Abnormal Events Causal Analysis from Spatio-temporal Human Mobility Network

POOREUMOE KIM

Technische Universität München

Fakultät für Informatik

Lehrstuhl für Connected Mobility

Garching: 29. October 2020

Sample questions

How did Hong Kong demonstration affect international human mobility?

How does controlling human mobility prevent COVID-19 outbreak?

A protest in Hong Kong(2019)

„Lockdown light“ bis Ende November (2020)

Research question

How did Hong Kong demonstration affect international human mobility?

How does controlling human mobility prevent COVID-19 outbreak?

How do social events affect and are affected by anomalies in a spatio-temporal human mobility network?

Contents

Research question

1. Spatio-temporal network

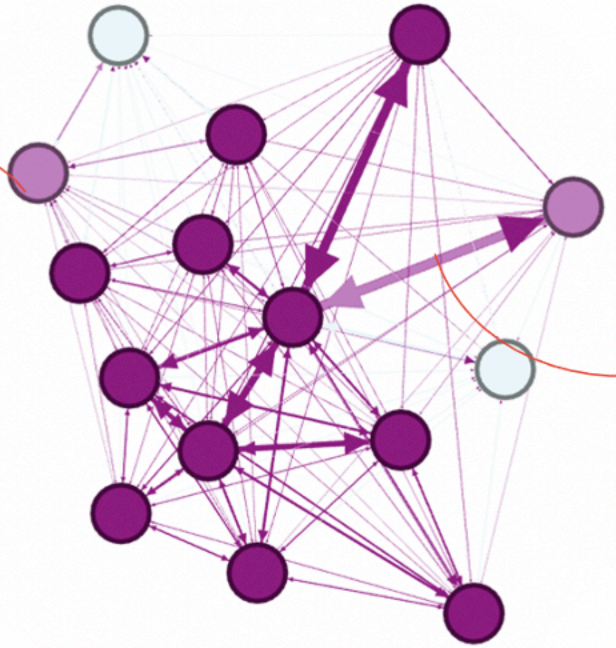
2. Causality between a mobility system and events

How do social events affect and are affected by anomalies in a spatio-temporal human mobility network?

3. Types of anomalies

Spatio-temporal network

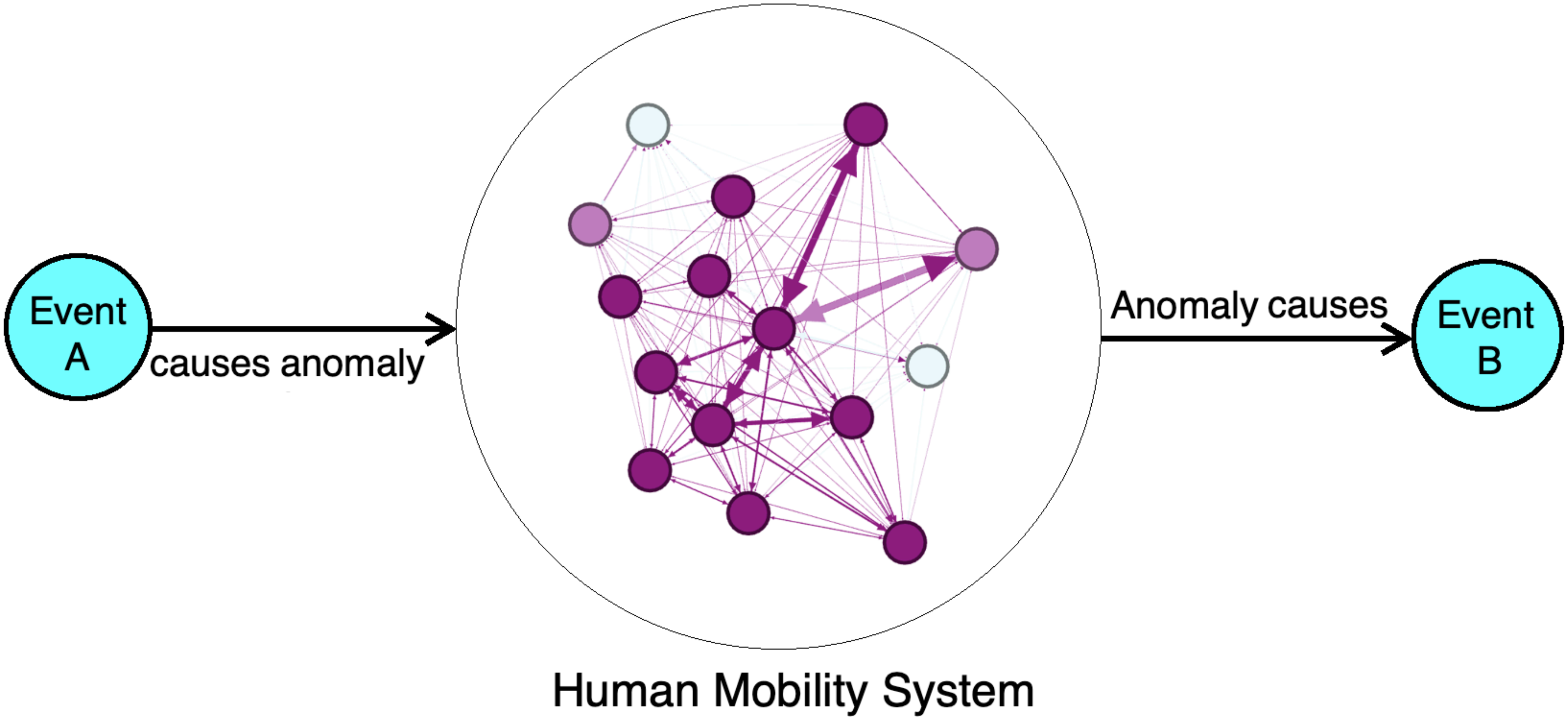
A node: spatial location



An edge:
- people's moving traffic.
- weighted, directed

Nodes & edges vary over time
- Dynamic graph(network)

Events causality



Anomaly types

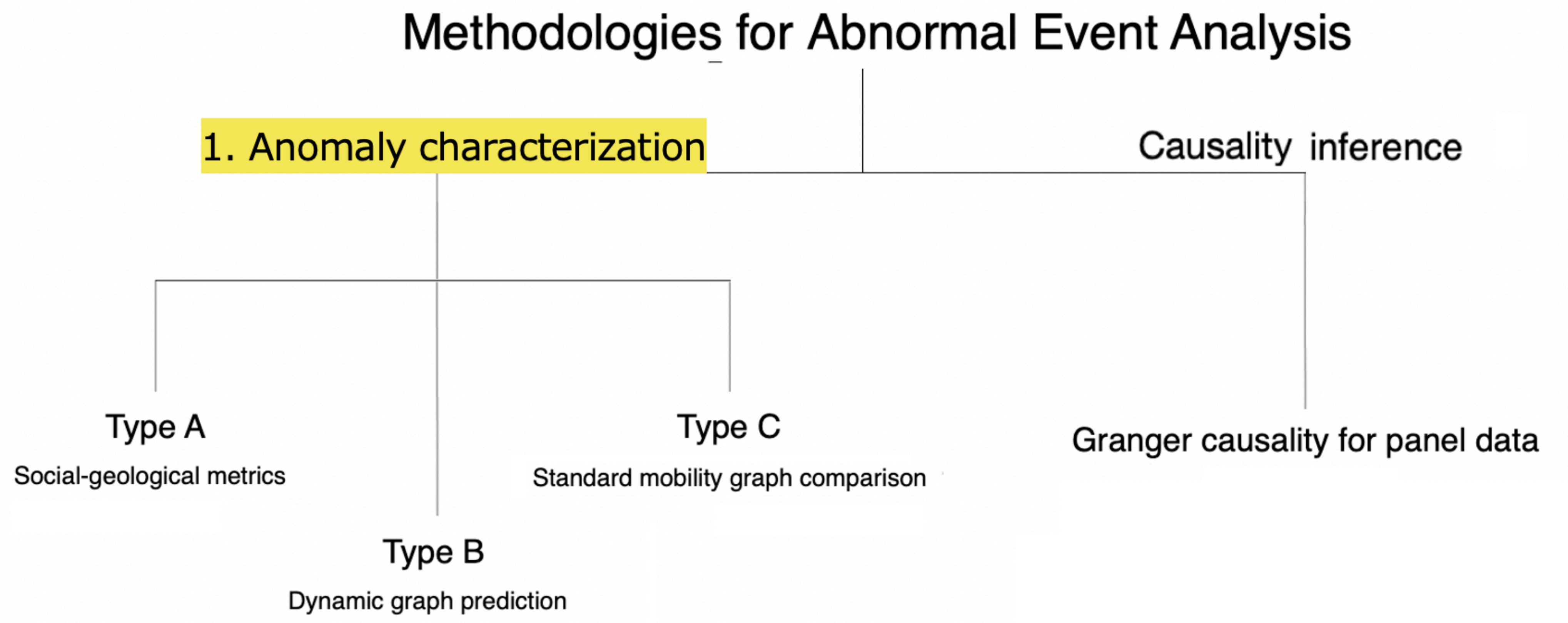
type A : An outlier regarding a graph metric

type B : An observation different from the past mobility pattern

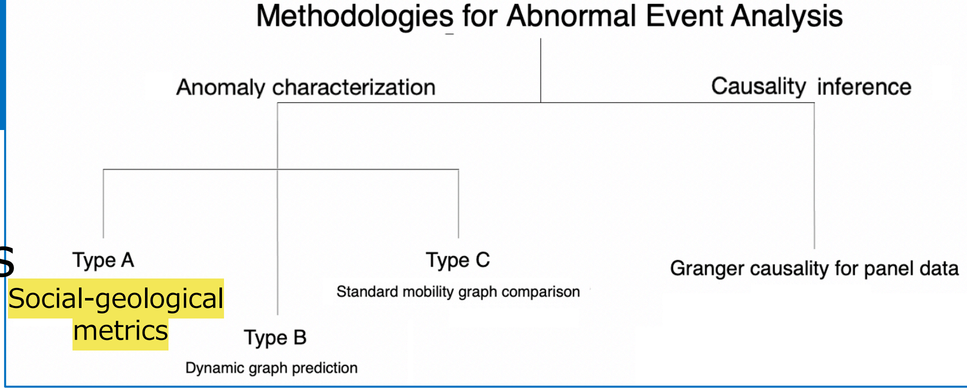
type C : An observation different from the standard mobility pattern

Contents

Methodologies: Anomaly characterization



Social-geological metrics



- Brokerage: between-centrality. High brokerage means the region actively intermediates other regions.

- Popularity: average incoming traffic volume

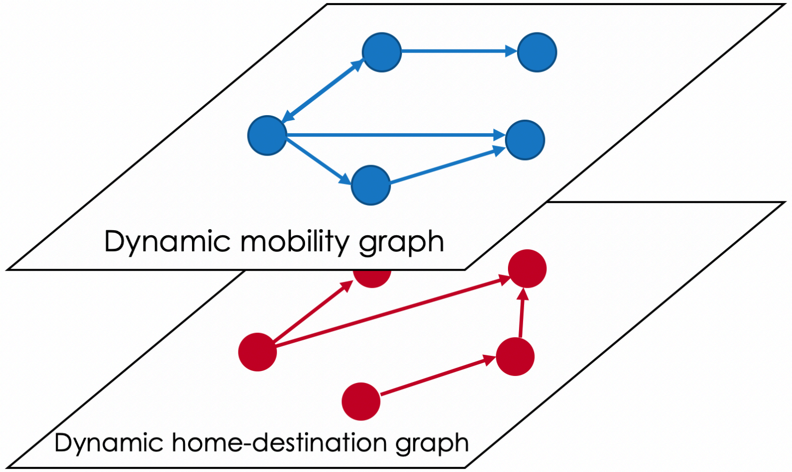
- Entropy: how diverse visitors' home places are. High entropy means heterogeneous people's gathering.

$$H_t(l) = -\sum_{u \in N} p_t(l,u) log p_t(l,u)$$

- Homogeneity: similarity between home regions' travel patterns

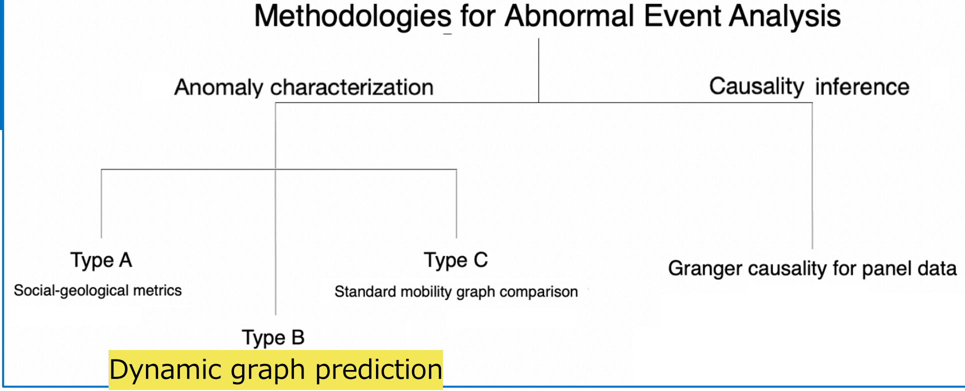
$$S_t(l) = \frac{1}{|V_t(l)|(|V_t(l)|-1)} \sum_{a,b \in V_t(l)} , sim(H^t_{(a,)}, H^t_{(b,)})$$

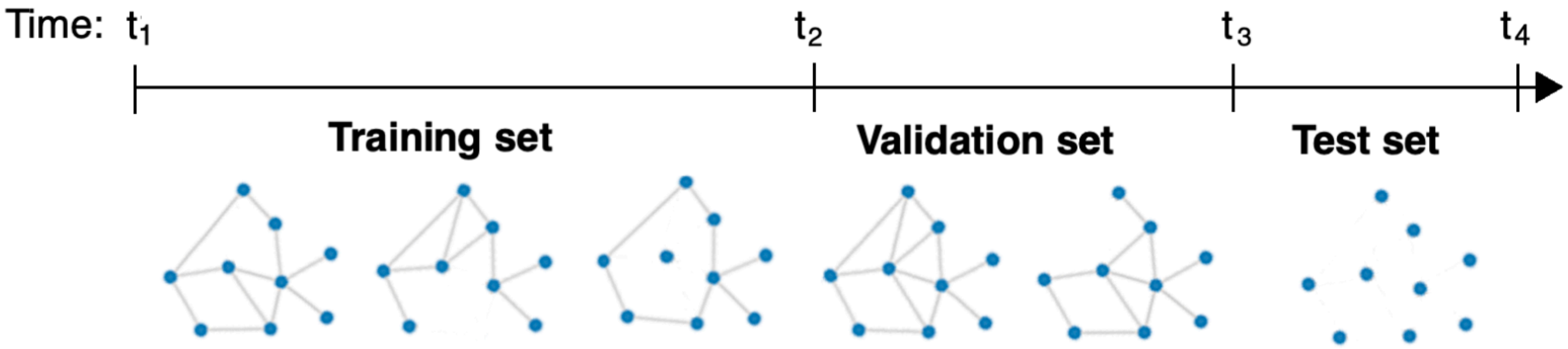
* entropy and homogeneity require travelers' home information



D. Hristova et al. "Measuring urban social diversity using interconnected geo-social networks". In: *Proceedings of the 25th international conference on world wide web*. 2016

Dynamic graph prediction



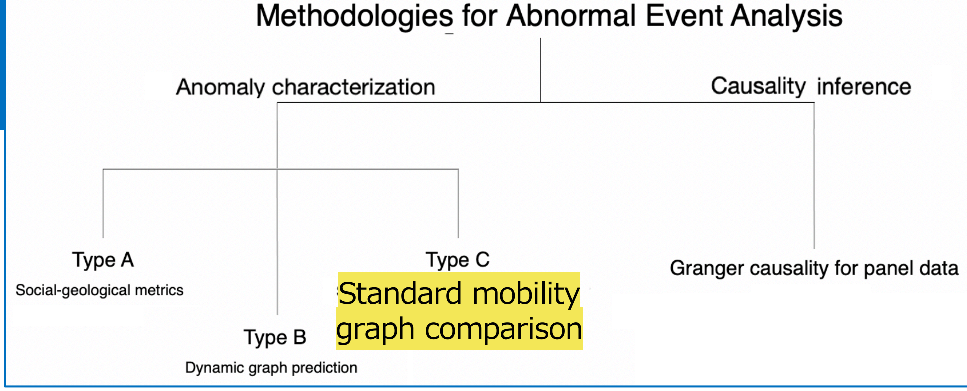


$$H_t^{(l+1)} = GCN(A_t, H_t^{(l)}, W_t^{(l)}) := \sigma(\hat{A}_t H_t^{(l)} W_t^{(l)})$$

T. N. Kipf and M. Welling. "Semi-supervised classification with graph convolutional networks". In: *arXiv preprint arXiv:1609.02907* (2016)

A. Pareja, G. Domeniconi, J. Chen, T. Ma, T. Suzumura, H. Kanezashi, T. Kaler, T. B. Schardl, and C. E. Leiserson. "EvolveGCN: Evolving Graph Convolutional Networks for Dynamic Graphs." In: *AAAI*. 2020, pp. 5363–5370.

Standard mobility graph comparison



Graphs over time

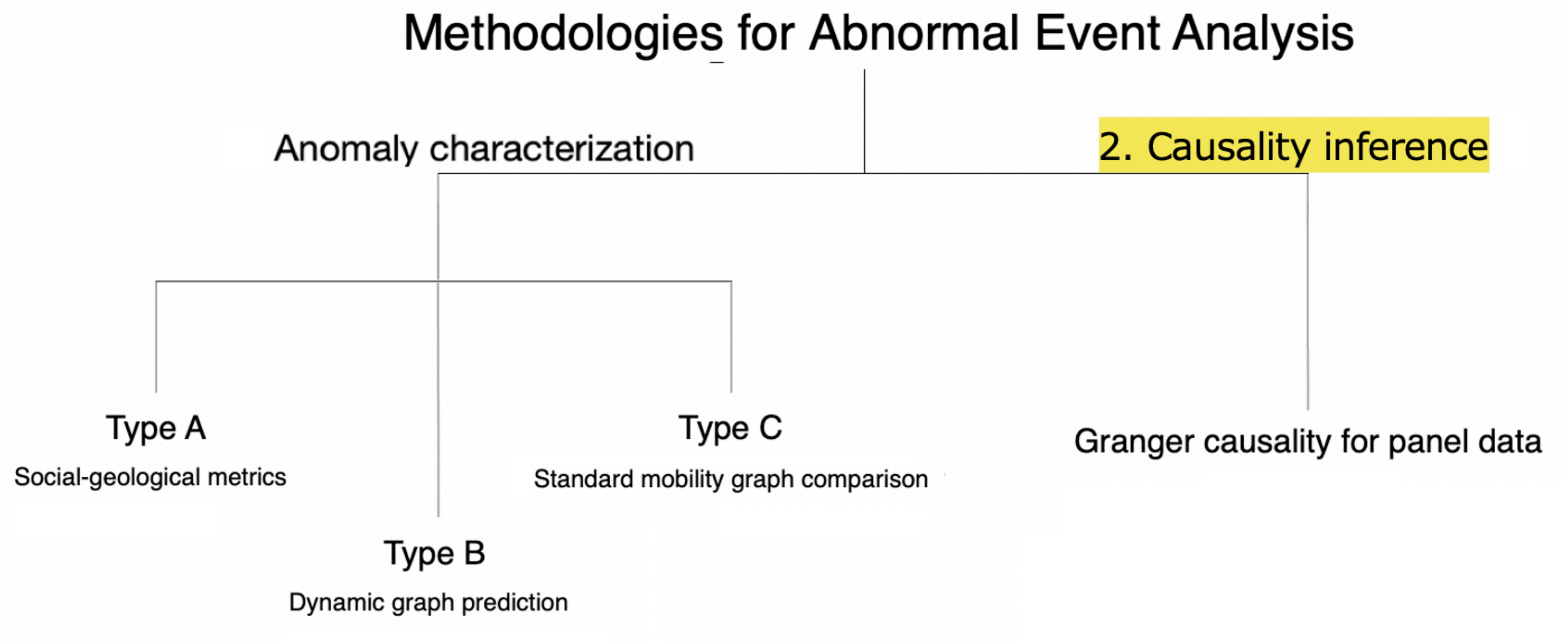
$$F_f(G_1, G_2, \ldots, G_T) = [z_1, z_2, \ldots, z_T]$$

Anomaly scores over time

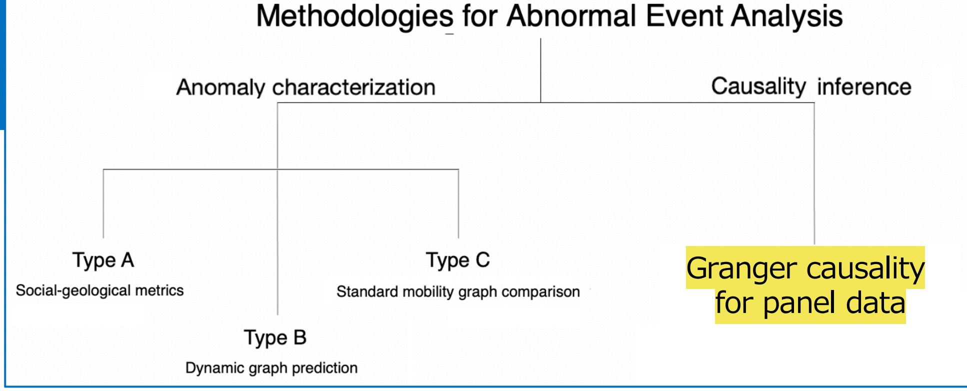
Compute a standard mobility graph
+
Local Outlier Factor

C. Lin, Q. Zhu, S. Guo, Z. Jin, Y.-R. Lin, and N. Cao. "Anomaly detection in spatiotem- poral data via regularized non-negative tensor analysis". In: *Data Mining and Knowledge Discovery* 32.4 (2018), pp. 1056–1073.

Methodologies: Causality inference



Granger causality



The Granger causality is a classical and intuitive causality concept for the pair of time series variables

$$Y_t = \alpha + \sum_{k=1}^{L} \beta_k Y_{t-k} + \sum_{k=1}^{L} \gamma_k X_{t-k} + \epsilon_t$$

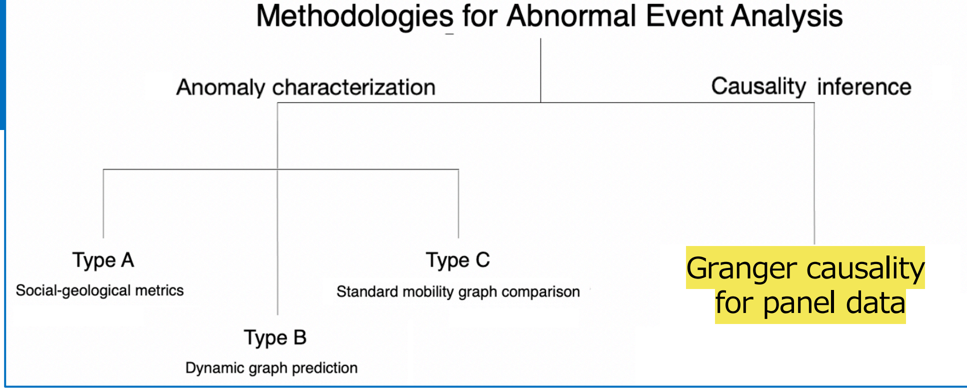
t : time notation

If discarding *Xt* reduces the predictive power regarding *Yt*, then *Xt* Granger-causes *Yt*.

H0: Xt does not Granger-causes *Yt*
H1: alternative hypothesis: *Xt* Granger-causes *Yt*

C. W. Granger. "Investigating causal relations by econometric models and cross-spectral methods". In: *Econometrica: journal of the Econometric Society* (1969), pp. 424–438.
E.-I. Dumitrescu and C. Hurlin. "Testing for Granger non-causality in heterogeneous panels". In: *Economic modelling* 29.4 (2012), pp. 1450–1460.

Granger causality for panel data



Panel data example

Country	Year	Y	X
1	2016	124	234
1	2017	217	52
1	2018	312	8
2	2016	178	352
2	2017	199	116
2	2018	308	78

If discarding *Xt* reduces the predictive power regarding *Yt*, then *Xt* Granger-causes *Yt*.

$$Y_{i,t} = \alpha + \sum_{k=1}^{L} \beta_k Y_{i,t-k} + \sum_{k=1}^{L} \gamma_k X_{i,t-k} + \epsilon_{i,t}$$

i : node number

H0: there is no granger causality

H1: alternative hypothesis: there is Granger causality from Xt to Yt in at least one region.

C. W. Granger. "Investigating causal relations by econometric models and cross-spectral methods". In: *Econometrica: journal of the Econometric Society* (1969), pp. 424–438.
E.-I. Dumitrescu and C. Hurlin. "Testing for Granger non-causality in heterogeneous panels". In: *Economic modelling* 29.4 (2012), pp. 1450–1460.

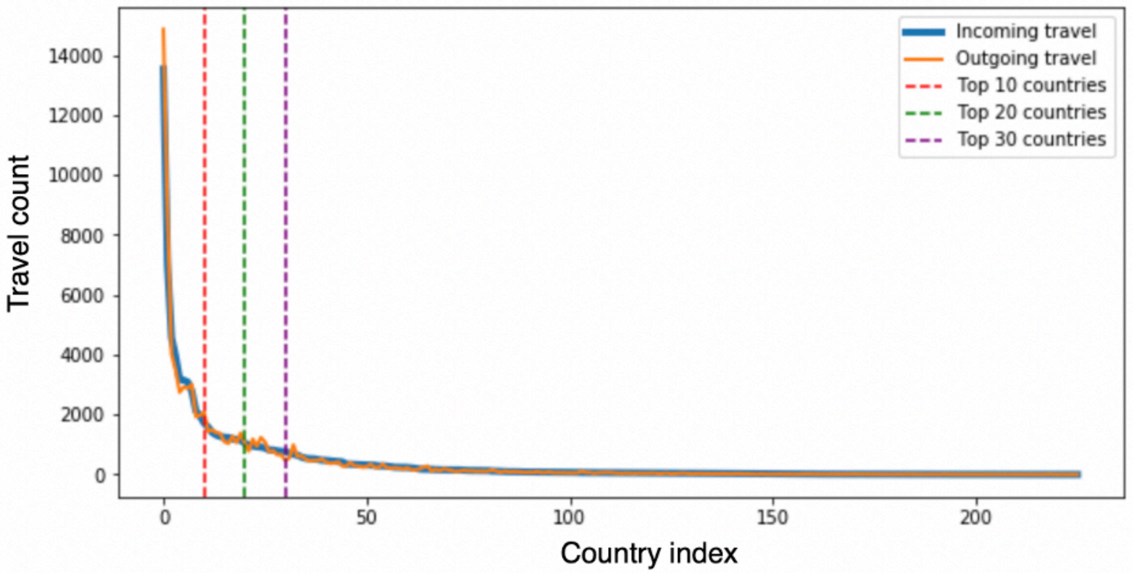
Contents

Heterogeneous mobility systems

Mobility system	Twitter users' travel	COVID-19 patients movement
Scale	International travels	Urban mobility in Seoul
Complexity	High	Low
Trend	Yes	No
Seasonality	Yes	Yes
Hub node	Yes(USA)	No
Graph density	Dense (x 6)	Sparse
Time interval	2016–2019	Jan–June.2020

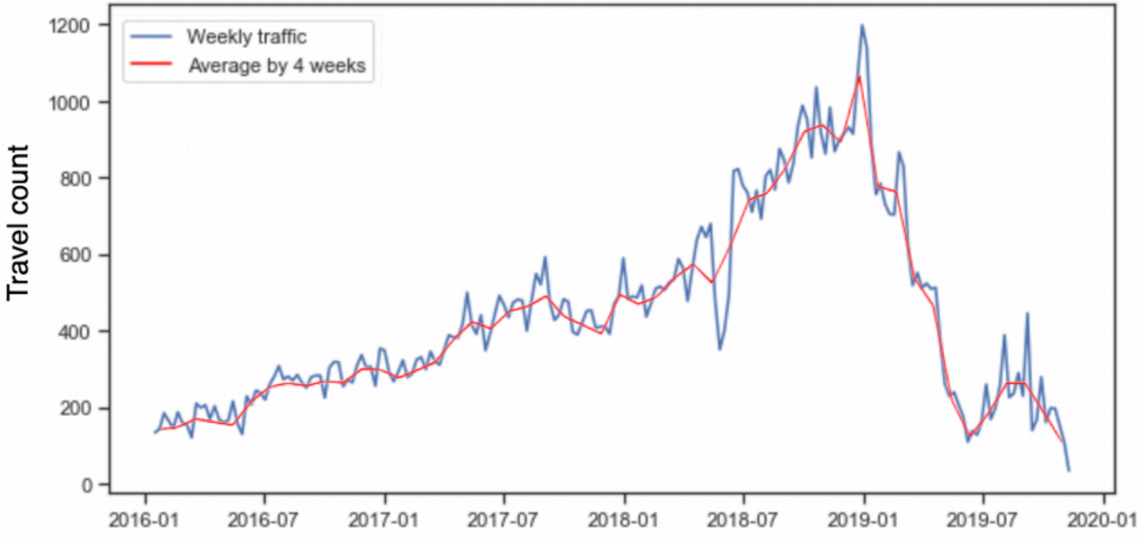
Travel of Twitter Users: EDA

(a) Total traffic volume by country

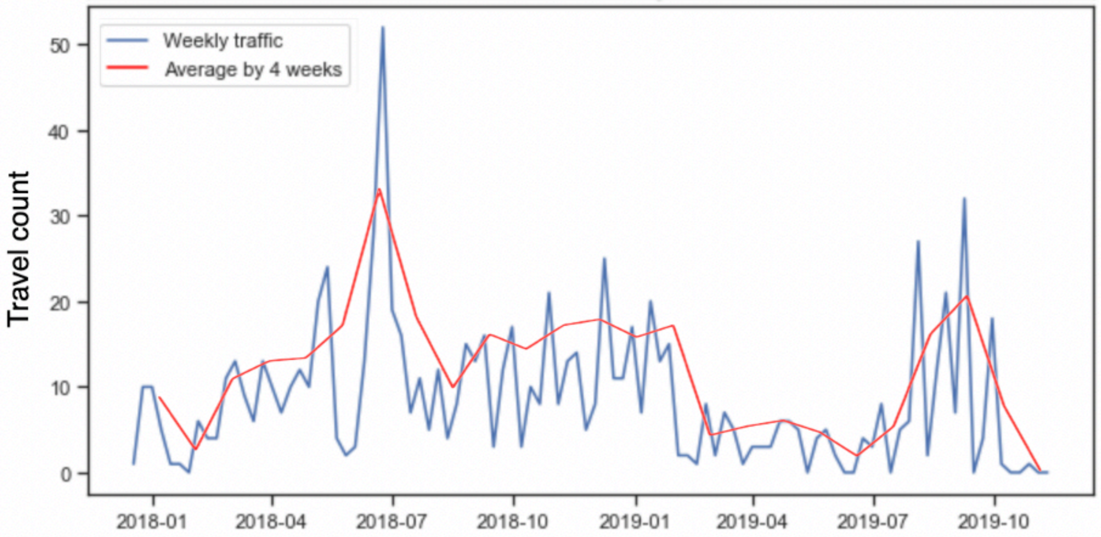


(b) Traffic on world map

(c) Global traffic volume by date

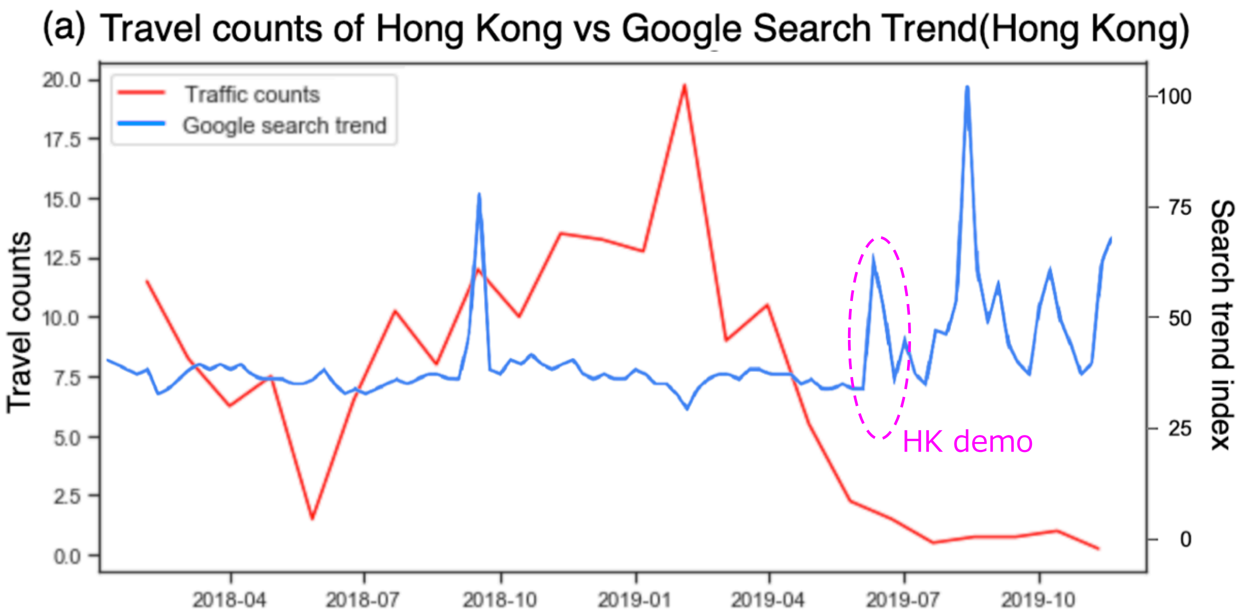


(d) Traffic volume by date in Russia

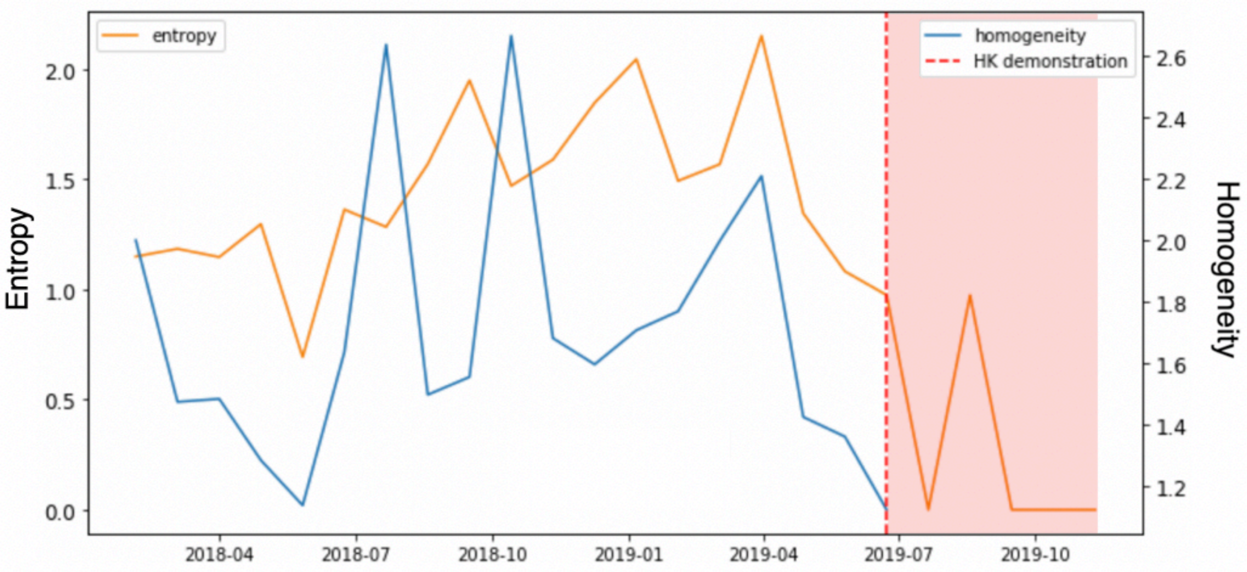


Events description

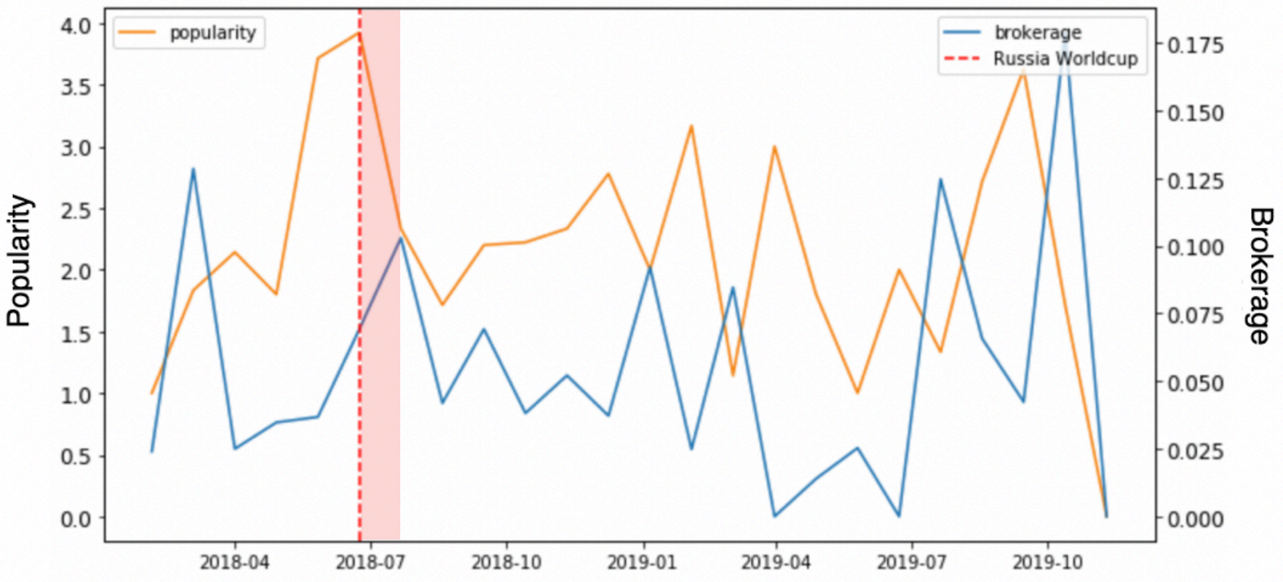
We are mainly interested in two events: Hong Kong protest(2019) and Russia World Cup(2018)



(a) Entropy and Homogeneity of Hong Kong

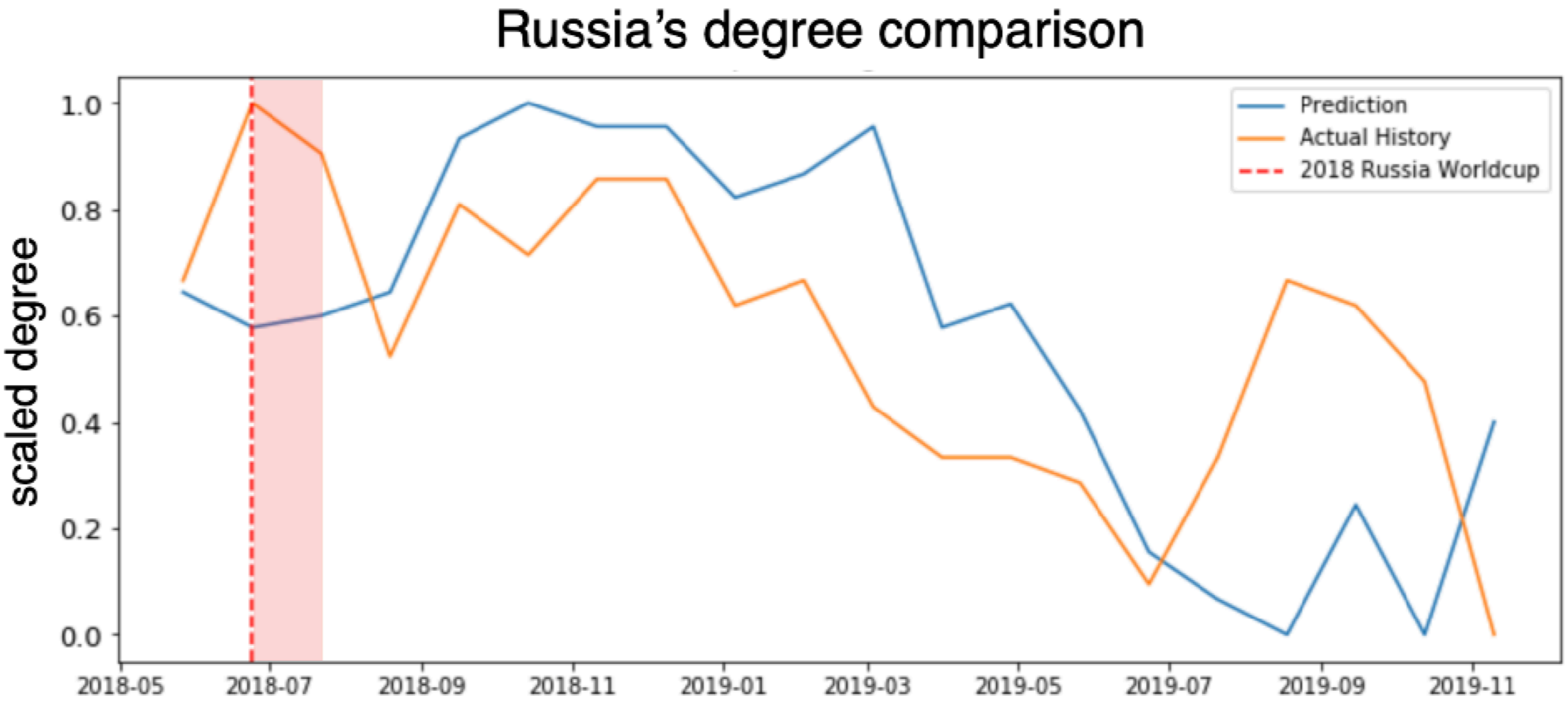


(b) Popularity and Brokerage of Russia



*The top 30 countries are considered for Hong Kong's metrics, while 40 countries are uesd for Russia.

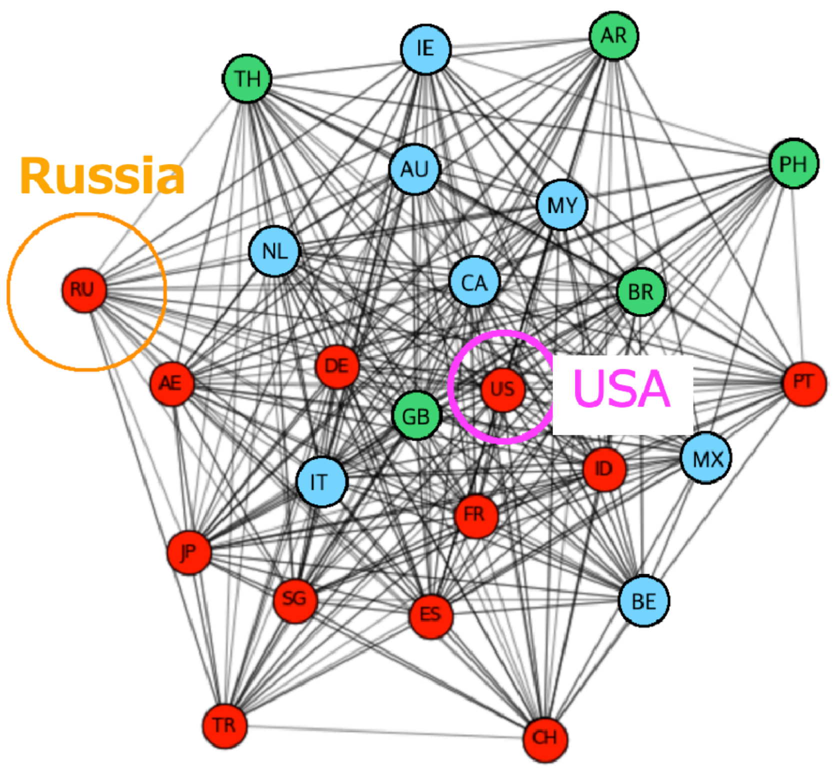
Dynamic graph prediction



*Russia's degree is scaled from 0 to 1

Dynamic graph prediction

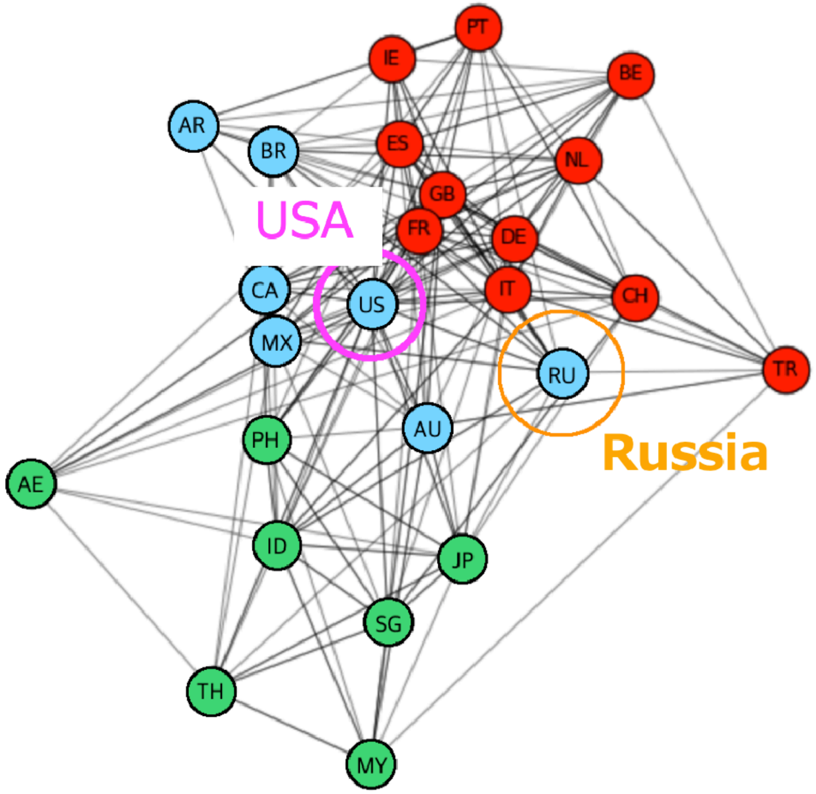
(a) The predicted graph
(24.June ~ 21.July in 2018)



In prediction
Russia is located on the outside of the graph.

Russia's between centrality
0.0045

(b) The historical graph
(24.June ~ 21.July in 2018)

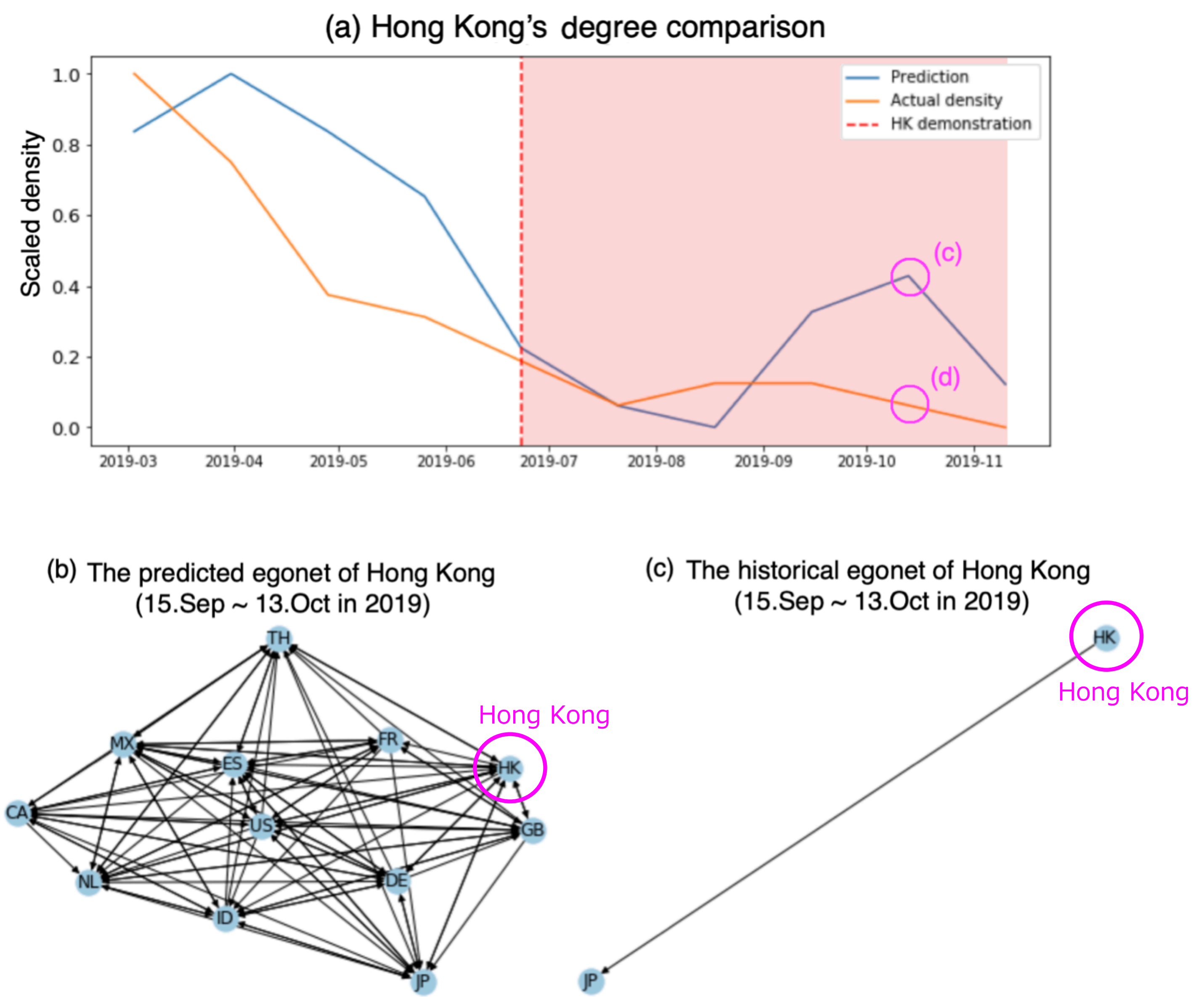


In actual graph
Russia was closer to the hub node(US)

Russia's actual between centrality
0.0195

*Top 25 countries used. Colors mean community clusers by Leiden algorithm. The red cluster is the biggest, followed by the light blue and green. Graph layout is by Fruchterman- Reingold force-directed algorithm.

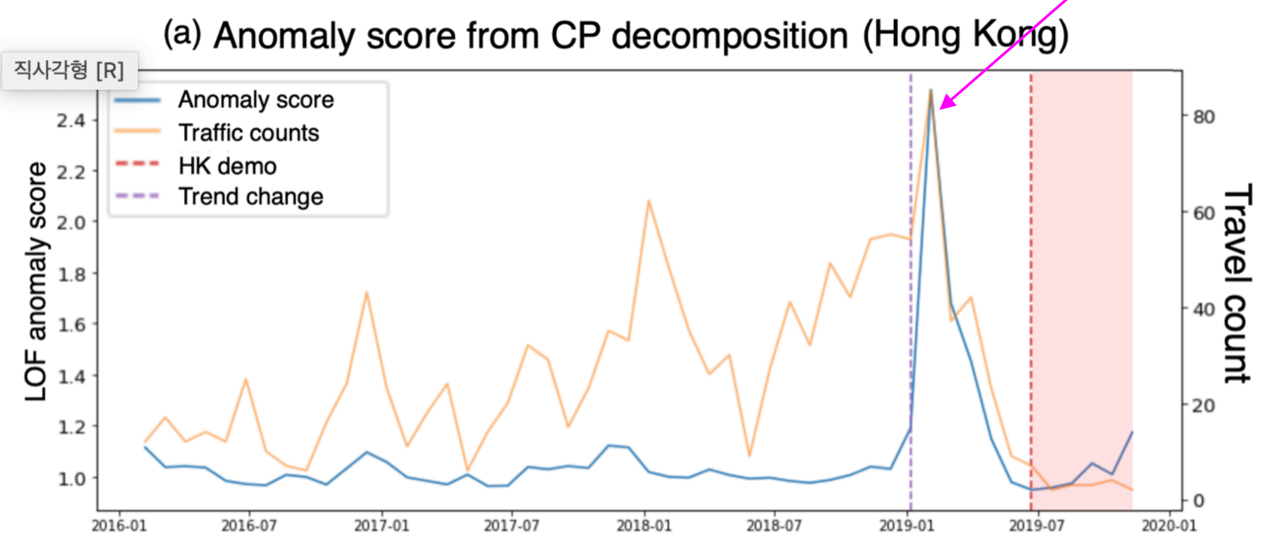
Dynamic graph prediction

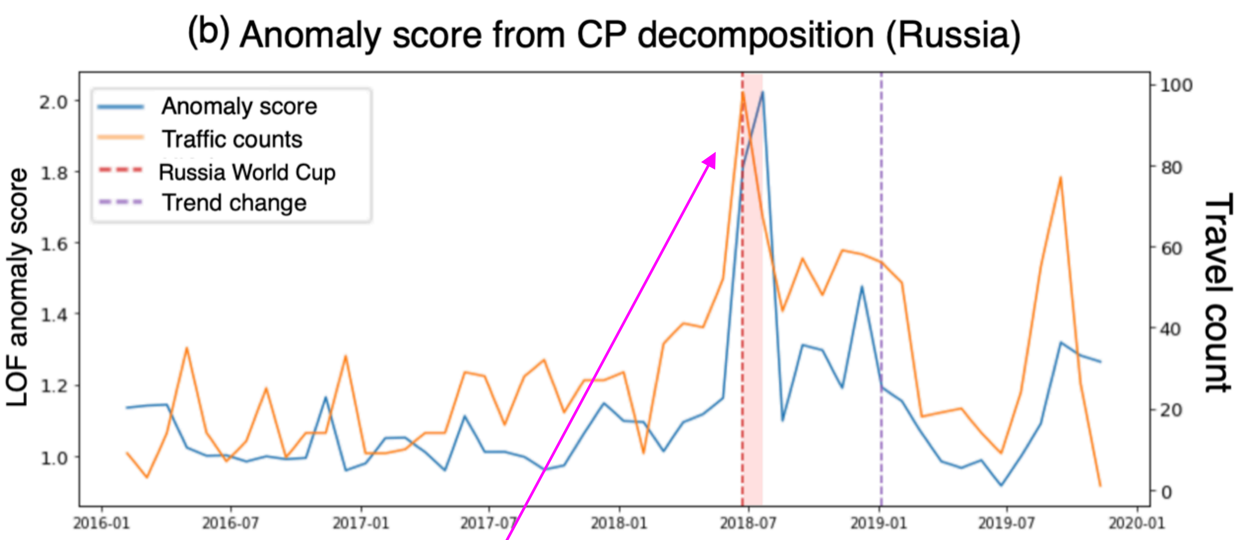


*Egonet refers to a group of nodes directly connected to Hong Kong.

Standard mobility graph comparison

Assign the highest anomaly score to the trend-changing point, not the event time





Successful anomaly evaluation despite the confusing trend and seasonality.

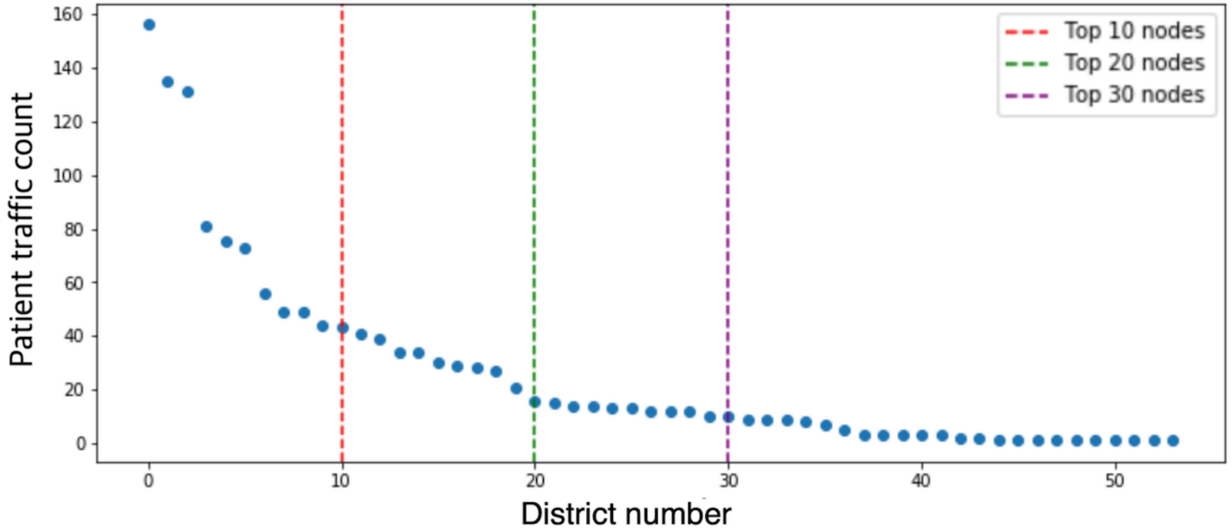
Contents

Heterogeneous mobility systems

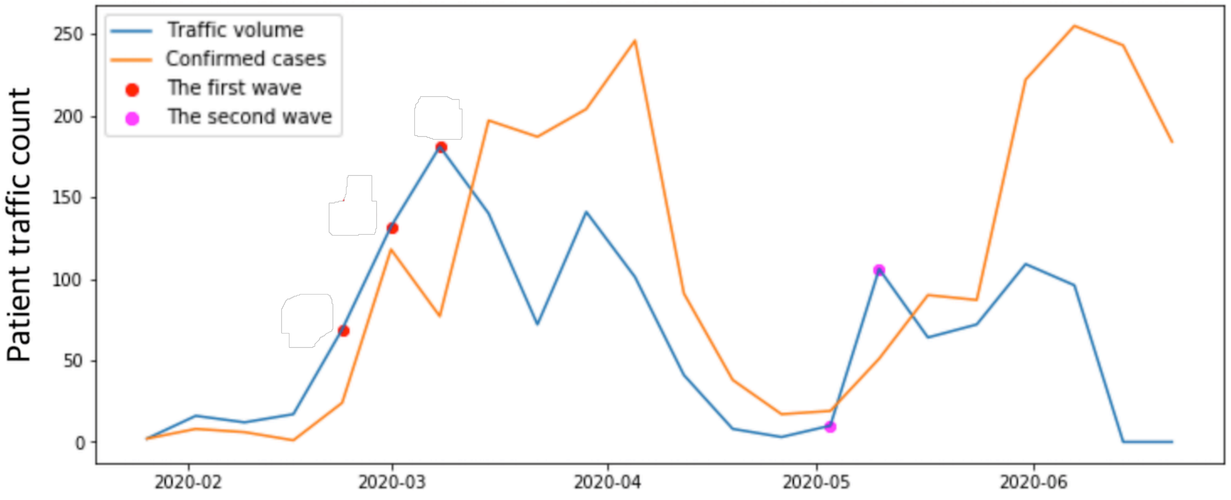
Mobility system	Twitter users' travel	COVID-19 patients movement
Scale	International travels	Urban mobility in Seoul
Complexity	High	Low
Trend	Yes	No
Seasonality	Yes	Yes
Hub node	Yes(USA)	No
Graph density	Dense	Sparse (x 1/6)
Time interval	2016–2019	Jan–June.2020

COVID-19 carriers mobility: EDA

(a) Total traffic by urban areas

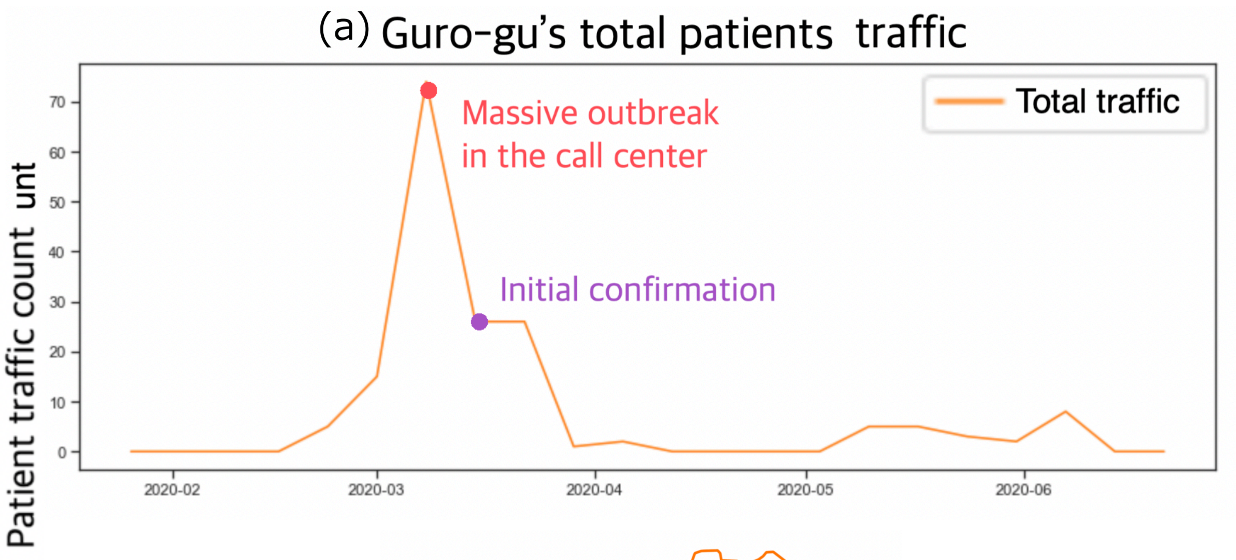


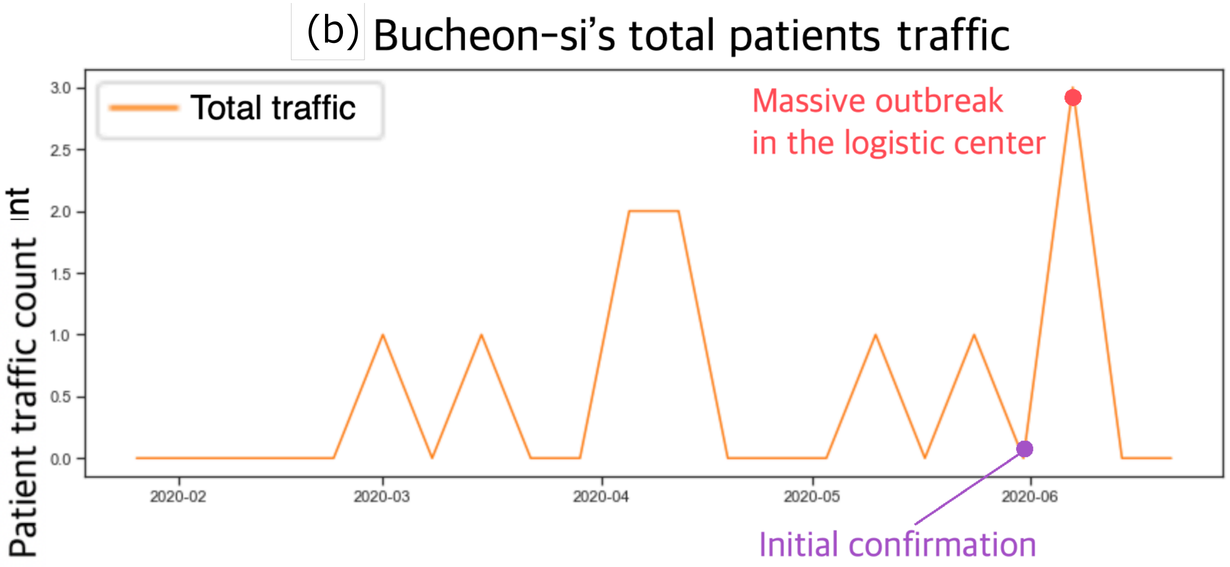
(b) Patients traffic vs confirmed cases

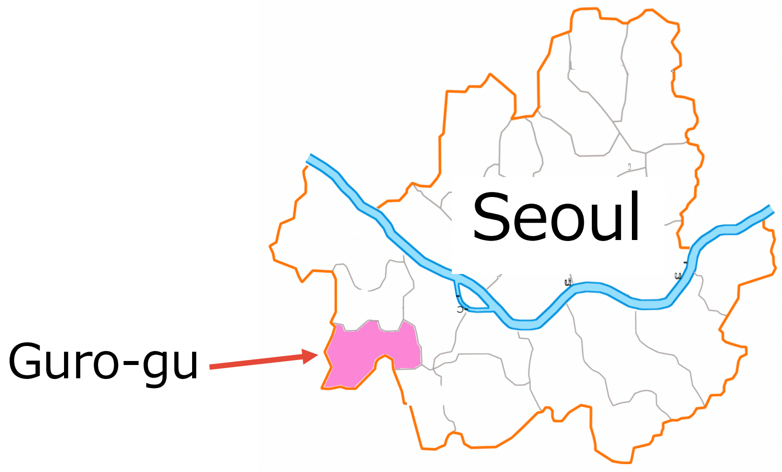


Events description

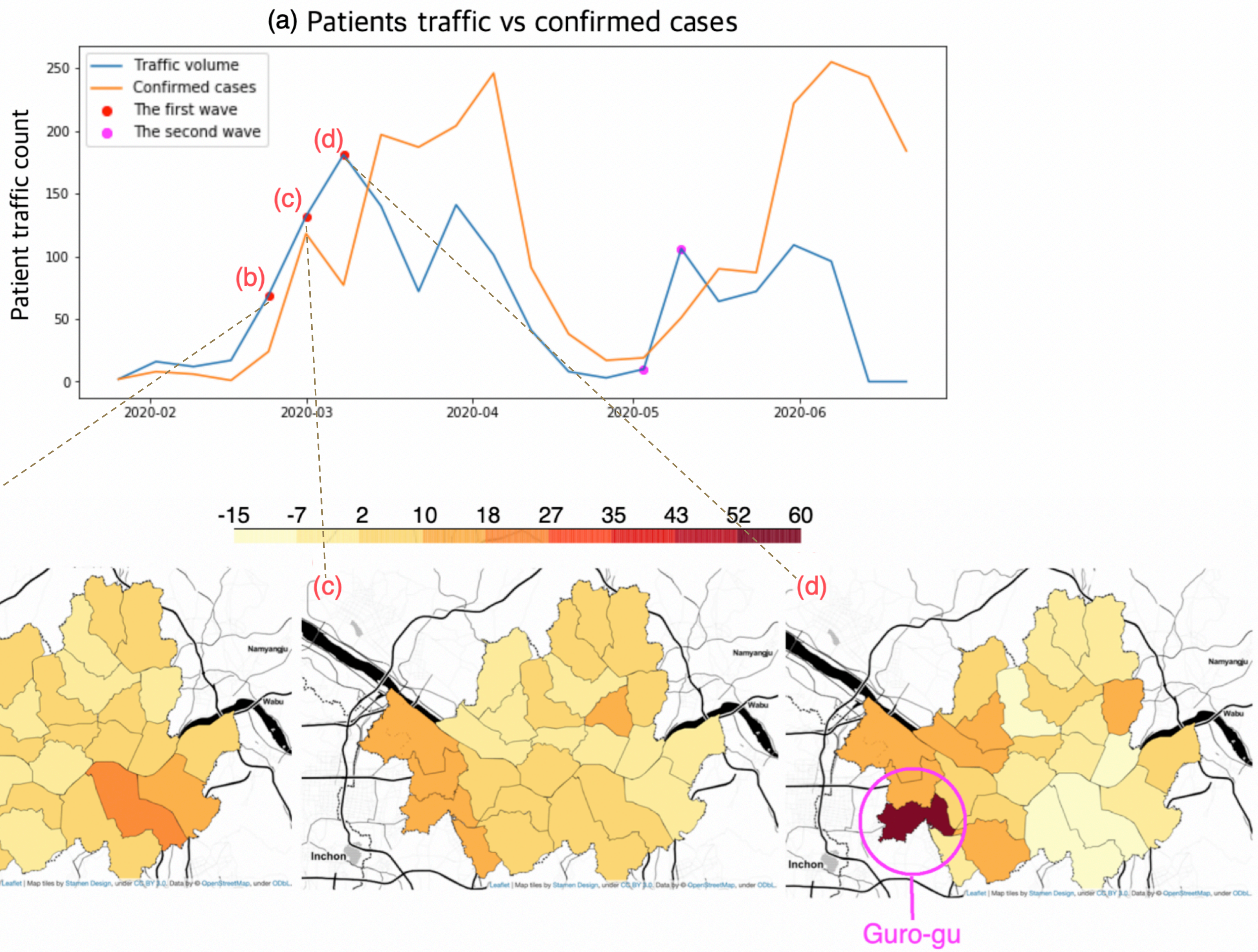
Two massive outbreaks in Guro-gu and Bucheon-si



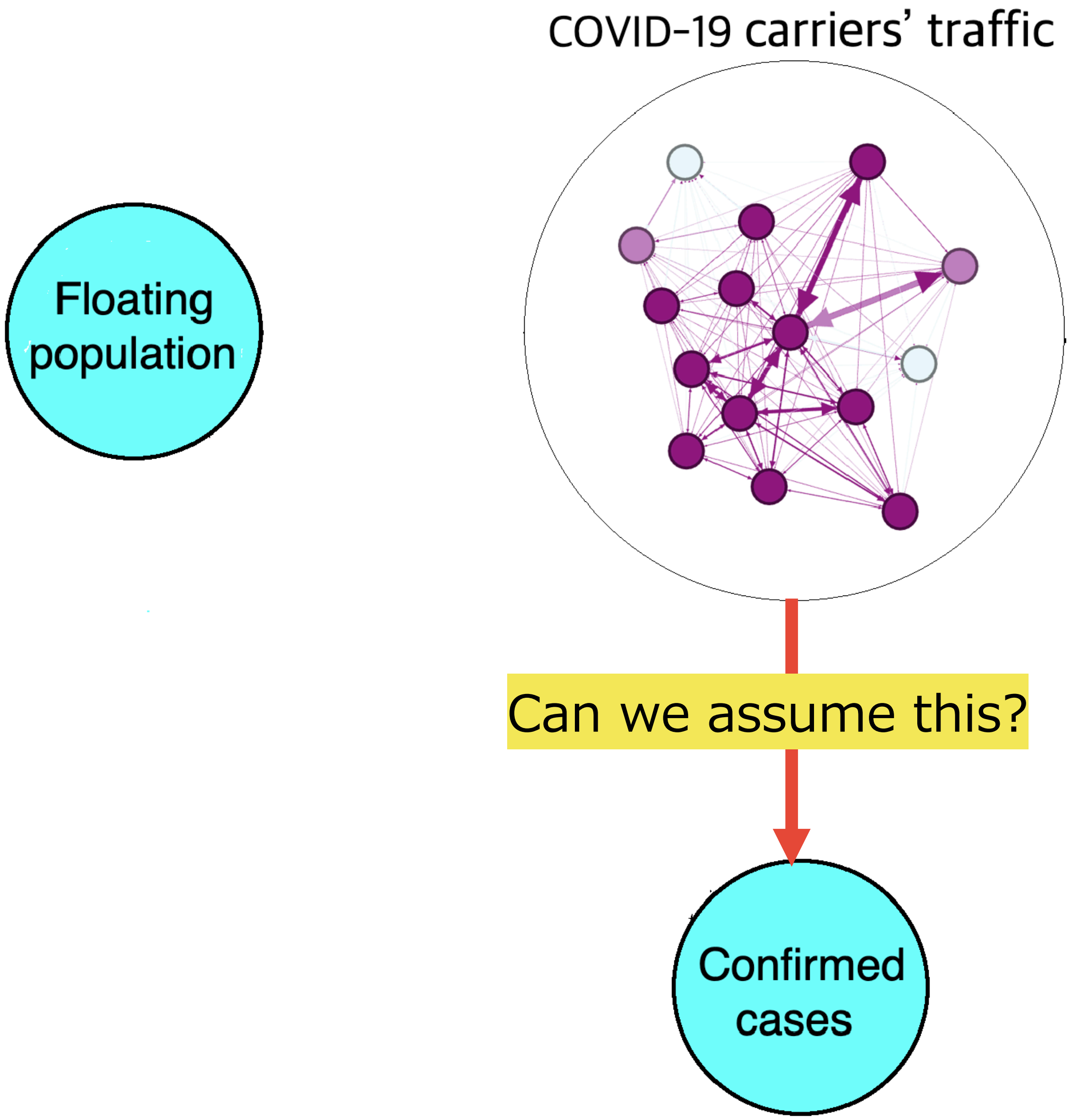




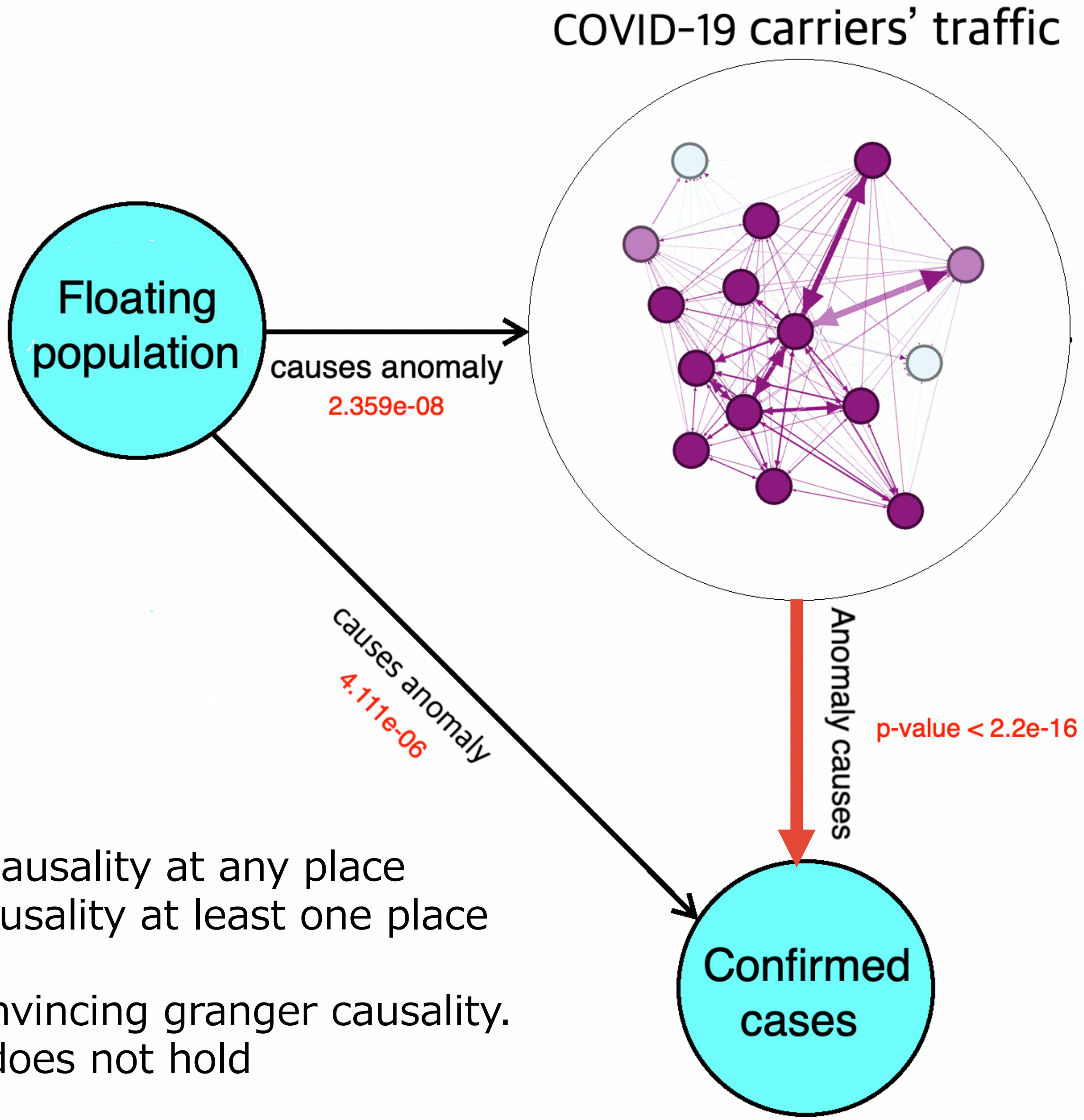
Events description



Causal inference

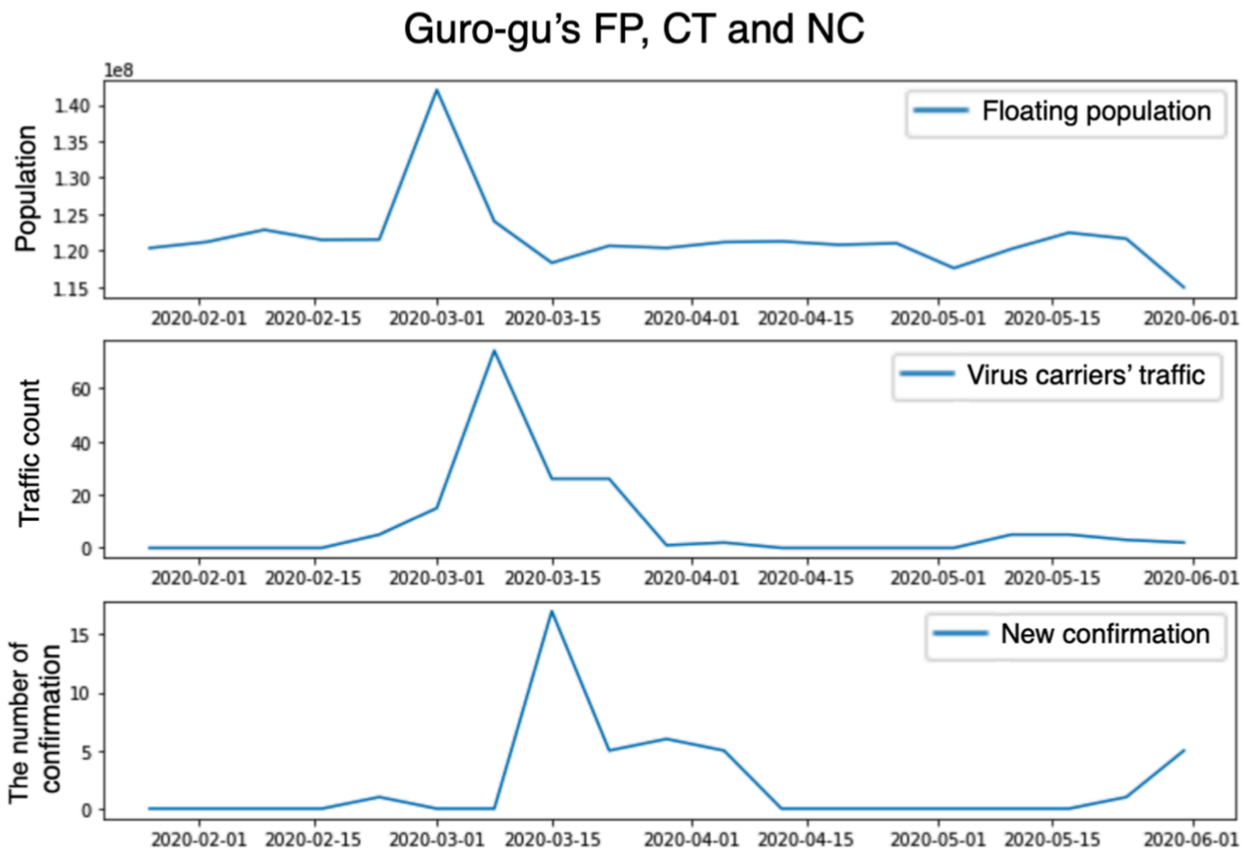


Causal inference



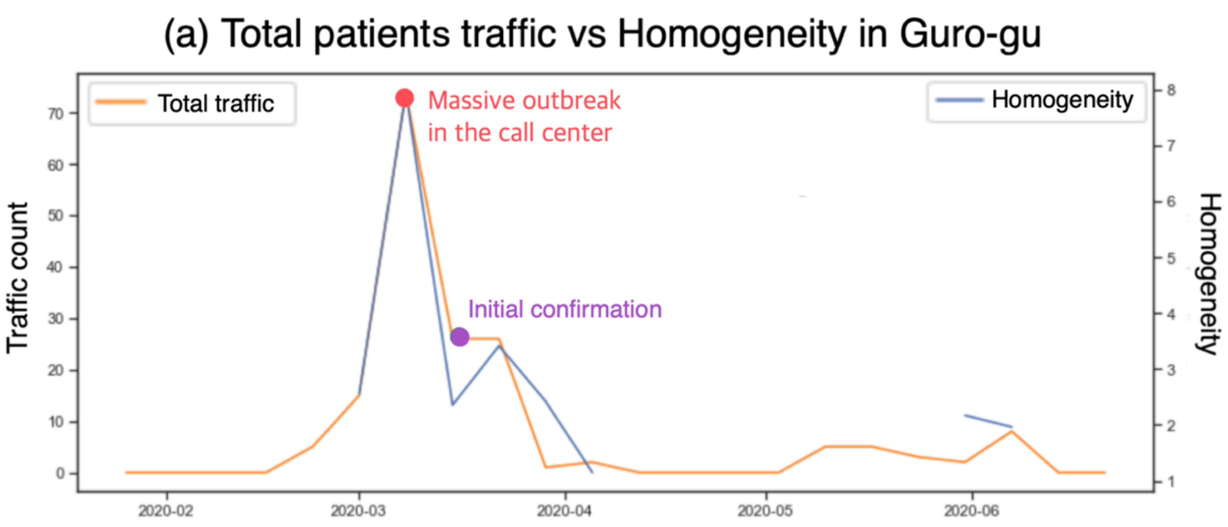
H0: There is no granger causality at any place
H1: There is a granger causality at least one place

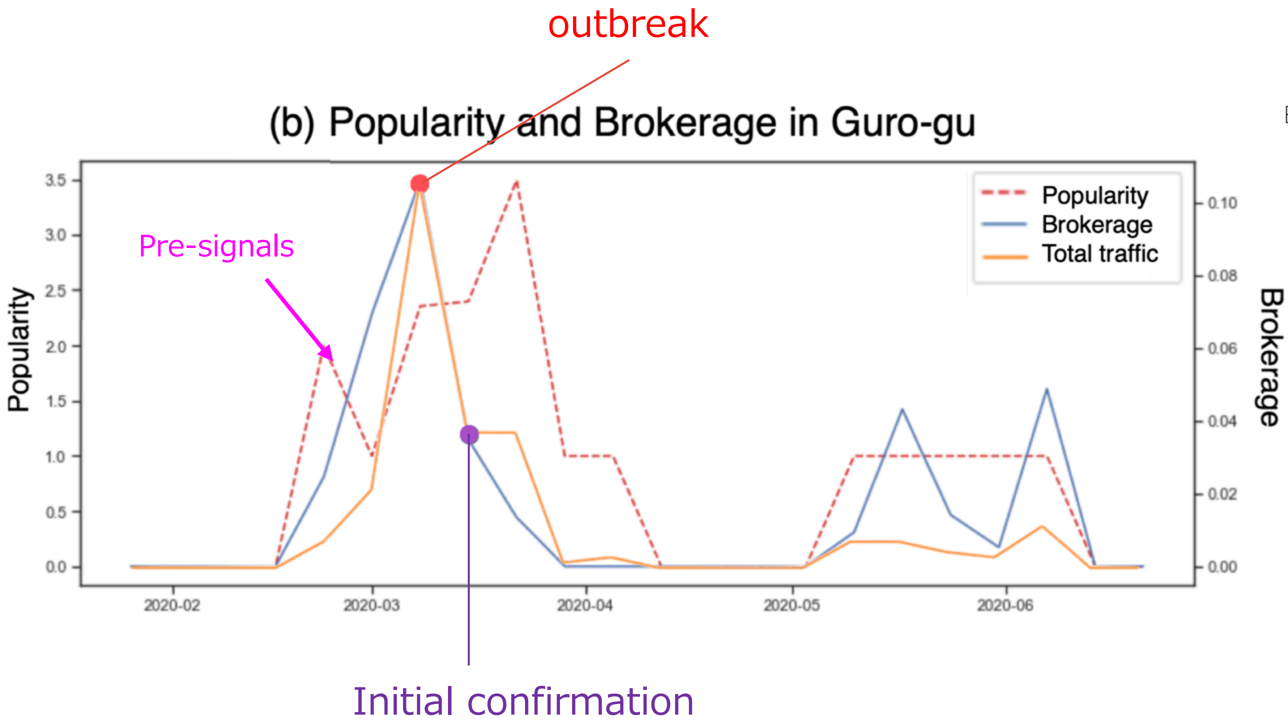
smaller p-value: more convincing granger causality.
* The opposite causality does not hold



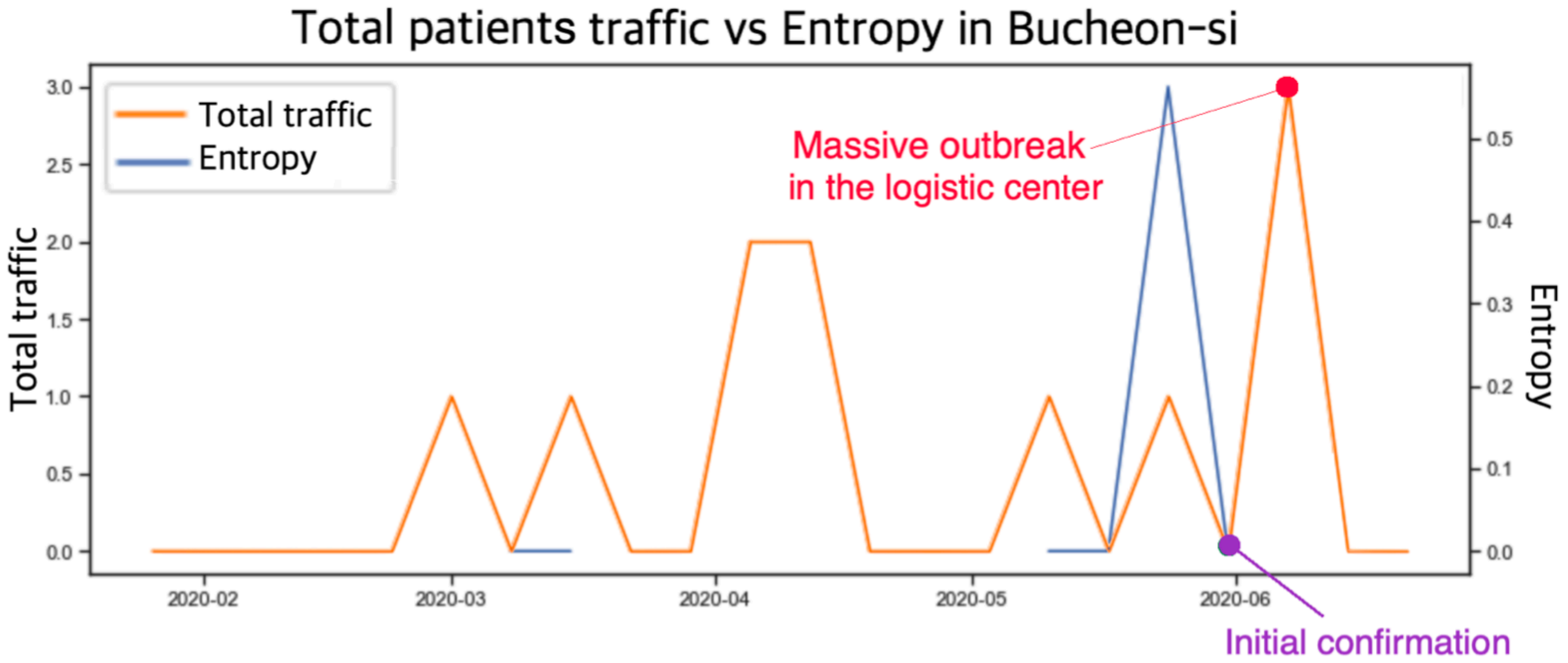
Cause	Effect	Proportion in Seoul
Floating population	Virus carriers' traffic volume	24/25
Floating population	The number of confirmation	22/23
Virus carriers' traffic volume	The number of confirmation	23/23

Social-geological metrics

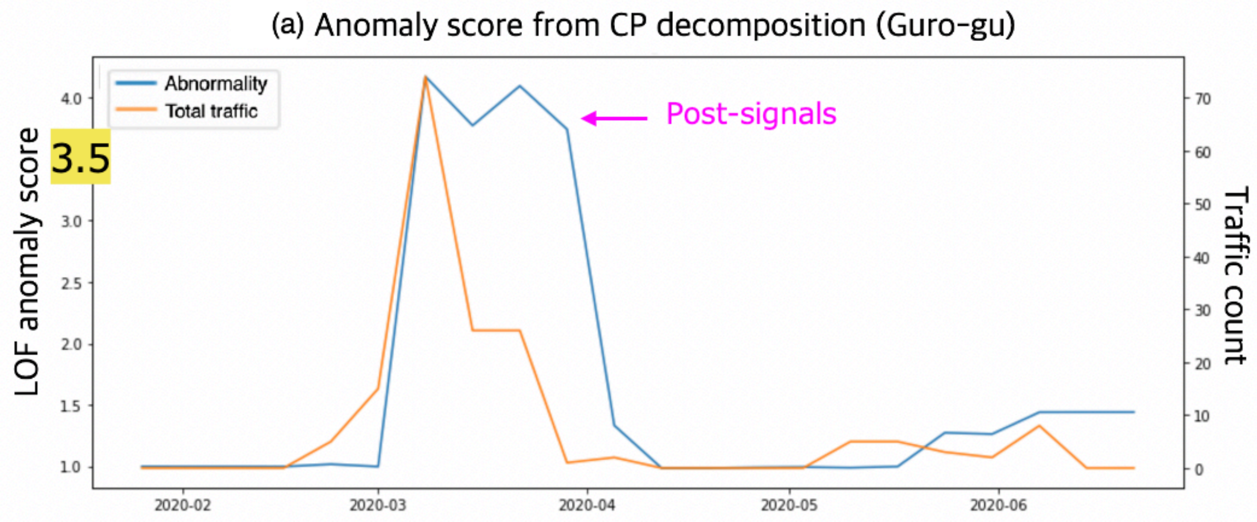


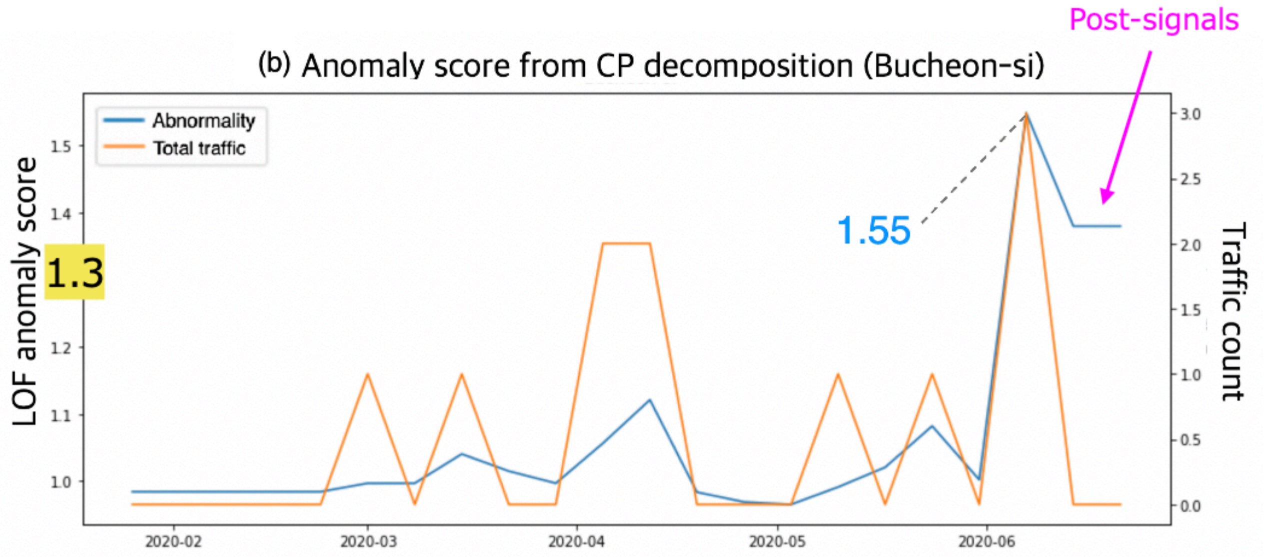


Social-geological metrics



Standard mobility graph comparison





Contents

Conclusion

Research Question

How do social events
affect and are affected by anomalies
in a spatio-temporal human mobility network?

Methodologies

Anomaly characterization

Causal inference

Case Studies

Twitter users' global travel

COVID-19 patients' urban mobility

Conclusion

Research Question

How do social events
affect and are affected by anomalies
in a spatio-temporal human mobility network?

Case study1

Events	HK Demonstration	Russia World Cup
Exogenous event type	Political protest	Sport event
Date	9.Jun.19	15.Jun.18
Traffic volume	decrease(0)	increase(24)
Mobility anomaly types	A,B	A,B,C
Impact period on mobility	Jun-Nov.2019	Jun-Oct.2018
Total traffic	min 0	maximum 24
brokerage	-	0.065
Popularity	-	4
Entropy	0	-
Homogeneity	none	-
Difference to the prediction	Fewer neighbors -	Closer to the hub node(USA) Centrality rise by 4.3 times

Case study2

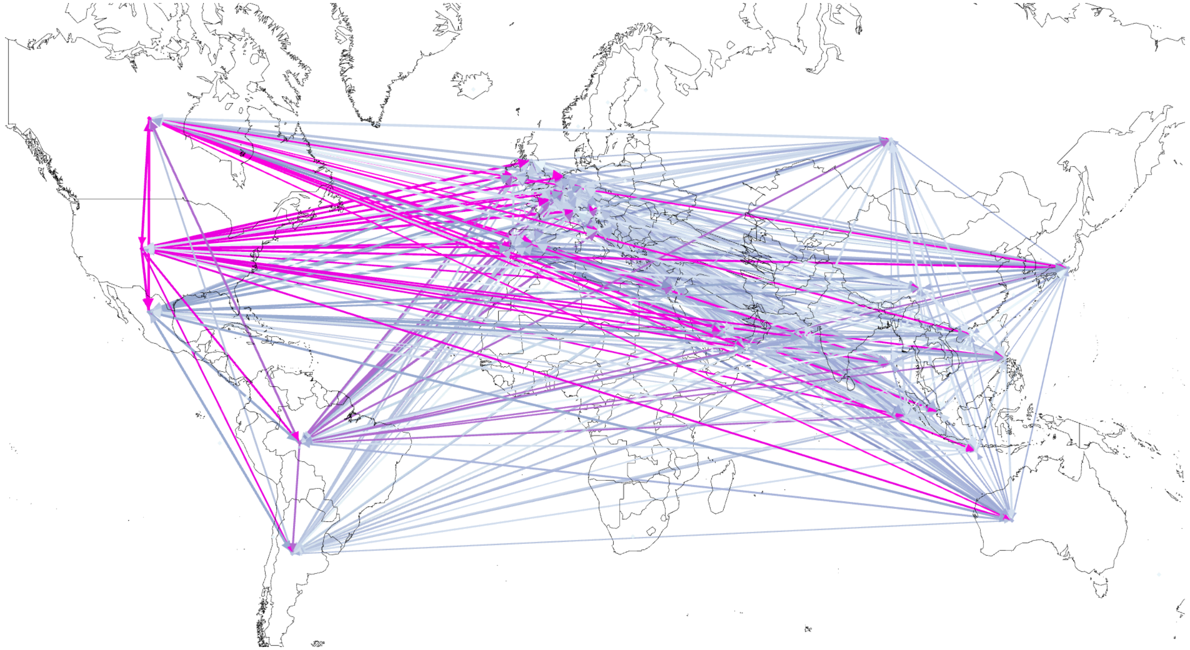
Events		COVID-19 massive outbreaks	
Location		Guro-gu	Bucheon-si
Patients traffic peak		01.Mar-07.Mar	31.May-06.June
Initial confirmation		8.May	24.May
Caused by		Patient traffic & Floating population	
Total traffic		70	-
Pre-signals	brokerage	0.1	-
	Popularity	3.5	-
	Entropy	-	0.5
Post-signal by standard graph comparison		3 weeks	>= 2weeks

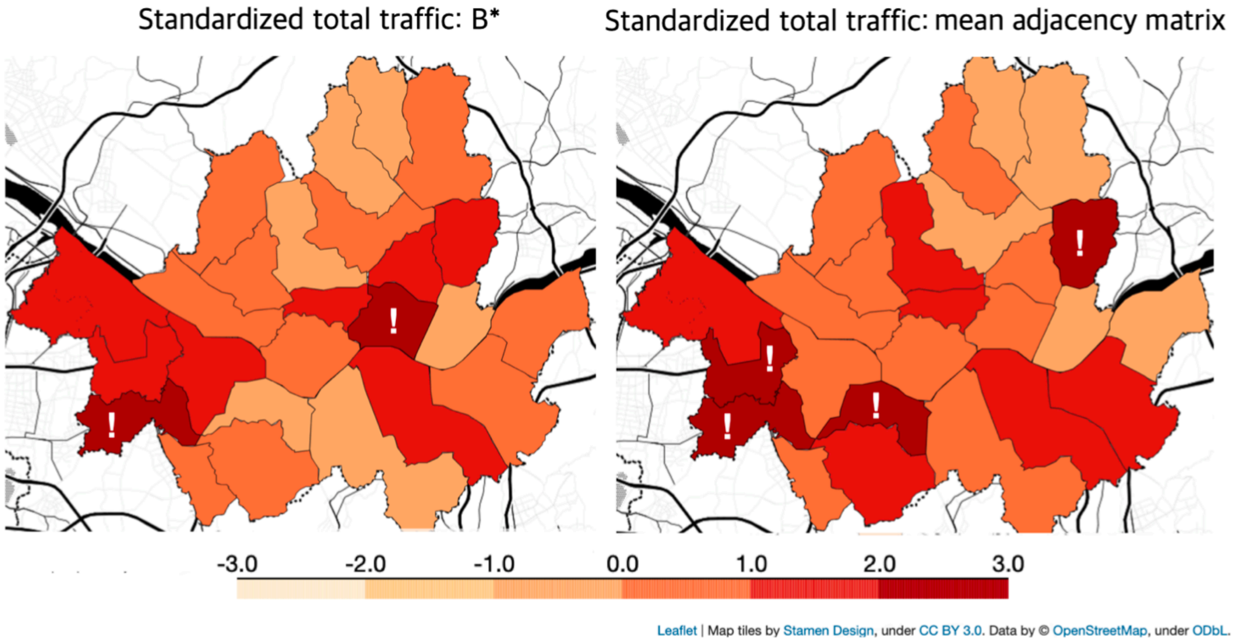
Questions?

Tools:

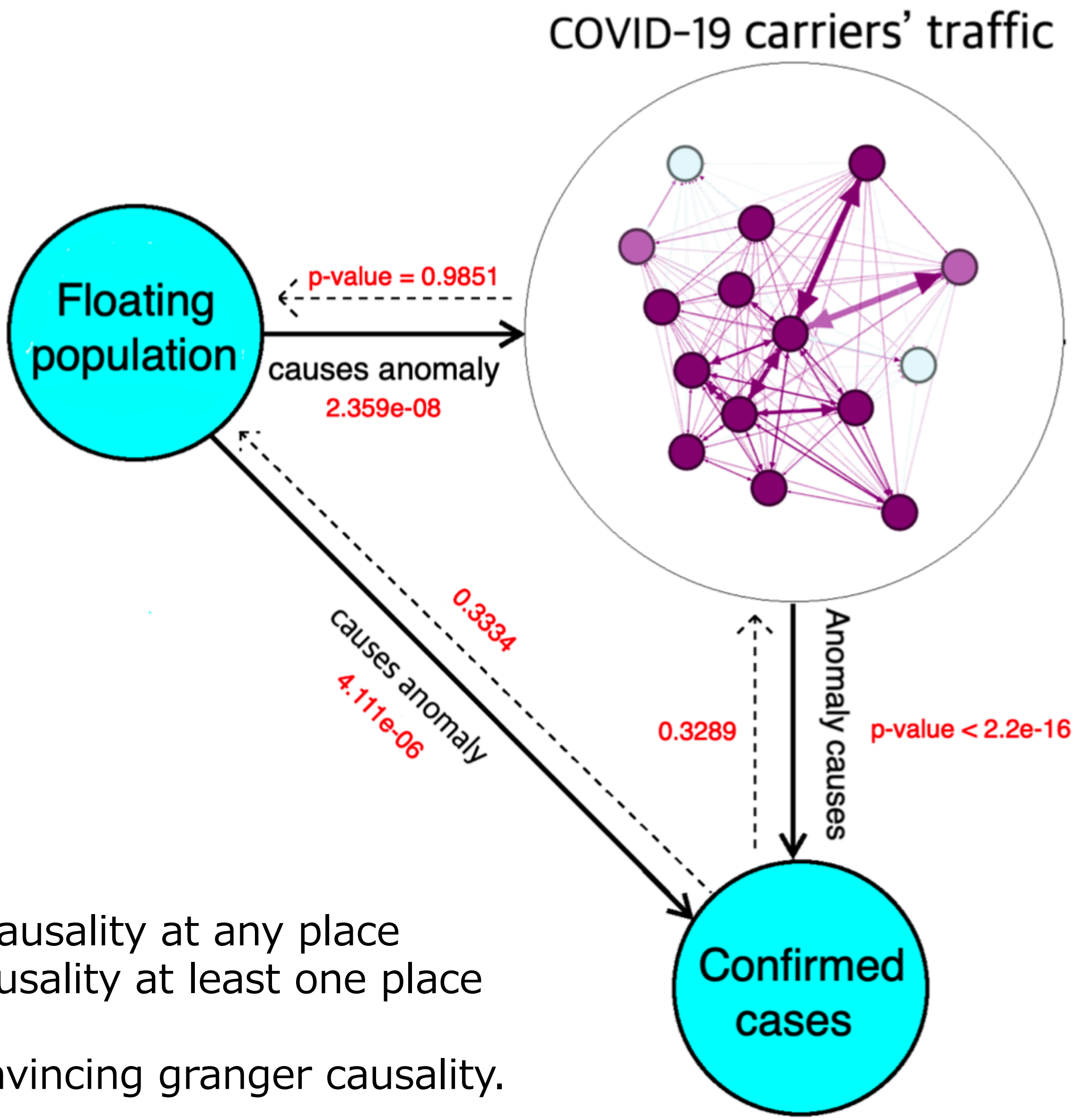
R python NetworkX Network Analysis in Python Gephi makes graphs handy CD·LIB PyTorch Folium statsmodels

Standard mobility graph comparison





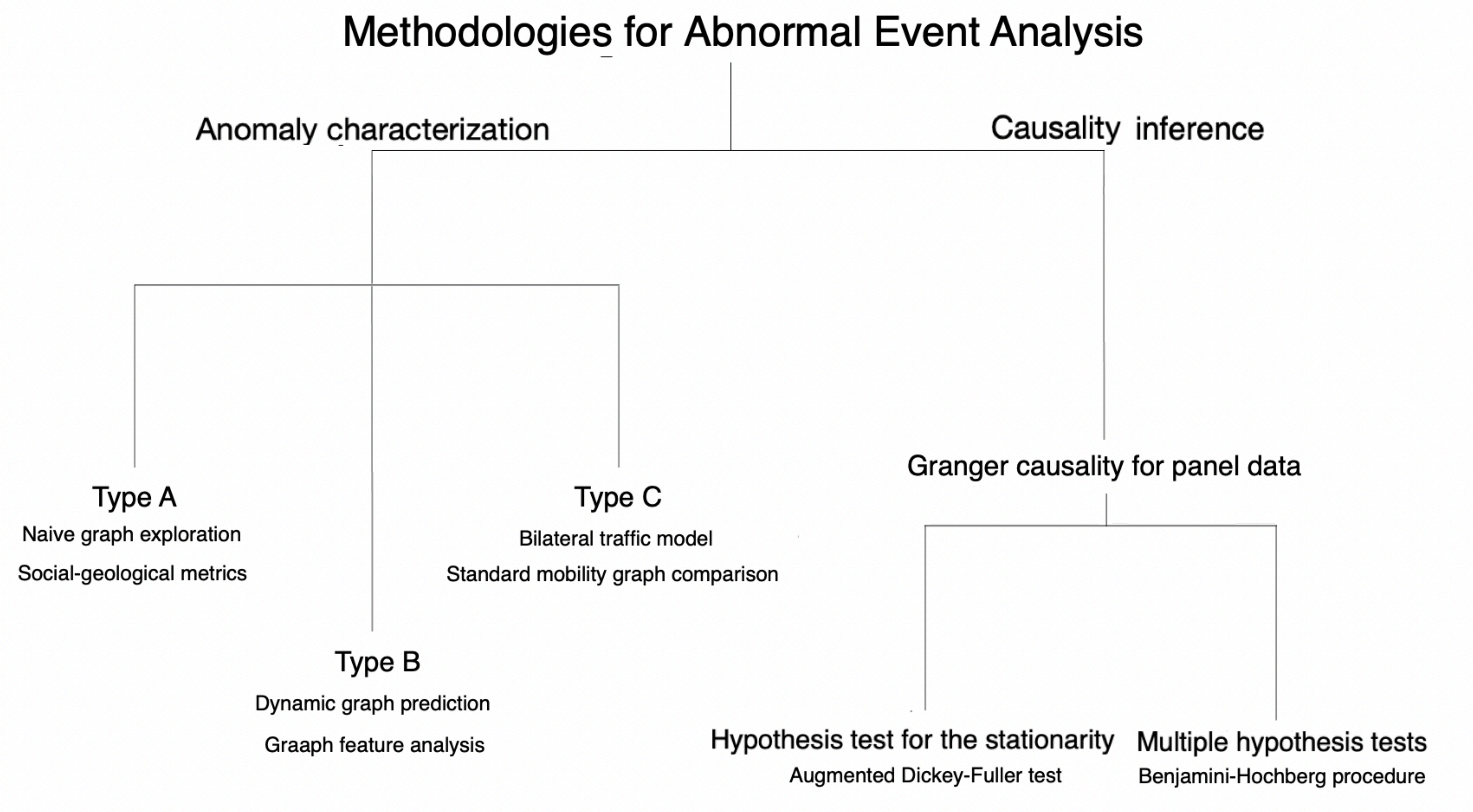
Causal inference



H0: There is no granger causality at any place
H1: There is a granger causality at least one place

smaller p-value: more convincing granger causality.

Methodologies: Anomaly characterization



Bilateral traffic model

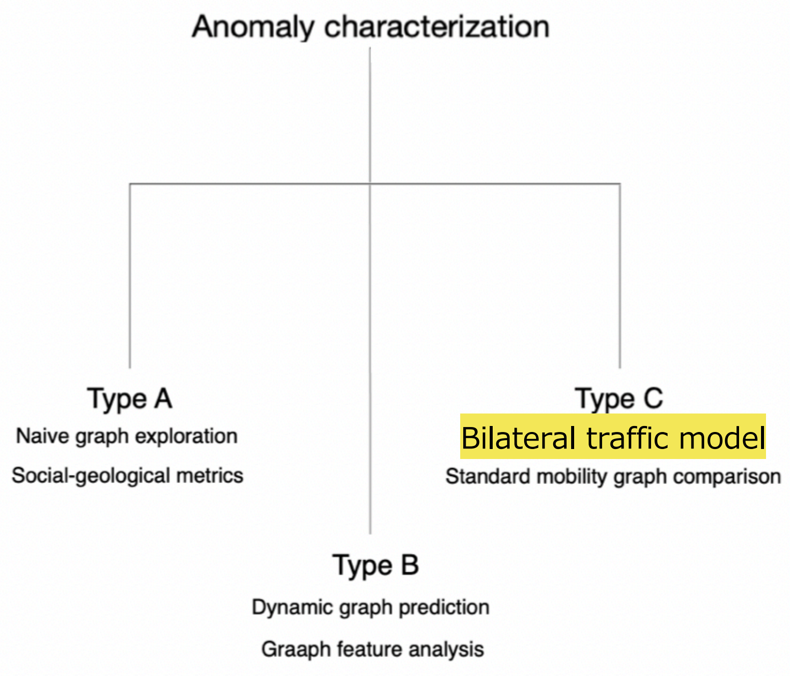
Gravity model: $$T_{ij} = k\frac{E_i^{\alpha} A_j^{\beta}}{d_{ij}^{\gamma}}$$

E_i : origin node's emission(outgoing traffic)
A_j : attraction(incoming traffic) of a destination node
d_{ij} : distance between i and j

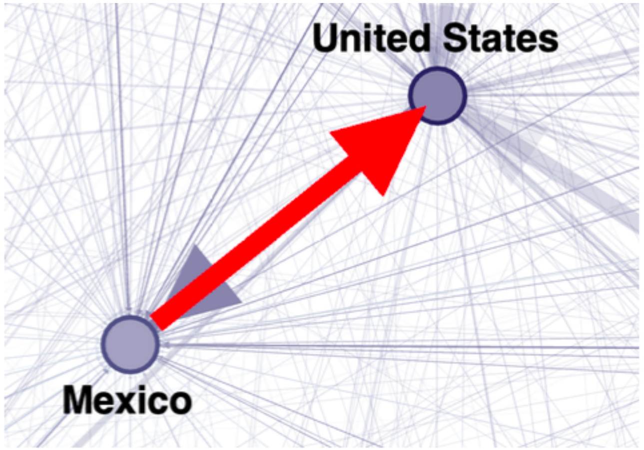
$$log(T_{ij}) = log(k) + \alpha \cdot log(E_i) + \beta \cdot log(A_j) - \gamma \cdot log(d_{ij})$$

Generalized linear modeling:

$$g(log(T_{ij})) = log(k) + \alpha \cdot log(E_i) + \beta \cdot log(A_j) - \gamma \cdot log(d_{ij})$$

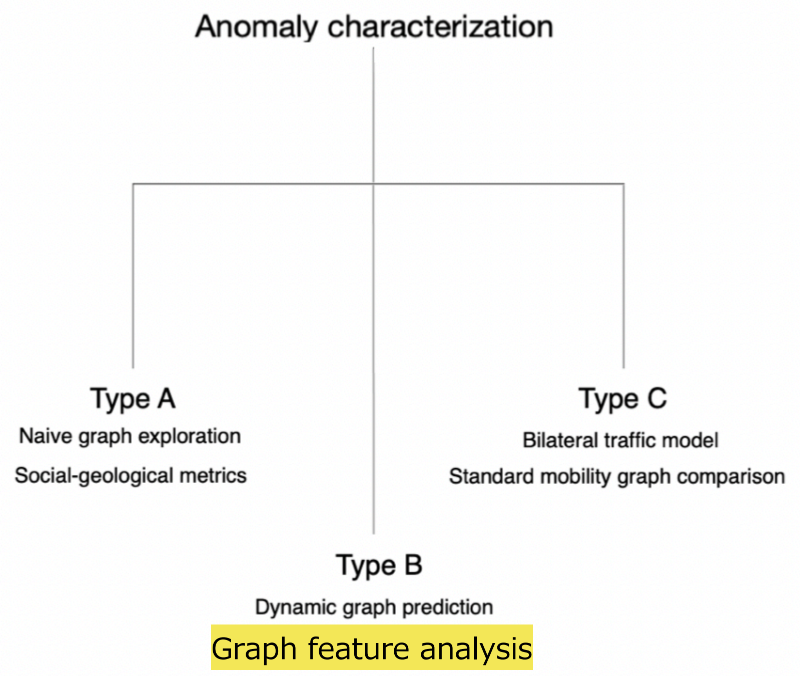


Bilateral traffic example
from Mexico to United States



A. S. Fotheringham and C. Brunsdon. "Local forms of spatial analysis". In: *Geographical analysis* 31.4 (1999), pp. 340–358.
T. M. Oshan. "A primer for working with the Spatial Interaction modeling (SpInt) module in the python spatial analysis library (PySAL)". In: *Region* 3.2 (2016), R11–R23.

Graph feature analysis



$$F_f(G_1, G_2, \ldots, G_T) = [z_{2W}, z_{2W+1}, \ldots, z_{T-1}, z_T]$$

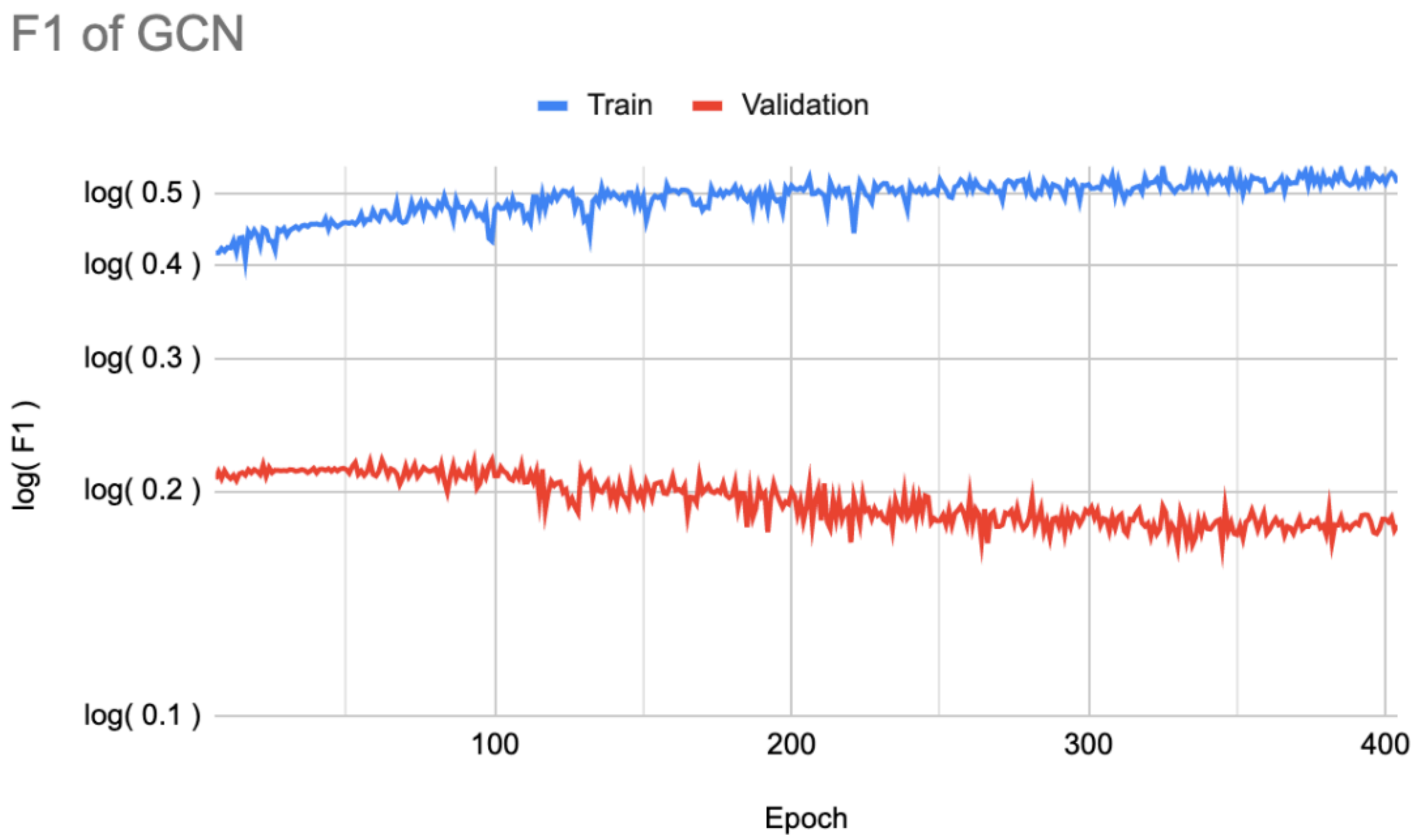
Z_t: Anomaly scores

G_t: one node's neighborhood graphs at different time

f: Graph features such as in-weight, the number of neighbors, triangles

L. Akoglu and C. Faloutsos. "Event detection in time series of mobile communication graphs". In: *Army science conference*. Vol. 1. 2010.

Dynamic graph prediction



One failure case: characterizing anomaly of patient mobility by graph prediction. The given patients' network is so sparse that we do not have sufficient training datasets to build the prediction model. The plot shows that the maximum validation F1 score is less than 0.25. The score only increases when the epoch is below 50 but starts to decrease after it exceeds 100, over-fitting problem.

To solve the issue, we tried a bootstrap method, but it only increased test F1 by 0.025

Scaled density

After making predictions, we will compare the result to the historical graph by a scaled graph density. The formula for the density is as follows:

$$density'(G_t) = \frac{density(G_t) - min(D_T)}{max(D_T) - min(D_T)} \tag{3.9}$$

where $D_T = \{density(G_t) | t = 1, ..., T\}$. In addition, $density(G_t) = \frac{1}{N(N-1)} \sum_{u,v \in G_t} e_t(u,v)$ where G_t refers to a graph at time t whose number of nodes is N, Also, $e_t(u,v)$ becomes 1 if there is a directed edge from u to v within G_t. The reason we scale the measure is to prevent abnormal difference between the predicted and actual values.