop a supervised learning model to predict the expected number of disruptions, per type, station and time of day. This supports public transport authorities and operators to prioritize what type of disruptions at what location to focus on, to potentially achieve the largest reduction in disruption exposure. Our clustering results show that start/terminal and transfer stations are most susceptible to disruptions, mainly due to operations- and vehicle-related disruptions.**

Keywords—clustering, exposure, prediction, vulnerability

I. INTRODUCTION

Disruptions in public transport (PT) can have major implications for passengers and the public transport operator. Disruptions can increase passengers' nominal travel time, due to additional waiting time, in-vehicle time or transfers. The travel time perceived by passengers potentially increases as well if crowding on remaining and alternative services increases, resulting in a more negatively perceived in-vehicle time [1], [2], [3]. Disruptions can also imply costs for the operator, due to overtime payments to personnel, possible fare reimbursement for delayed passengers, and if contractual agreements between PT operator and authority result in fines. In the long term, disruptions result in a loss of revenue if ridership levels decrease as a result of unreliability and vulnerability of the PT system [4]. It is thus important to analyse the vulnerability of PT networks, to create a better understanding of the frequency, location, duration and impact of different disruption types occurring on the network. A more accurate prediction of the occurrence and impact of disruptions supports PT authorities and operators to prioritize the locations and disruption types they need to focus on, to achieve the largest robustness benefits and to get most value for money from potential measures aimed at improving PT robustness. A more robustness PT network in turn can improve the attractiveness of a PT system and drive ridership levels.

While many scientific studies to transport vulnerability are performed over the last years, most studies focus on the *impact* of a disruption, once a disruption occurs at a certain location in the network. After several studies to road network vulnerability, e.g. [5], this topic drew the attention for public transport networks as well. For example, [6] performed a study to quantify the value of spare capacity in a PT network in terms of robustness, and [7] investigated the value of providing real-time information during disruptions on urban PT networks. A study to metro network robustness based on network topology was performed by [8], whereas [9] evaluated the robustness of railway timetables once a disruption occurs. In [10], different rerouting strategies are evaluated in response to railway disruptions. A vulnerability analysis related to the impact of partial rather than complete track closures was performed by [11]. PT vulnerability is however influenced by both the *frequency* and the *impact* of disruptions. Focusing solely on vulnerability in relation to disruption impacts, so-called conditional vulnerability, can incorrectly put the emphasis on very severe, but very rare disruptions. Incorporating how often different network locations are exposed to different disruption types can therefore shift the attention to the most vulnerable locations once both disruption exposure and impact are considered. Studies to PT disruption exposure are however limited. An important reason is often a lack of disruption log data, as data over a longer period of time is required given the relatively infrequent occurrence of disruptions. In [12] a data-driven method is developed to detect atypical events in PT networks using anomaly detection. However, this method does not explicitly provide what type of disruption at which location initiated this anomaly, making it difficult to formulate policy recommendations how to tackle PT vulnerability. In [13] and [14] a database consisting of logged disruptions on the PT network for a period of 2.5 year was used to perform a vulnerability analysis. In these studies, relatively simple predictors such as the number of trains or train-kilometres were used to translate the network-wide number of disruptions to expected disruption exposure per station or link. This implies that location-specific characteristics - such as the type of stock serving a station, the state and complexity of the infrastructure at different stations, the passenger load or the geographical area where a station is located – are not considered, while these are generally important to predict disruption exposure.

The objective of our study is to develop a generic approach to accurately predict disruption exposure at different stations of a PT network, thereby incorporating the location-specific characteristics of the different stations. We apply our approach to the Washington metro network, using a 13-month database with logged disruptions received from the Washington Metropolitan Area Transit Authority (WMATA). Our study contribution is threefold:

- Analyse disruption exposure characteristics of a PT network based on empirical disruption log data.
- Develop a prediction model to predict future disruption exposure to different disruption types, for different stations and times of the day.
- Obtain insights in the susceptibility of different types of stations to disruptions, to support prioritizing locations and type of stations to be considered for potential mitigation measures.

978-1-5386-9484-8/19/$31.00 ©2019 IEEE

II. METHODOLOGY

A. Definitions

In our study we apply the following definition of *vulnerability*, obtained by combining definitions from [15] and [16] with *robustness* being its antonym: vulnerability is the degree of susceptibility of a PT network to disruptions and the ability of PT network to cope with these disruptions. This definition highlights the two components vulnerability consists of: *exposure*, the degree to which a PT system is exposed to disruptions, and the *impact* once a disruption occurs. Moving from a network level to individual elements, we define *criticality* as the degree an individual station contributes to vulnerability. Criticality again refers to both disruption exposure and impact: it considers both *weakness*, the degree of disruption exposure for an individual station, and *importance,* the impact of disruptions occurring at a station [13]. The most critical stations thus contribute most to PT vulnerability in terms of both weakness and importance.

For a given PT network, let us define each station $s \in S$, with $|S|$ being the total number of stations in the considered network. Each disruption type is defined by d, with D indicating the total set. Each time period considered in the vulnerability analysis is indicated by $t \in T$. When we define the disruption frequency f and the disruption impact w, the expected station criticality c for different disruption types in a certain amount of time is defined in (1), measured in passenger delay time. PT network vulnerability V can then be formulated as sum of the station criticality, expressed as fraction of the total passenger travel time u as per (2). The emphasis of this study lies on predicting the expected disruption exposure for different stations $E(f_{d,t,s})$.

$$E(c_s) = \sum_{t\in T}\sum_{d\in D} E(f_{d,t,s}) * E(w_{d,t,s}) \quad (1)$$

$$V = \sum_{s\in S} E(c_s) \,/\, \sum_{t\in T}\sum_{s\in S}\sum_{s\in S} u_{t,s,s} \quad (2)$$

B. Disruption Classification

The input for our approach is incident log data. This type of data is usually available at the PT authority or operator, based on logged incident notifications from train operators, station operators, control room staff, police, and the general public. An example of this incident data is provided in Table 1. While this data usually contains some information about the nature, location and time of the incident, this data is generally not intended for vulnerability analysis purposes or to draw policy recommendations from. Instead, this is used for real-time control purposes in the control room to recover services. This also entails there is only a limited degree of consistency in the description and classification of incident notifications, as it strongly depends on manual actions from controllers whose main priority is solving the incident. As a result, reassuring the incident database is fit for our study purpose requires two data processing steps: a) defining disruptions from incidents, and b) classifying disruptions.

First, we define disruptions from incidents. The database also contains incidents which did not result into a disruption. For example, a driver not able to perform its duty due to sickness is reported in the incident database, even if a stand-by driver took over the shift without any delays. In the database we use, the minutes of train delay (delays for an individual train) and line delay (delay for the entire line) are indicated. We define a disruption as any incident where either the train delay or line delay is 2 minutes or more. Incidents with both the train and line delay being smaller than 2 minutes are considered regular variability instead.

TABLE 1. EXAMPLE INCIDENT LOG DATA

ID	Start Time	Line	Train	Stop	Type	Description
11	16-08-17 8:30	Blue	419	C07	AIRL	Air leak
12	23-08-17 9:13	Red	231	A11	PUBL	Sick customer

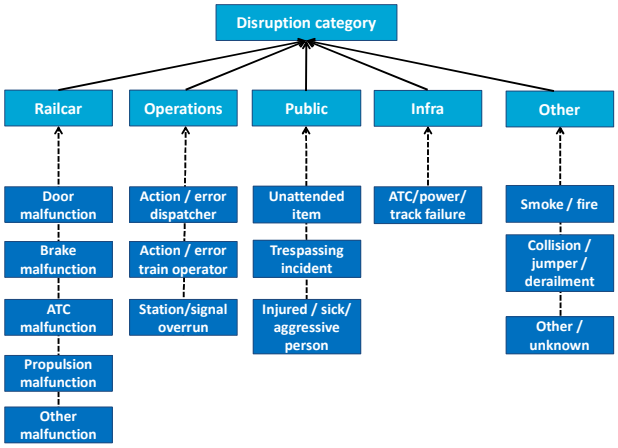

Fig. 1. Classification of disruptions.

Second, disruptions are classified into a selected number of distinctive disruption types. In the provided database, 114 different disruption types are logged. When considering the distribution over different stations $s \in S$ and time periods $t \in T$,this would result in an insufficient number of observations per station and time period to make a prediction model for. Besides, there is strong overlap between some of the disruption types, due to differences in classification by different controllers. For example, in the used database a train car motor overload is indicated by both disruption type *MOLD* ('motor overload') and *MOLF* ('flashing motor overload'). The disruption types in the database also do not always reflect the root disruption cause. As illustration, one can find an incident registered as *ONEC* ('operational necessity') with the description 'late dispatch due to door not closing'. In this case, the root cause is a door malfunctioning, resulting in an operational action from the control room. In a manual exercise, all disruptions in the database are classified based on their root cause following their description. Consequently, all disruptions are classified in 15 different distinctive types $d \in D$, which occur frequently enough to be able to develop a prediction model for. The distinguished disruption types are visualized by the dark-blue rectangles in Fig. 1. As can be seen in the light-blue rectangles, these disruption types are classified into five main categories *railcar-related, operations-related, public-related, infra-related* and *other*. The category with railcar-related disruptions for example consists of *door malfunctioning, brake malfunctioning, ATC malfunctioning, propulsion malfunctioning* and *other* disruption types. Similarly, public-related disruptions are categorized as *left unattended item, trespassing incident* and *injured/sick/aggressive passenger.*

C. Prediction of Exposure to Disruptions

We adopt a supervised learning approach to predict exposure to different disruptions $d \in D$ at stations $s \in S$ during each time period $t \in T$. This allows us to find linear and non-linear relations between presumed disruption

predictors and the exposure to disruptions. As each disruption type occurs relatively infrequent at a specific station and in a specific time period, our study objective implies predicting the occurrence of relatively rare events. For that reason we do not use $f_{d,t,s}$ as our target, as a model always predicting $f_{d,t,s}$ equals zero would still result in a low *MSE-score* and high average *F1-score* due to the overrepresentation of samples with zero disruptions, without providing any useful prediction for disruption exposure. Neither applying different weights for false positive and false negative predictions, nor applying a technique to correct the dataset imbalance such as a *Synthetic Minority Oversampling Technique* (SMOTE) did sufficiently improve the quality of disruption predictions. Instead, we therefore use the probability of each disruption type $P_{d,t,s}$ occurring within each considered time period (e.g. all AM, PM, Inter Peak and Evening periods for each day of the year) as target for the prediction. $E(f_{d,s})$ is then calculated by multiplying the predicted probabilities by the number of time periods, as formulated in (3).

$$E(f_{d,s}) = |T| * P_{d,t,s} \quad (3)$$

The number of samples in our model therefore equals $|S| * |T|$. To predict disruption probabilities we apply a classification algorithm, which calculates disruption probabilities for each $d \in D$ and then assigns each sample to one of our 15 defined disruption categories or to the category *no disruption* based on the highest probability. The shape of the target vector thus equals $(|S| * |T|, 1)$, where column values can take $|D| + 1$ different values. In our case, this value equals 0 if no disruption is predicted to occur in the considered time period, and this value ranges between 1 and 15 depending on which disruption type is predicted to occur in that time period. By binarizing this target vector, a vector with shape $(|S| * |T|, |D| + 1)$ results which consists of the predicted probabilities on each disruption type per sample.

Several location-specific station characteristics are identified as predictor in our machine learning model (see Fig. 2). *Weekday* equals 1 if the time period is during a weekday, and 0 if during the weekend. *Time of Day* considers if the time period is during the peak (7-10AM or 3-7PM: only during weekdays), daytime off-peak (weekdays: hours outside peak until 7PM; weekend: all hours until 7PM) or evening (hours after 7PM). The aim of these predictors is particularly to capture the possible influence of differences in mixture of passenger types and purposes between peak, off-peak, evenings and weekends on disruption probabilities. The predictor *Seasons* aims to capture differences in disruption probabilities for different seasons. One can think of potentially more vehicle defects due to leaves in Autumn, or more passenger-related incidents due to slippery surfaces in Winter. *Lines* refers to the different metro lines serving each station, as different stock types on different lines potentially influence especially railcar-related disruption probabilities. The possible difference in state and age of infrastructure between different lines can also play a role here. One-hot encoding is applied for the categorical predictors *Time of Day, Seasons* and *Lines*, resulting in separate binary predictors for each category. If a station is served by multiple lines, for example being part of a trunk, the binary predictor equals one for each of these lines. Two separate binary predictors *Start station* and *Transfer station* are added, being equal to one if the station is a start/terminal or a transfer station, respectively. It is expected that the occurrence of some disruptions is related to a station being a start/terminal, as problems such as a malfunctioning train or a late / sick train operator often arise here. It is hypothesized that transfer stations might be more susceptible to disruptions due to more complex infrastructure (such as switches) and large passenger transfer volumes. *Passenger volume* refers to the number of boarding plus alighting passengers for each station and time period, based on AFC data for an average day. This predictor is added to capture primarily passenger-related disruption probabilities. *Train frequency* is equal to the scheduled number of trains serving a stop during each time period and day of the week. This predictor is calculated based on timetable data, and is aimed at capturing railcar-related disruption probabilities. *Disruption frequency* refers to the total number of disruptions what occurred during a certain time period, weekday/weekend at the considered train station in the previous month, for each disruption type separately. This predictor in fact auto correlates in time with the target, and assumes the availability of disruption data of the previous month can be used to predict disruption exposure for each $t \in T$ in the next month. Effectively 15 separate predictors are used for the *Disruption Frequency* predictor, for each disruption type $d \in D$. *Passenger volume, Train frequency* and *Disruption frequency* are all scaled between 0 and 1, so all predictors use the same range.

Given our target to predict the probability of different disruption types in a certain time period at a certain station, we test two different machine learning algorithms suitable for this purpose: logistic regression and a multilayer perceptron (MLP) classifier, a class of feedforward artificial neural networks. The total dataset is split into a 80% training set and 20% testing set, applied in a randomized 5-fold cross validation. Applying a higher 10-fold cross-validation did not significantly improve prediction accuracy. Log loss, or cross entropy loss, is used as evaluation metric. This function as shown by (4) calculates the negative log-likelihood of the true label y, given the predicted probability that a sample equals this true label $\tilde{y}$.

$$-logP_{y|\tilde{y}} = -(y * log(P_{\tilde{y}}) + (1 - y) * log(1 - P_{\tilde{y}})) \quad (4)$$

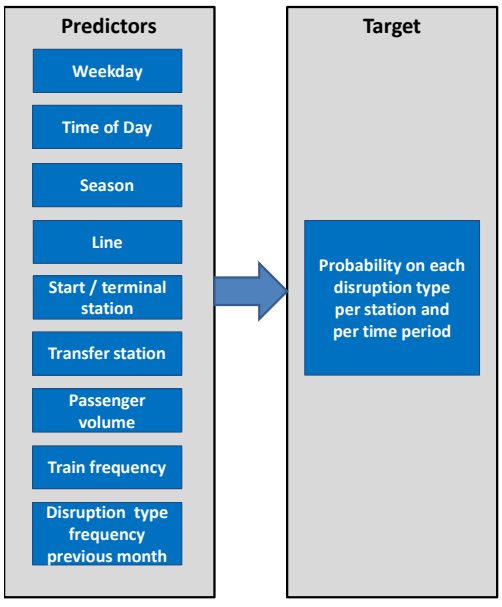

Fig. 2. Framework prediction model.

For logistic regression we perform a multiclass regression with a maximum of 100 iterations. *Sag* is used as solver method, as this is fast for relatively large datasets. For the MLP classifier one hidden layer is used. Furthermore, *adam* is used as solver method being fast for large datasets, with the number of iterations being capped at 200. A logistic sigmoid function is used as activation function for the hidden layer. The number of neurons of the hidden layer is determined by hyperparameter tuning: for all number of neurons between the number of neurons of the input layer and output layer the log loss score is calculated, thereby selecting the number of neurons for the hidden layer minimizing this value. Python is used to execute the machine learning models [17].

D. Clustering Stations Based on Disruption Exposure

Based on the predicted number of disruptions per type and station in a given time period as done in *C.*, we apply an unsupervised learning method to cluster stations based on disruption exposure. This provides insight in differences in susceptibility for different disruption types between stations, and shows clusters of stations with a similar susceptibility.

As our aim is to cluster all stations $s \in S$ without outliers, and no number of clusters k is known on beforehand, we apply hierarchical agglomerative clustering. Input for the clustering is a matrix consisting of values $E(f_{d,s})$ with shape $(|S|, |D|)$, which results from our supervised learning prediction model. The distance matrix is determined by calculating the $|D|$-dimensional Euclidean distance between all points. Ward is used as linkage criterion during the clustering, thereby minimizing the within-cluster variance. We use the cophenetic correlation coefficient to assess the degree the clustering reflects the input data. The optimal number of clusters k is, after visual inspection of the dendrogram, determined using the average silhouette coefficient. The silhouette coefficient for each sample is calculated by taking the difference between the Euclidean distance to the nearest cluster this sample is not part of, and the intra-cluster distance. This difference is then divided by the maximum value of these two. The average silhouette coefficient results if this calculation is repeated for all $|S|$ samples. k is considered optimal if the value of the average silhouette coefficient is maximized.

III. CASE STUDY

We apply our proposed methodology to the Washington D.C. metro network as case study. The Washington Metro, administrated by WMATA, consists of 6 lines indicated by different colours: the Red line (R), Green line (G), Yellow line (Y), Blue line (B), Orange line (O) and Silver line (S). At the time of consideration, 95 different metro stations are operational, thus $|S|$=95. The total length of the metro network is about 190 km. During AM and PM peak hours, the Red line runs 15 trains per hour (tph), of which every other train is a short-turning service to Silver Spring. The other lines run 7.5tph during peak hours. During daytime off-peak periods all lines run 5tph. The Blue, Orange and Silver line share a substantial part of their routes between Rosslyn and Stadium-Armory. The joint frequency on this trunk equals 22.5tph during peak hours.

A 13-month incident database for the Washington metro network is provided by WMATA, covering all 21,868 reported incidents from August 1st 2017 to August 31st 2018. When applying our disruption definition, 7,263 disruptions remain. After combining disruption registrations for multiple trains caused by the same disruption, 5,835 distinguishable disruptions remain. As one of the predictors in our model equates disruption exposure in the previous month, only the 4,935 disruptions for the 12-month period from September 1st 2017 to August 31st 2018 are used in our prediction and clustering models. Disruption data for August 2017 is however used to quantify values for the predictor *Disruption frequency* for disruption probabilities in September 2017.

IV. RESULTS AND DISCUSSION

A. Analysis of Disruptions

In this section we present the results for our study contribution to analyse disruption exposure characteristics of a PT network based on empirical disruption log data, applied to the Washington metro network. Fig. 3 presents the relative frequency of the 15 different disruption types, categorized in the five main categories as shown in Fig. 1. The top 3 of most frequent disruption types consists of actions / errors of dispatchers, terminal supervisors or interlocking operators (25%), injured/sick/aggressive passengers (20%) and door malfunctioning (12%). Two third of all disruptions are operations-related (action / error dispatcher or train operator) or vehicle-related (primarily malfunctioning of doors and brakes). Infrastructure-related disruptions have only a relatively small share in the total number of disruptions.

The spatial distribution of disruptions over the Washington metro network is shown in Fig. 4. The weakest stations, being most susceptible to disruptions, can be found at start/terminal stations and in the central area of the network where train frequencies and passenger volumes are generally highest. The least weak stations are intermediate stations (non-terminal and non-transfers) at the line branches, often served by one line only. Largo Town Center (red circle in Fig. 4) suffers from most disruptions, whereas Cheverly (green circle in Fig. 4) is least susceptible.

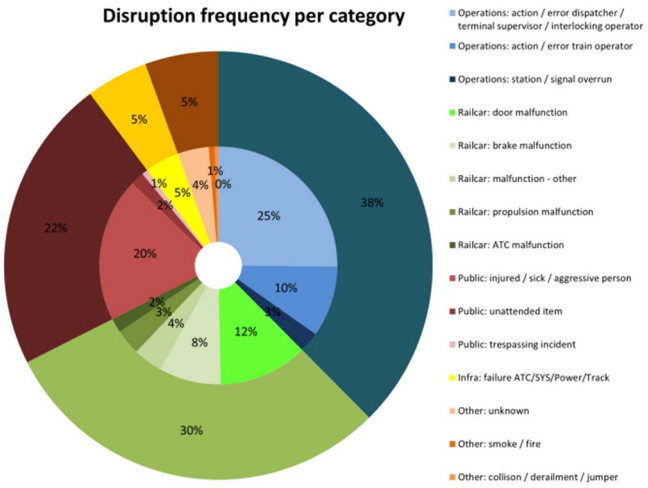

Fig. 3. Relative frequency of each disruption type.

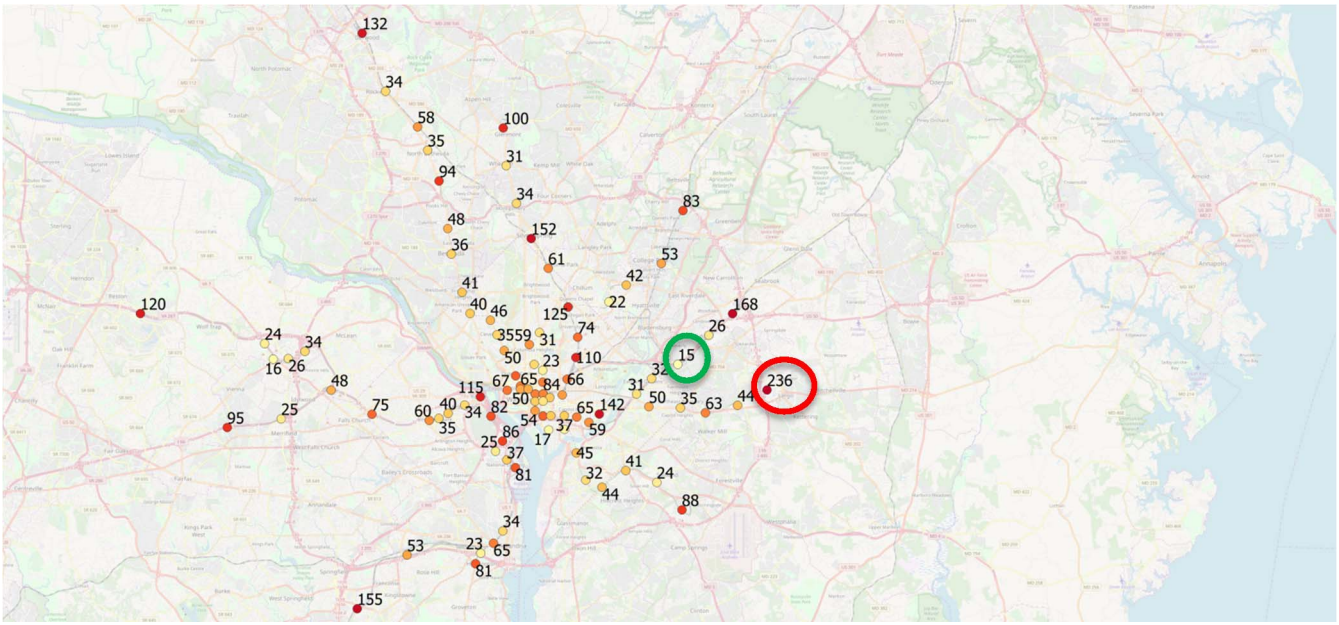

Fig. 4. Spatial distribution of disruptions.

In Fig. 5 one can see the distribution of disruptions over the six metro lines in Washington. It can be seen that most disruptions (31%) occur on the Red line, whereas only 7% of the disruptions occurs on the Yellow line. When the disruption share per line is contrasted to the share of train-kilometres of each line, we can conclude that both shares are of the same magnitude for the Blue, Orange, Green and Yellow line. The disruption share of the Red line is somewhat higher than its share in train-kilometres, while the opposite is true for the Silver line. This suggests the Red line is relatively weak compared to other lines. Causes might be found in the state and characteristics of rolling stock and infrastructure of this line. For the Silver line, the most recently opened line, an opposite explanation applies.

B. *Prediction of Disruptions*

In this section results are presented for our study contribution to develop a prediction model to predict future exposure to different disruption categories for each station for different times of the day. Applied to the Washington case study network, we use our developed model to predict the probability a certain disruption type occurs at each station during each time period (peak, daytime off-peak and evening) for one full future year. By multiplying the predicted disruption probabilities for each time period with the number of time periods per year (using Eq.3), the expected number of disruptions per type and station is calculated.

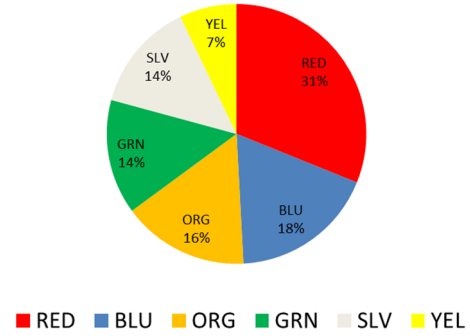

Fig. 5. Disruption distribution over different metro lines.

Model estimation

Based on the number of predictors and one-hot encoding, our final feature vector consists of (991 time periods per year * 95 stations) 94,145 samples and 34 columns. The shape of the target vector is (94,145; 1), respectively (94,145; 16) when binarized into the 16 disruption classes (15 disruption types plus no disruption). The optimal number of neurons of the hidden layer for the MLP classifier is therefore sought between 16 and 34 neurons. From Fig. 6 can be concluded that the log loss is minimized when 29 neurons are used. Table 2 compares the performance between the logistic regression and MLP classifier, showing both the log loss score and area under the ROC curve (AUC) score. Although the model performances do not differ substantially, we can conclude that the MLP classifier does outperform the logistic regression model in both log loss and AUC score. Further results are therefore based on the MLP classifier.

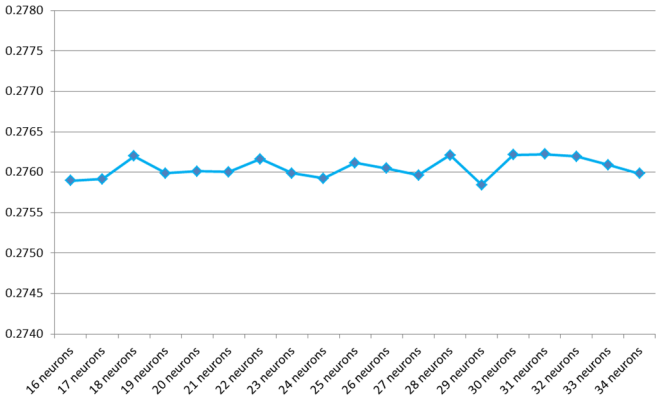

Fig. 6. Log loss for MLP as function of number of neurons of hidden layer.

TABLE 2. PERFORMANCE PREDICTION MODELS

Prediction model	**Log loss**	**AUC**
Logistic regression (random 5-fold cross validation)	0.2817	0.6779
Multilayer perceptron classifier (random 5-fold cross validation)	0.2758 (-2.1%)	0.7107 (+4.8%)

Model validation

The expected number of disruptions (per type, station and time period) is predicted for one future year using the MLP multiplier. In Fig. 7 a comparison is made between the expected exposure to each disruption category per year, aggregated over stations and time periods, and the observed yearly disruption exposure from our empirical dataset. There is a very high correlation (>0.995) between our predicted numbers and observed values. Especially predictions of exposure to injured/sick/aggressive passengers, door malfunctioning, actions / errors of the train operator and collisions are accurate. It can however also be noted that our prediction model tends to underestimate disruption exposure somewhat. On average the expected number of disruptions is underestimated by 8% using our model, indicating there is still room for further improvement of our model.

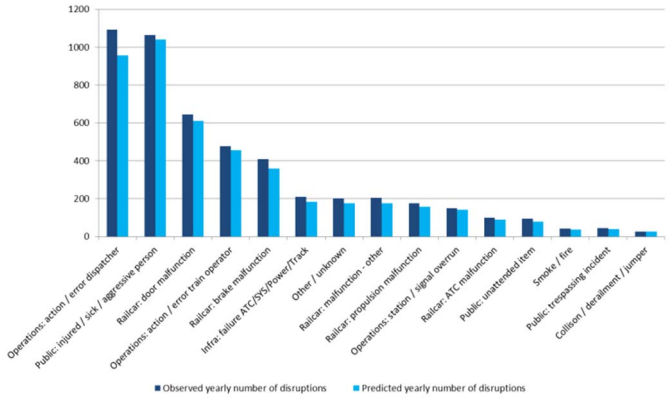

Fig. 7. Prediction accuracy MLP classifier.

C. *Metro Station Clustering*

This section discusses our research contribution to obtain insights in disruption exposure and vulnerability of different types of stations in the network. The predicted disruption exposure per station and per disruption category from our MLP classifier are used as input to cluster the 95 metro stations according to their susceptibility to disruptions.

The clustering results are visualized using the dendrogram shown in Fig. 8. For interpretation purposes, 10 distinctive and intuitive station clusters are shown using different colours. The optimal number of clusters is equal to four based on maximizing the silhouette coefficient, as indicated by the dotted line in Fig. 8. The cophenetic correlation coefficient equals 0.845, what can be considered a good clustering quality. The first and second cluster from these four only contain terminal stations. In line with the empirical observations previously, start/terminal stations show to be exposed substantially more frequent to disruptions than other stations. An explanation for this is that several disruptions often arise at the first station of the line: a railcar malfunctioning when testing the train, a late or sick driver not arriving on time, or a late movement of the train from the yard to the first station. Cluster 1 (the lowest green labels) only consists of the start/terminal stations of the Blue Line (Largo Town Center and Springfield), which stations are even more susceptible to disruptions than the 10 other start/terminal stations of the other lines, which form cluster 2 (the red and light-blue labels). Cluster 3 (the purple and gold labels) is formed of all 8 transfer stations of the Washington metro network (Metro Center, L'Enfant Plaza, Gallery Place and Fort Totten). The separate clustering of these stations shows that exposure patterns for transfer stations form a separate group, possibly related to the high volume of transferring passengers and more frequent occurrence of train operator and dispatcher actions in relation to transfers. Cluster 4 contains all 75 stations which are not a start/terminal or transfer location, which are in general less susceptible to disruptions than stations from the other three clusters. Interesting patterns can however also be found within this cluster. Intermediate stops mostly served by one line are grouped per line: the majority of the intermediate stops of the Red line are grouped together (purple labels), as well as the stops of the Green line including the few stops shared with the Yellow line (light-blue labels). Also stops on trunks served by two or three lines are grouped together: one cluster consists of stops of the Silver/Orange/Blue trunk (red labels), the other cluster contains stops of the Blue/Yellow and Blue/Silver trunk (green labels). At last, stops on the branches of the Orange and most recently constructed Silver line, as well as the Orange/Silver trunk, are grouped together as being least susceptible to disruptions.

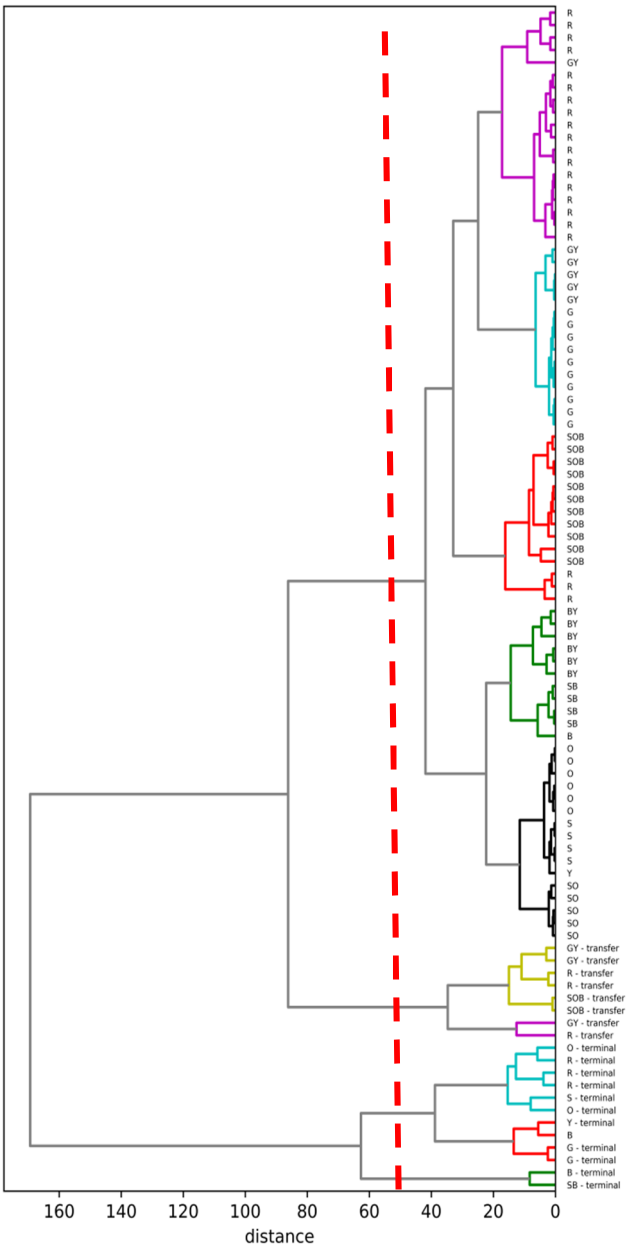

Fig. 8. Dendrogram with resulting clustering of metro stations.

V. CONCLUSIONS

The objective of our study is to develop a generic approach to accurately predict disruption exposure at different stations of a PT network, thereby incorporating the location-specific characteristics of different stations. For this end, we used a 13-month incident database consisting of all disruptions between August 1st 2017 and August 31st 2018 on the Washington metro network as input. Based on the empirical data we can conclude that the most important causes for disruptions are related to actions / errors of dispatchers, terminal supervisors or interlocking operators (25%), injured/sick/aggressive passengers (20%) and railcar door malfunctioning (12%). Two third of all disruptions are operations-related (action / error dispatcher or train operator) or vehicle-related. Based on the incident database we developed a supervised learning prediction model, which predicts the probability on a certain disruption category for different stations and times of the day. Predicting the expected number of disruption for one year using a MLP classifier shows a strong correlation between predicted and empirical values. Our model allows for a reasonably accurate prediction of disruption exposure, although exposure is somewhat underestimated. Future research will focus on further improvement of the accuracy of this prediction model. Using predicted values from our supervised learning model, all metro stations are clustered based on their susceptibility to disruptions. Four clear groups of stations are found in this clustering. The first two clusters only consist of start/terminal stations, which show to be substantially more susceptible to disruptions than other stations. All transfer stations of our case study network are grouped into a separate cluster, indicating a distinctive exposure pattern related to the particular characteristics of transfer stations. All other intermediate, non-terminal and non-transfer stops are grouped together in one cluster as being least susceptible to disruptions.

Our study results provide PT authorities and operators insight to the type and location of disruptions which contribute most to total network disruption exposure. This supports them in prioritizing what type of disruptions at what location to focus on, to potentially achieve the largest reduction in disruption exposure. It is recommended to give priority to reducing disruptions at start/terminal and transfer stations, thereby particularly focusing on operations- and vehicle-related disruptions. While this study merely focuses on disruption exposure, in future research these results will be integrated with predicted impacts of disruptions, so that the integral contribution to network vulnerability can be quantified.

ACKNOWLEDGMENT

This research was performed as part of the TRANS-FORM (Smart transfers through unravelling urban form and travel flow dynamics) project funded by NWO grant agreement 438.15.404/298 as part of JPI Urban Europe ERA-NET CoFound Smart Cities and Communities initiative. The authors thank WMATA for their valuable cooperation and data provision.

REFERENCES

[1] Hörcher, D., D.J. Graham, and R.J. Anderson, “Crowding cost estimation with large scale smart card and vehicle location data,” Transportation Research Part B, vol. 95, pp. 105-125, 2017.

[2] Tirachini, A., R. Hurtubia, T. Dekker, and R.A.Daziano, “Estimation of crowding discomfort in public transport: Results from Santiago de Chile,” Transportation Research Part A, vol. 103, pp. 311-326, 2017.

[3] Yap, M.D., O. Cats, and B. van Arem, “Crowding valuation in urban tram and bus transportation based on smart card data,” Transportmetrica A: Transport Science, DOI: 10.1080/23249935. 2018.1537319, 2018.

[4] Yap, M.D., S. Nijenstein, and N. van Oort, “Improving predictions of public transport usage during disturbances based on smart card data,” Transport Policy, vol. 61, pp. 84-95, 2018.

[5] Jenelius, E., T. Petersen, and L-G. Mattsson, “Importance and exposure in road network vulnerability analysis,” Transportation Research Part A, vol. 40, pp. 537-560, 2006.

[6] Cats, O. and E. Jenelius, “Planning for the unexpected: the value of reserve capacity for public transport network robustness,” Transportation Research Part A, vol. 81, pp. 47-61, 2015.

[7] Cats, O. and E. Jenelius, “Dynamic Vulnerability Analysis of Public Transport Networks: Mitigation Effects of Real-Time Information,” Networks and Spatial Economics, vol. 14, pp. 435-463, 2014.

[8] Derrible, S. and C. Kennedy, “The complexity and robustness of metro networks.” Physica A, vol. 389, pp. 3678-3691, 2010.

[9] Corman, F., A. D’Ariano, and I.A. Hansen, “Evaluating disturbance robustness of railway schedules,” Journal of Intelligent Transport Systems, vol. 18, pp. 106-120, 2014.

[10] Corman, F., A. D’Ariano, D. Pacciarelli, and M. Pranzo, “A tabu search algorithm for rerouting trains during rail operations,” Transportation Research Part B , vol. 44, pp. 175-192, 2010.

[11] Cats, O., and E. Jenelius, “Beyond a complete failure: the impact of partial capacity degradation on public transport network vulnerability,” Transportmetrica A: Transport Dynamics, vol. 6, pp. 77-96, 2018.

[12] Tonnelier, E., N. Baskiotis, V. Guigue, and P. Gallinari, “Anomaly detection in smart card logs and distant evaluation with Twitter: a robust framework,” Neurocomputing, vol. 298, pp. 109-121, 2018.

[13] Cats, O., M.D. Yap, and N. van Oort, “Exposing the role of exposure: Public transport network risk analysis,” Transportation Research Part A, vol. 88, pp. 1-14, 2016.

[14] Yap, M.D., N. van Oort, R. van Nes, and B. van Arem, “Identification and quantification of link vulnerability in multi-level public transport networks: a passenger perspective,” Transportation, vol. 45, pp. 1161-1180, 2018.

[15] Rodriguez-Nunez, E., and J.C. Garcia-Palomares, “Measuring the vulnerability of public transport networks,” Journal of Transport Geography, vol.35, pp. 50-63, 2014.

[16] Oliveira, E.L., L. Silva Portugal, and W. Porto Junior, “Indicators of reliability and robustness: Similaries and differences in ranking links of a complex road system,” Transportation Research Part A, vol. 88, pp. 195-208, 2016.

[17] Pedregosa et al., “Scikit-learn: Machine Learning in Python,” JMLR, vol. 12, pp. 2825-2830, 2011.