# Abnormal Situation Simulation and Dynamic Causality Discovery in Urban Traffic Networks

**Yadi Wang**
School of Software and Microelectronics, Peking University, Beijing, China

**Yicheng Pan**
School of Electronics Engineering and Computer Science, Peking University, Beijing, China

**Meng Ma**[1]
National Engineering Research Center for Software Engineering, Peking University, Beijing, China

**Ping Wang**[2]
School of Software and Microelectronics, Peking University, China
National Engineering Research Center for Software Engineering, Peking University, Beijing, China

## Abstract

Various participants in urban traffic systems intertwine a highly complicated coupling network. An interpretable analysis of underlying correlations is one of the keys to understanding traffic anomalies. Unfortunately, abnormal situation analysis in real scenarios faces severe limitations in negative sample deficiency, data integrity, and verifiability. In view of this, we developed a simulation tool – the Traffic Anomaly Situation Simulator (TASS). Through configurable scripts, TASS simulates real traffic networks by road editing, data collection, and fault injection. Given the generated cases, we designed a dynamic causal discovery algorithm, Dycause-Traffic, to demonstrate the features of causality in traffic anomalies.

**2012 ACM Subject Classification** Applied computing → Transportation

**Keywords and phrases** SUMO simulation, dynamic causality discovery, congestion propagation

**Digital Object Identifier** 10.4230/LIPIcs.COSIT.2022.22

**Category** Short Paper

**Funding** The National Natural Science Foundation of China (92167104, 62072006) and Science and Technology on Communication Networks Laboratory (6142104200103) support this work.

## 1 Introduction

Abnormal traffic situation such as congestion keeps increasing in metropolitan areas, contributing to great economic losses and causing delay, compromised levels of service, and discomfort during traveling [10]. Since the 1980s, enormous progress has been made in predicting traffic, but current studies lack the analysis of important flow relationships and evolution trends of important urban structures [3, 8], and the dynamic causal association information hidden in the dataset has not been fully explored. Besides, most algorithms only consider the existence of causalities while ignoring their strength.

The traditional traffic analysis primarily relies on traffic time series data. For example, by redesigning the distance measurement and clustering methods in the clustering algorithm, the urban traffic flow patterns and urban structures are revealed from the spatiotemporal data [9, 12, 17]. The association rule and clustering method are often combined into a

[1] Corresponding author. Meng Ma is also a visiting researcher with the Science and Technology on Communication Networks Laboratory.
[2] Corresponding author. Ping Wang is also with the Key Laboratory of High Confidence Software Technologies (PKU), Ministry of Education, China.



multi-step model [15]. But less consideration is given to the topological relationship between roads and the interaction between road segments. Graph-based algorithms that can be used to discover such associations [18] mainly include CGC [5] and PSTE [14] based on Granger causality theory [4], SVAR-FCI [11] and PCMCI [16] based on conditional constraints, and other models [13, 19] built on GNN. However, the upper limit of the number of conditional independence tests of the algorithm increases with the maximum degree of any node.

High-quality urban traffic flow data is the basis for analyzing abnormal urban traffic situations. Some papers provide relevant research with good actual data [2, 6, 7]. But the data generated in real scenarios are insufficient. The timely feedback of traffic intervention cannot be achieved quickly and it's challenging to excavate the influencing factors under the dynamically changing traffic conditions. Analyzing urban traffic anomalies requires controllable and reproducible data, and simulation tools provide this possibility.

Our contributions are manifolds: *1.Urban traffic simulation.* We build an urban traffic simulation system based on the relevant technical framework of SUMO [1], which provides a basis for the design and implementation of simulation experiments. *2.Traffic abnormal situation injection.* During the simulation, controllable anomalies, such as signal lights, lanes, and flow, are injected into the traffic road network to reproduce controllable traffic congestion. *3.Abnormal situation propagation analysis.* We use the dynamic causal correlation algorithm to mine the abnormal situation propagation path and intensity of traffic and analyze road congestion's impact on the overall road network and other roads.

## 2 Traffic abnormal situation simulation

Given the requirement of adjusting the traffic lights, intersections, road facilities, and lanes in the abnormal urban traffic situation, we design the Traffic Anomaly Situation Simulator (TASS) based on SUMO [1], and additionally various options for traffic anomaly injection in the case of simulating a given traffic demand consisting of a single-vehicle.

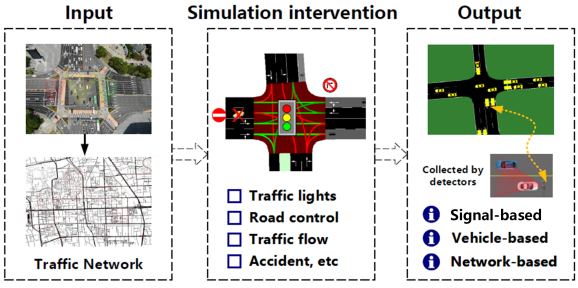



**Figure 1** Overall pipeline of TASS.

The overall simulation flow of TASS is shown in Fig. 1. Please find the source code, sample cases, and dataset of TASS in this anonymous `github` repository.

### 2.1 TASS input

The input of TASS includes the following elements. In this study, we leverage the OSMWebwizard program to generate the network from the satellite map information of OpenStreetMap (OSM), including the traffic network structure of selected area as well as the subway, light rail, and other public transport systems.

(1) *Road attributes.* The basic network structure of the road, the allowed and limited direction, the types of vehicles that are allowed to pass, etc.
(2) *Vehicle attributes.* Vehicle type, route planning, driving preference and proficiency of the driver, etc., are mainly described by the dynamic model of the vehicle.
(3) *Interaction between vehicles.* The driving behavior of vehicles during operation is not only affected by road conditions and route planning but also by other vehicles, resulting in various traffic phenomena, such as traffic anomalies and traffic congestion.

## 2.2 Simulation intervention

The core task of TASS is the reproduction and intervention of traffic demand, that is, the data and related information of traffic flow. By intervening in the simulation process through the Traci script, TASS can realize the fixed-point timing injection of the abnormal situation. Urban traffic anomalies can be generally divided into three categories:

(1) *Abnormal state of lanes or edges.* Lanes or edges cannot meet the normal traffic needs of vehicles and pedestrians. For example, traffic accidents, vehicle lane changes, traffic control, and other abnormal driving behaviors cause lane blockages, resulting in lane prohibition. The maximum travel speed for fixed lanes can be set to 0 via Traci in TASS.
(2) *Abnormal traffic facilities.* For example, the traffic light is damaged, or the alternating cycle of the traffic light is abnormal. The Traffic Lights States (TLS) can be modified after specifying the number of simulation steps during the simulation. After recovery, the TLS can be changed to the normal state.
(3) *Abnormal driving state of vehicles.* It is generally reflected in the abnormal increase of pedestrians and vehicles caused by extreme weather, peak travel period, road control, and low-speed vehicles. The maximum driving speed of the lanes in the area can be reduced as required by Traci.

## 2.3 TASS output

In practice, the traffic simulation focuses on three output values to solve the traffic problem: vehicle-based information, network-based element (such as stations, traffic-lights, cross-linking, and other factors, which are related to the geometric layout of roads), and hidden cost (such as waiting time, time loss and depart delay). In this paper, we collect the three outputs but consider the vehicle-based data primarily. The vehicle-based information includes positions, speeds, acceleration of all vehicles for every simulation step, emission values, battery usage, collisions, lane-changing events, and trajectory data. Data collected from detectors provide lane/edge-based performance measurements such as speed, road occupancy, and traffic flow.

# 3 Dynamic Causality Discovery – Dycause-Traffic

## 3.1 Problem statement

We formally denote a traffic network as a directed graph $G = (V, E)$, where $V$ is the set of $N = |V|$ vertices representing nodes on the road network (i.e., detectors in the following experiments), and $E$ is the set of edges representing the connectivity among vertices. At each time step $t$, there is a graph signal $\mathbf{X}^{(t)} \in \mathbb{R}^{N \times C}$ on graph G, where $C$ is the number of features of the input signal (such as traffic flow, speed, occupancy). The superscript $(t)$ is the sample time index in a short period. To reveal the dynamic causalities between locations in traffic anomalies, we design the Dycause-traffic algorithm to examine Granger causal intervals with sliding windows, generating the dynamic causality curves $C_{i,j}(t)$ for each pair

of detectors. The dynamic causality curve $C_{i,j}(t)$ depicts the time-varying strengths of the correlation from detector $v_i$ to $v_j$. Thus, we transform the anomaly propagation analysis task into the computing of the dynamic causality dependencies between vertices denoted as $\hat{Y} = (\hat{\mathbf{X}}^{t_{p+1}}, \hat{\mathbf{X}}^{t_{p+2}}, ..., \hat{\mathbf{X}}^{t_{p+q}})$ from the historical simulation output of time series of $p$ time steps. Note that the calculation is limited by the overall simulation steps $q$.

## 3.2 Temporal dynamic causality discovery

We build Dycause-traffic based on the Granger causality test. Particularly, we denote the time series of the node $i$ and node $j$ as $x_i$ and $x_j$. Our test constructs two linear regression models (the partial $\mathcal{M}_p$ and the full $\mathcal{M}_f$) to predict $x_j(t), t = 1, 2, ..., T$ with the past observations. Please note that one of them has more independent variables $x_i(t-l), 1 \leq l \leq L_{lag}$, representing the additional information from the node $i$.

$$\mathcal{M}_p \colon \hat{x}_j(t) = \sum_{l=1}^{L_{lag}} \alpha_l^p x_j(t-l) + b^p, \mathcal{M}_f \colon \hat{x}_j(t) = \sum_{l=1}^{L_{lag}} \left( \alpha_l^f x_j(t-l) + \beta_l^f x_i(t-l) \right) + b^f \quad (1)$$

Here, $\hat{x}_j(t)$ denotes the prediction and $\alpha_l^p$, $b^p$, $\alpha_l^f$, $\beta_l^f$ and $b^f$ are the parameters. By computing the sums of squared errors ($SSE_p, SSE_f$) from the two models, which is calculated by $\sum_t (\hat{x}_j(t) - x_j(t))^2$, we test the null hypothesis using the F-distribution. When the null hypothesis cannot be rejected, we accept that the extra information can not improve the prediction and $x_i$ has no causal impact on $x_j$. To do that, we calculate the value $F = \frac{(SSE_p - SSE_f)/(d_f - d_p)}{SSE_f/(T - d_f - 1)}$ to estimate the probability of the null hypothesis based on the F-distribution of $\mathcal{F}_{d_f - d_p, T - d_f - 1}$ and reject it lower than a significant threshold $\alpha$, which in turn indicates the Granger causality from $x_i$ to $x_j$. Here, $d_p$ and $d_f$ are the degrees of freedom of the two models, which are $L_{lag}, 2L_{lag}$, respectively. $T$ is the total number of samples. Given the fact that the causalities change dynamically in anomalies, we conduct the test on enumerated sliding windows $\{[s_q : e_q] \mid 0 \leq s_q < e_q \leq T\}$ instead of a single-window $[0, T]$. Finally, Dycause-traffic generates the causality curve $C_{i,j}$ according to Algorithm 1.

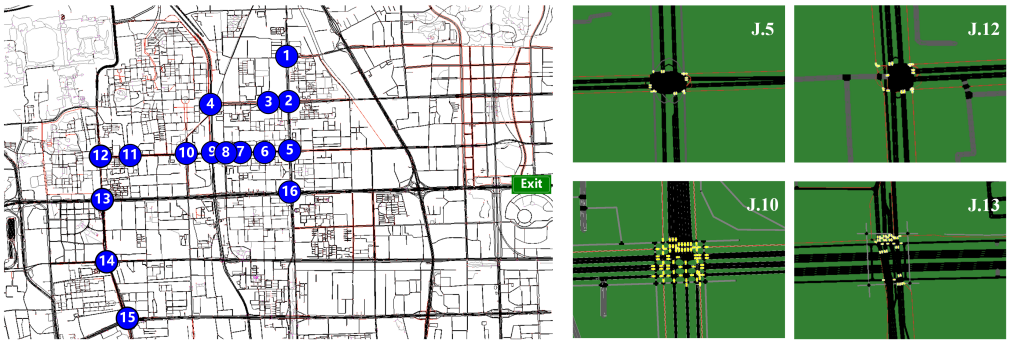
**Algorithm 1** Dycause-Traffic.

**Input:** Metric data during time intervel $T$: $\mathbf{X} \in \mathbb{R}^{N \times T}$, Granger causality significance value $\alpha$, basic interval length $L_b$
**Output:** Dynamic causality curves $C \in \mathbb{R}^{N \times N \times T}$.

```
begin
    for each node pair i → j do
        C_{i,j} ⟵ {0}^T
        for s = 0; s ≤ T − L_b; s += L_b do
            for e = s + L_b; e ≤ T; e += L_b do
                Estimate the ⌈F⌉ and ⌊F⌋ using previous regression models
                if ⌈F⌉ and ⌊F⌋ indicate pruning then
                    Continue
                else if ⌈F⌉ and ⌊F⌋ indicate causality then
                    Increase C_{i,j}(s: e) by 1
```

## 4 Experiments

We experimentally select a 50 $km^2$ dense traffic network in Beijing, China. As shown in Fig. 2(a), we number the main junctions from J.1 to J.16. The horizontal and vertical lanes at J.13 and J.16 are urban arterial roads. The horizontal line is the Fourth Ring Road in Beijing, and the surrounding infrastructure density is low. Schools, residential areas, and commercial areas are scattered around other junctions. A certain number of detectors are distributed at the traffic inflow position at each junction, which can provide traffic flow, vehicle speed, and lane occupancy data during the simulation period.




(a) Map and intersection number (b) Distribution of traffic detectors

**Figure 2** Map and detectors in our experiment.

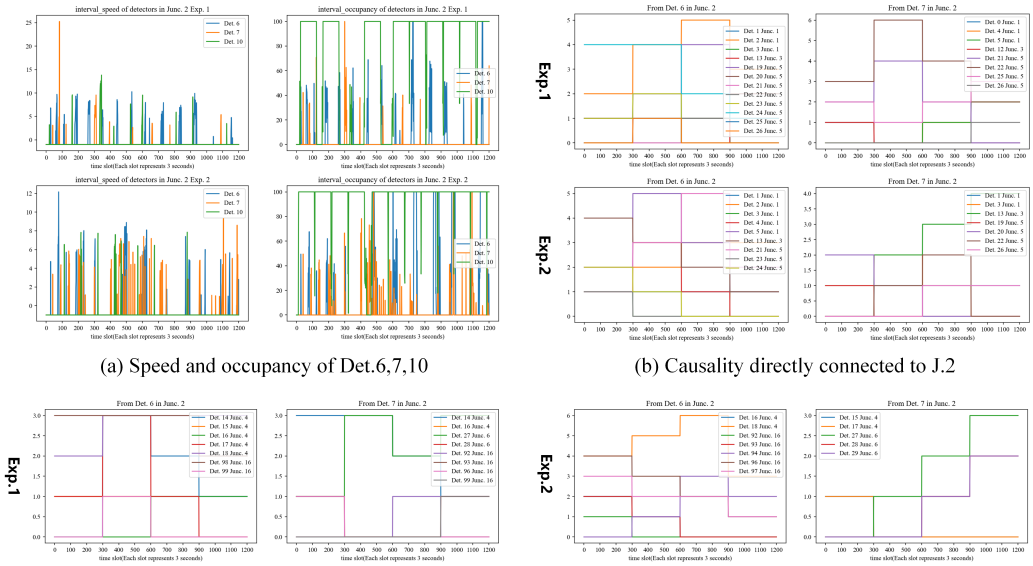
The distribution of detectors at junctions is shown in Fig. 2(b). The detectors are placed in the direction of the traffic entering the junction. J.5 and J.12 are on the urban arterial road. There are fewer connections to the secondary roads around the junction. In other words, the traffic network structure is simpler where the detector density is also lower. Correspondingly, the road network structure of J.10 and J.13 is more complicated, which is related to the surrounding infrastructure to a certain extent.

We conduct TASS to the following experimental simulations, considering traffic light failure, road failures, and abnormal traffic flow. Details of each experiment are summarized in Table 1. The simulation time is 3600 seconds, i.e., one hour in total. The fault injection period is 600 to 1800 seconds, lasting for 20 minutes. To improve the efficiency of model training, we select the raw data of 100 detectors, from Det.0 to Det.99.

**Table 1** Simulation details of each experiment.

| Experiment ID | Simulation Details | Failure Type |
|---|---|---|
| Exp. 1 | Original traffic flow, the detector sampling frequency is set to 3 seconds | No failure. |
| Exp. 2 | The traffic light at intersection Junc. 2 is faulty, only green | Traffic light failures |
| Exp. 3 | The traffic light at intersection Junc. 2 is faulty, only red | |
| Exp. 4 | The traffic light at intersection Junc. 12 is faulty, only green | |
| Exp. 5 | The traffic light at intersection Junc. 12 is faulty, only red | |
| Exp. 6 | The inbound route of the north-south road at junction Junc. 2 is faulty | Lane failures. Only the north-south edges are modified. |
| Exp. 7 | The outbound route of the north-south road at junction Junc. 2 is faulty | |
| Exp. 8 | The inbound route of the north-south road at junction Junc. 5 is faulty | |
| Exp. 9 | The outbound route of the north-south road at junction Junc. 5 is faulty | |
| Exp. 10 | Double the traffic flow of some lanes in Experiment 1 | Abnormal traffic flow. The lanes selected are the same. |
| Exp. 11 | Reduce the traffic flow of some lanes in Experiment 1 | |

Fig.3 shows part of the experimental results where we observe several features of anomaly propagation in traffic networks.




(a) Speed and occupancy of Det.6,7,10

(b) Causality directly connected to J.2

(c) Causality secondary connected to J.2

**Figure 3** Example of experimental results.

*No traffic lights or a single permit(red or green only)can cause traffic anomalies.* Fig. 3(a) shows the speed and occupancy rate of Det.6,7,10 in Exp.1,2. When there is only a green light, the vehicle's average speed decreases. During the time slot 200 to 600 (i.e., 600s to 1800s), the speed has a short-term increase compared to Exp.1. Still, the green light increases the traffic volume while also increasing the road load, and the final result is a negative effect. When there are red lights at J.2, the vehicle runs normally only when there is no abnormal increase in the early period at J.2. Even if the accident is cleared, it does not quickly return to the initial condition.

*Anomaly propagation is restricted by road capacity and road structure.* J.3 is a junction directly linked with J.2, and the maximum road load is smaller than that of J.2. Therefore, the failure propagation from J.2 to J.3 is relatively stable, as shown in Fig. 3(b), and the results of all experiments have relatively minor differences. Since the capacity from the main arterial road to the secondary arterial road is limited by the secondary arterial road, the J.2 continues to transfer the flow to the J.3. Still, at the same time, once the traffic flow exceeds the threshold, the increase will not cause additional congestion for J.3.

*Upstream and downstream traffic can affect each other.* The upstream traffic flow parameters have a transmission effect on the downstream traffic flow parameters through the road carrier, and vice versa. As a result, they both will experience a time delay. Dotted failure can be achieved by injecting specified road anomalies or adding specific traffic flows. It will be propagated with the overflow or outflow mode, thereby creating linear or area failure. But it should be noted that the reproduction of overflow or outflow mode needs to be better reproduced in the case of meeting the road conditions.

*Secondary road connection structure affects anomaly's propagation.* Fig. 3(c) shows how the secondary junction connected with J.2 is affected. Traffic anomalies in the same place caused by different factors have inconsistent characteristics. Considering the diffusion of traffic flow, the problem of secondary anomalies caused by local anomalies cannot be ignored. When severe traffic anomalies occur locally, the traffic conditions of upstream and downstream sections and junctions will be affected along with secondary road links. A worse situation could lead to the collapse of large areas of the road network.

## 5 Conclusion

This study presents a novel tool, TASS, to simulate anomalies in urban traffic networks, including traffic light failures, lane failures, and abnormal flow. We design the Dycause-Traffic algorithm to reveal the anomaly's propagation mechanism. Experiments verified that the length of lanes, the connection of roads, and the location of faults are the main factors affecting the propagation of anomalies.

## References


1 Michael Behrisch et al. Sumo–simulation of urban mobility: an overview. In *Proceedings of SIMUL 2011, The Third International Conference on Advances in System Simulation*. ThinkMind, 2011.

2 Gleb Beliakov et al. Measuring traffic congestion: an approach based on learning weighted inequality, spread and aggregation indices from comparison data. *Applied Soft Computing*, 67:910–919, 2018.

3 Etienne Come et al. Spatio-temporal analysis of dynamic origin-destination data using latent dirichlet allocation: Application to vélib'bike sharing system of paris. In *TRB 93rd Annual meeting*, page 19p. Transportation Research Board, 2014.

4 Clive WJ Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: journal of the Econometric Society*, pages 424–438, 1969.

5 Meng Hu et al. A copula approach to assessing granger causality. *NeuroImage*, 100:125–134, 2014.

6 Bin Jiang et al. Topological analysis of urban street networks. *Environment and Planning B: Planning and design*, 31(1):151–162, 2004.

7 Zihan Kan et al. Traffic congestion analysis at the turn level using taxis' gps trajectory data. *Computers, Environment and Urban Systems*, 74:229–243, 2019.

8 Felix Kling et al. When a city tells a story: urban topic analysis. In *Proceedings of the 20th international conference on advances in geographic information systems*, pages 482–485, 2012.

9 Xi Liu et al. Revealing travel patterns and city structure with taxi trip data. *Journal of transport Geography*, 43:78–90, 2015.

10 Qiong Lu et al. The impact of autonomous vehicles on urban traffic network capacity: an experimental analysis by microscopic traffic simulation. *Transp. Lett.*, 12(8):540–549, 2020.

11 Daniel Malinsky et al. Causal structure learning from multivariate time series in settings with unmeasured confounding. In *Proceedings of 2018 ACM SIGKDD workshop on causal discovery*, pages 23–47. PMLR, 2018.

12 Feng Mao et al. Mining spatiotemporal patterns of urban dwellers from taxi trajectory data. *Frontiers of Earth Science*, 10(2):205–221, 2016.

13 Meike Nauta et al. Causal discovery with attention-based convolutional neural networks. *Machine Learning and Knowledge Extraction*, 1(1):312–340, 2019.

14 Angeliki Papana et al. Detecting causality in non-stationary time series using partial symbolic transfer entropy: Evidence in financial data. *Computational economics*, 47(3):341–365, 2016.

15 Felix Rempe et al. Spatio-temporal congestion patterns in urban traffic networks. *Transportation Research Procedia*, 15:513–524, 2016.

16 Jakob Runge et al. Detecting and quantifying causal associations in large nonlinear time series datasets. *Science advances*, 5(11):4996, 2019.

17 Xiaoying Shi et al. Exploring the evolutionary patterns of urban activity areas based on origin-destination data. *IEEE Access*, 7:20416–20431, 2019.

18 Huijun Sun et al. Urban traffic congestion spreading in small world networks. *International Journal of Modern Physics B*, 19(28):4239–4246, 2005.

19 Zhang Zhang et al. A general deep learning framework for network reconstruction and dynamics learning. *Applied Network Science*, 4(1):1–17, 2019.