Urban Structure and Mobility as Spatio-temporal complex Networks

How do socio-economic attributes of urban locations drive mobility?

Gevorg Yeghikyan

Supervisors: M. Nanni, A. Facchini, A. Passarella, M. Conti, B. Lepri

5 June, 2020

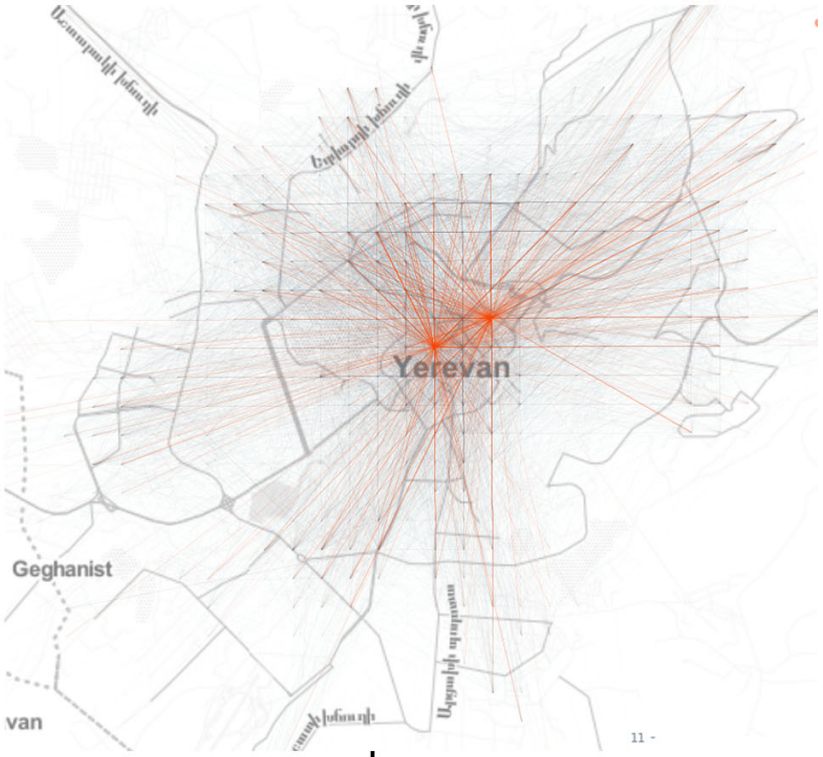
Motivation

How is urban mobility explained?
Maybe networks?

Motivation

a. Urban street network b. Urban mobility Origin-Destination (OD) network

a.



b.

Overarching question: **How is urban mobility related to urban socio-economic attributes?**

Thesis structure

- Data
- Urban attribute-informed network centrality measures
- Centrality-informed urban mobility modelling
- Network-based urban mobility prediction

Motivation

Network centrality measures

Closeness centrality:

$$C_i^C = \frac{N-1}{\sum_{j \in G, j \neq i} d_{ij}}$$

Betweenness centrality:

$$C_i^B = \frac{1}{(N-1)(N-2)} \sum_{s \neq t \in V} \frac{\sigma_{st}(i)}{\sigma_{st}}$$

Eigenvector centrality:

$$C_i^E = \frac{1}{\lambda} \sum_{j=1}^{n} A_{ij} x_j$$

PageRank centrality:

$$C_i^P(n) = c \sum_{j \to i} \frac{\pi_j(n)}{outdeg\ j} + \frac{1-c}{|V_n|}$$

Motivation

Human mobility models[1]: **poor performance for intra-urban flows**

(Constrained) gravity model:

$$T_{ij} = O_i \frac{m_j f\left(r_{ij}\right)}{\sum_k m_k f\left(r_{ik}\right)}$$

Intervening opportunities model:

$$T_{ij} = O_i \frac{e^{-LV_{ij-1}} - e^{-LV_{ij}}}{1 - e^{-LV_{in}}}$$

Radiation model:

$$T_{ij} = O_i \frac{m_i n_j}{\left(m_i + s_{ij}\right)\left(m_i + n_j + s_{ij}\right)}$$

Generalized linear models:

$$T_{ij} = O_i D_j \exp\left(\sum_{k=1}^{m} \beta_k X_{k_{ij}}\right)$$

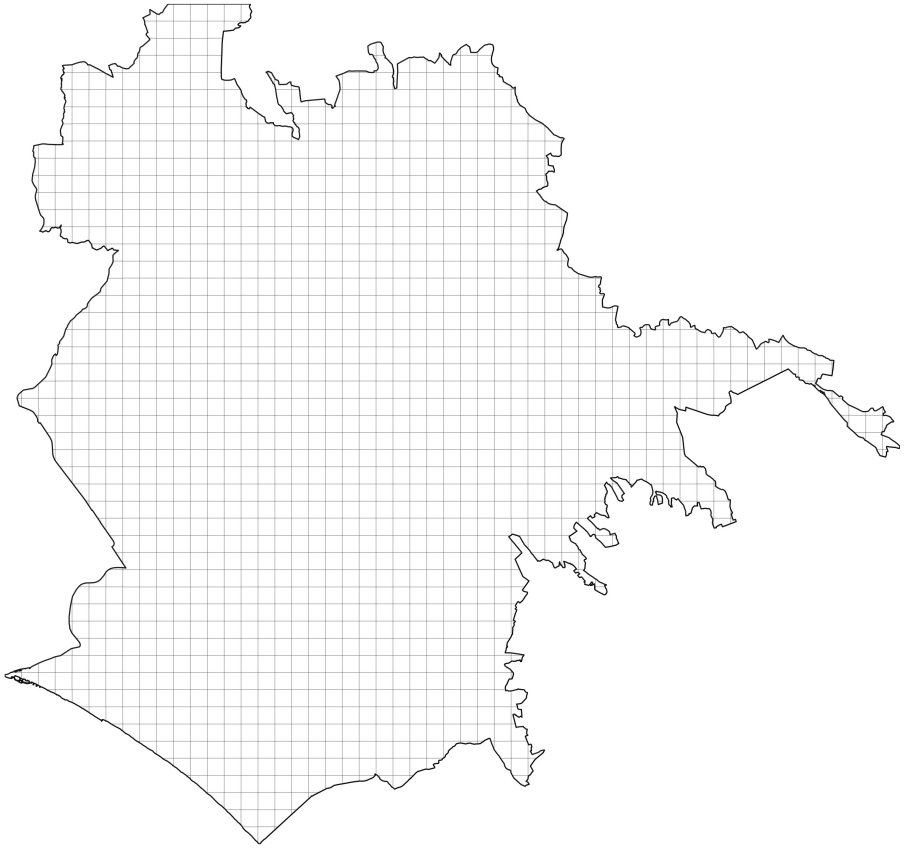
[1] H. Barbosa, et al., 'Human Mobility: Models and Applications', Physics Reports 734 (2018): 1-74

Data

Building the urban mobility network

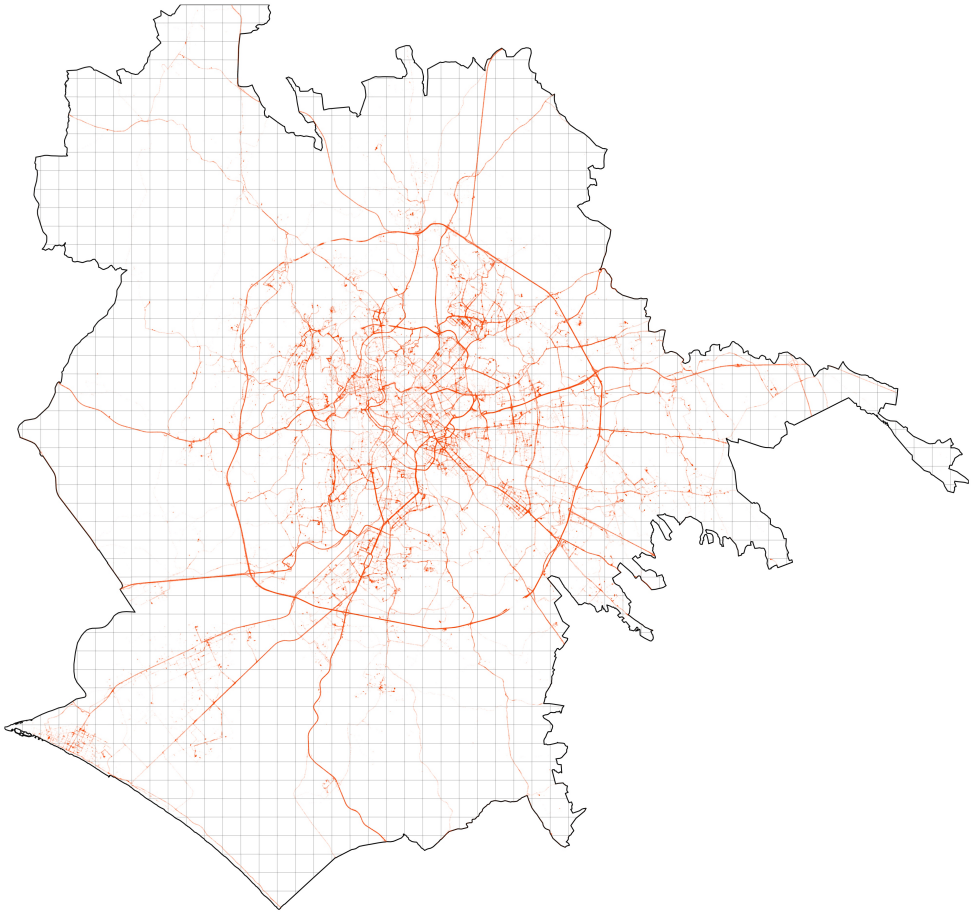
Building the mobility network

Rome: subdividing the city into a homogeneous grid (500×500m, 1000×1000m, etc.):



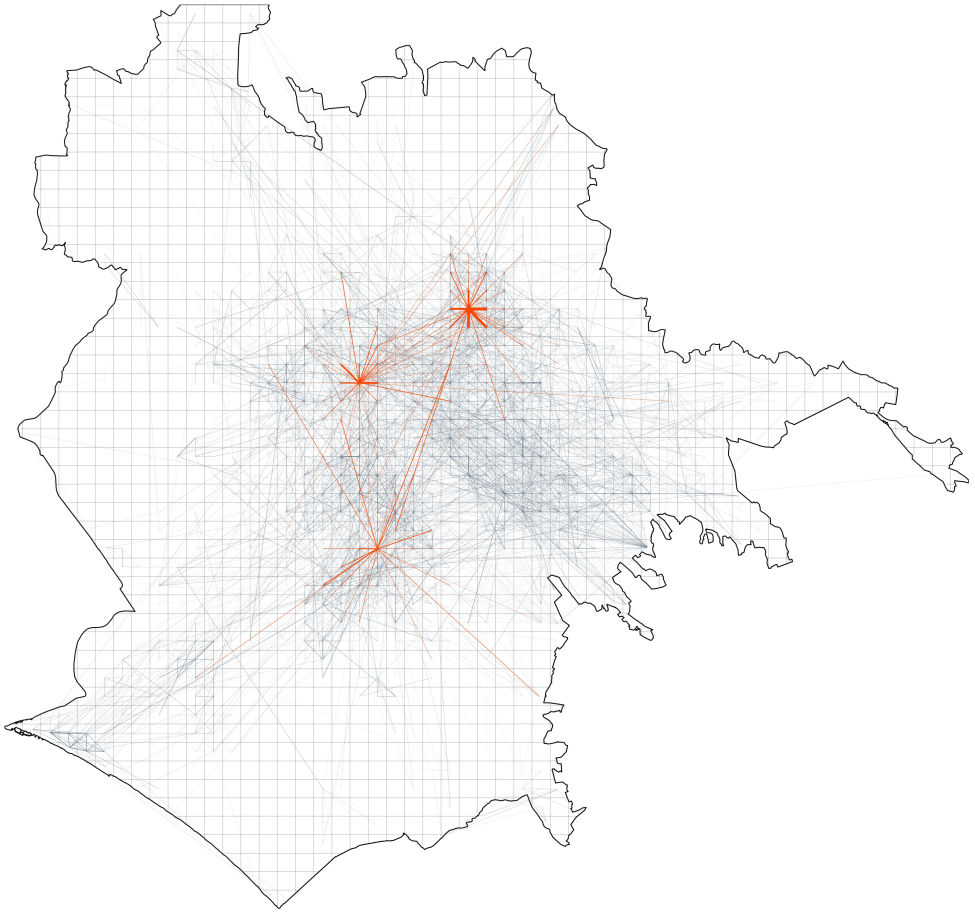
Building the mobility network

Overlaying car GPS trajectories over city grid:



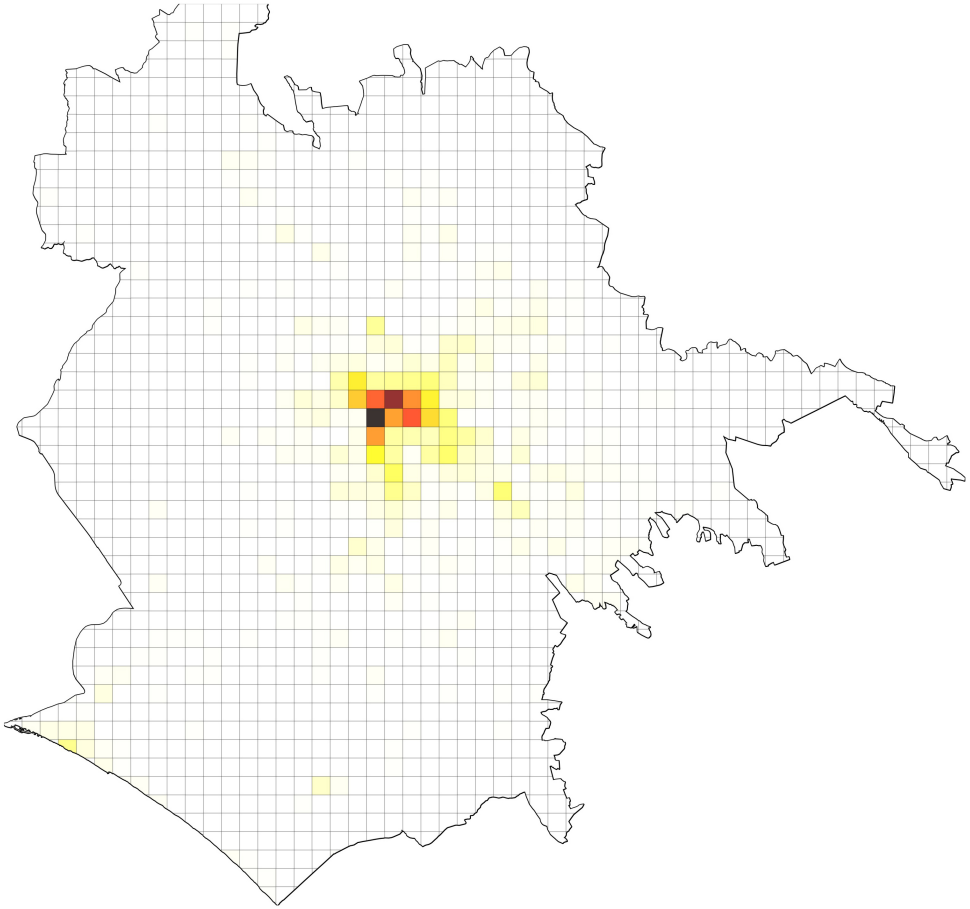
Building the mobility network

Extracting trip origins & destinations, and building the OD network:



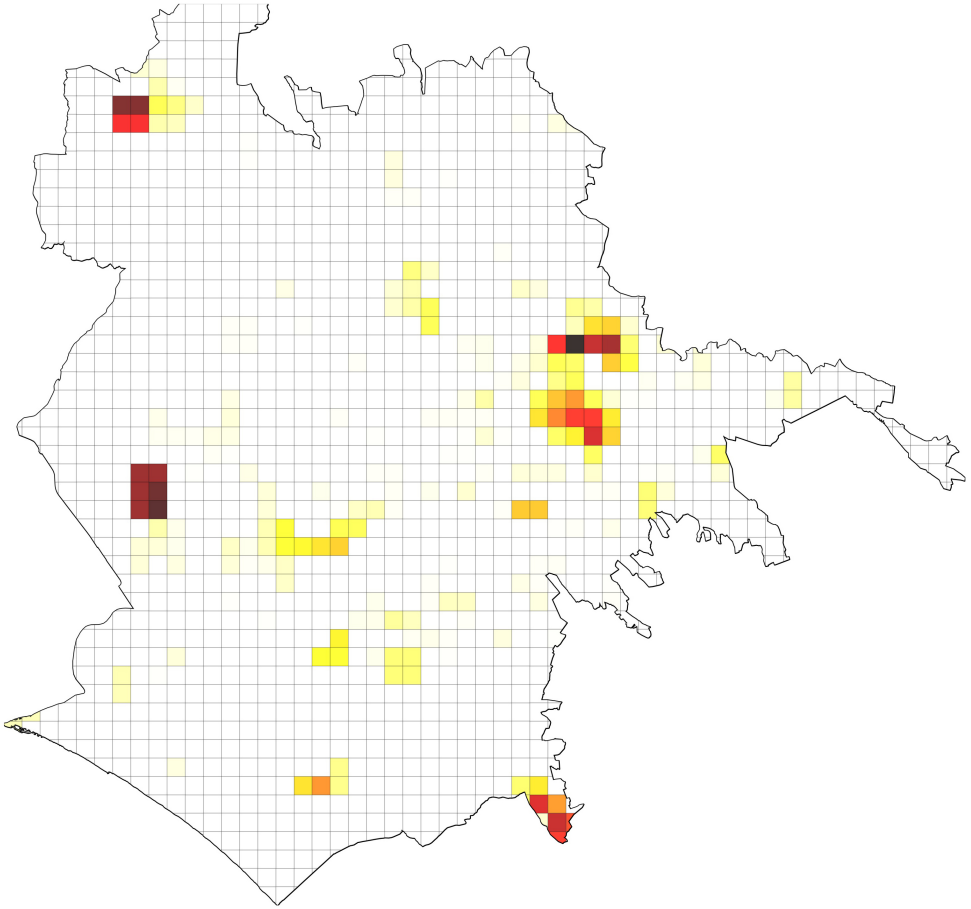
Urban attributes

Rome restaurants per cell:



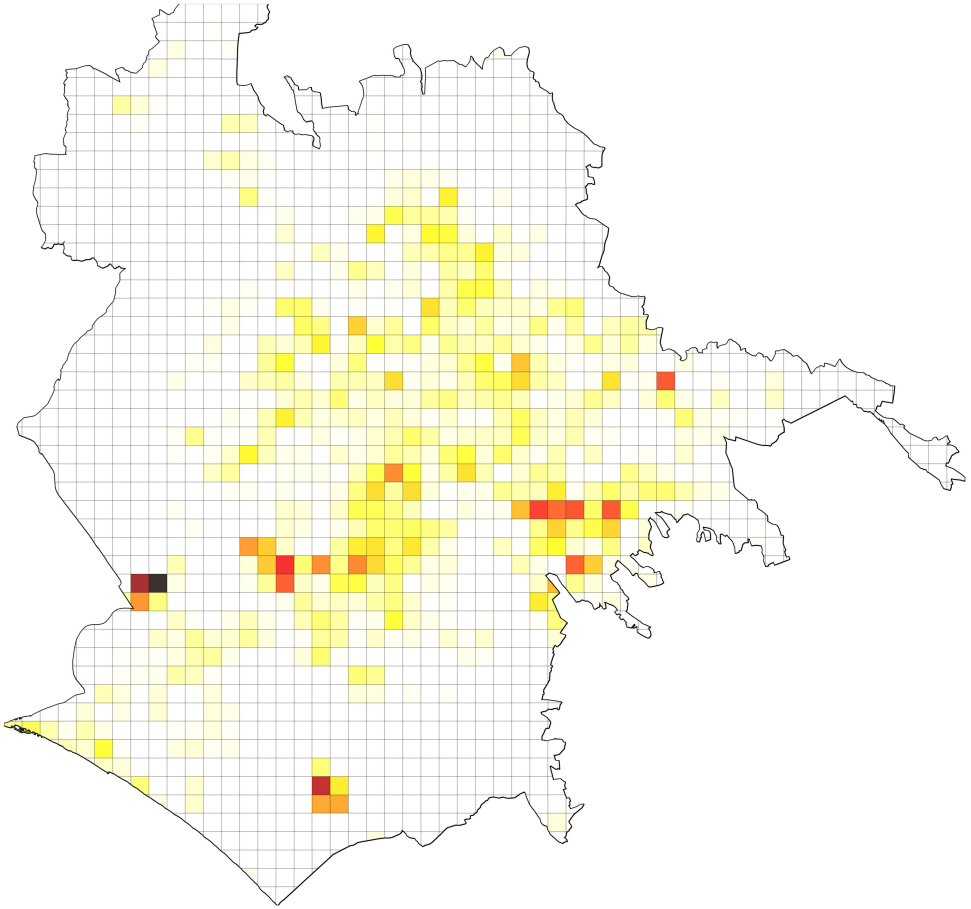
Urban attributes

Rome industrial facility area per cell:



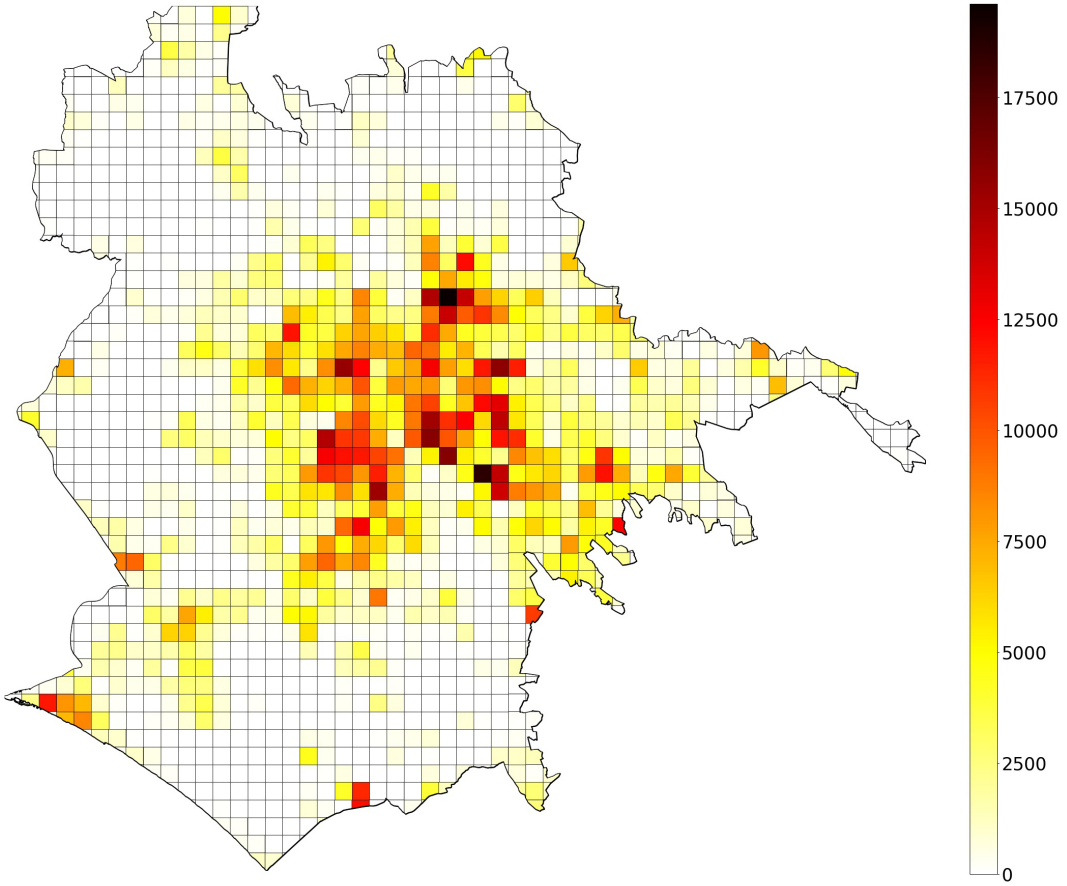
Urban attributes

Rome parking area per cell:



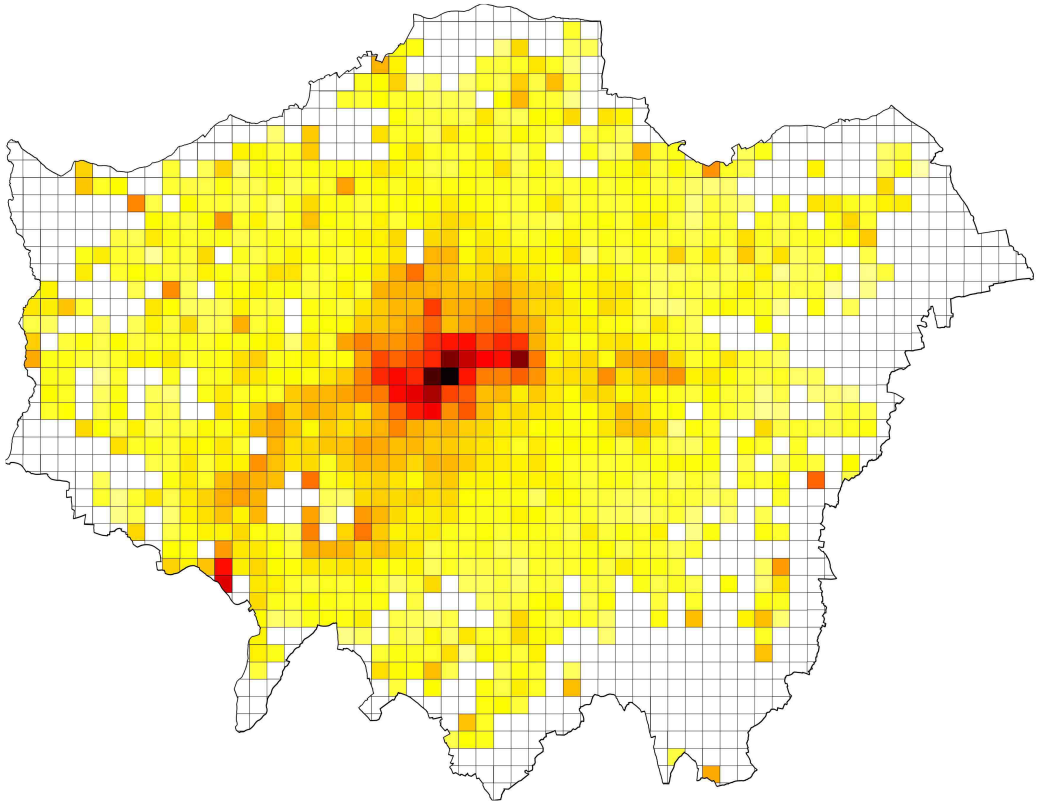
Urban attributes

Rome total car in-flow per cell:



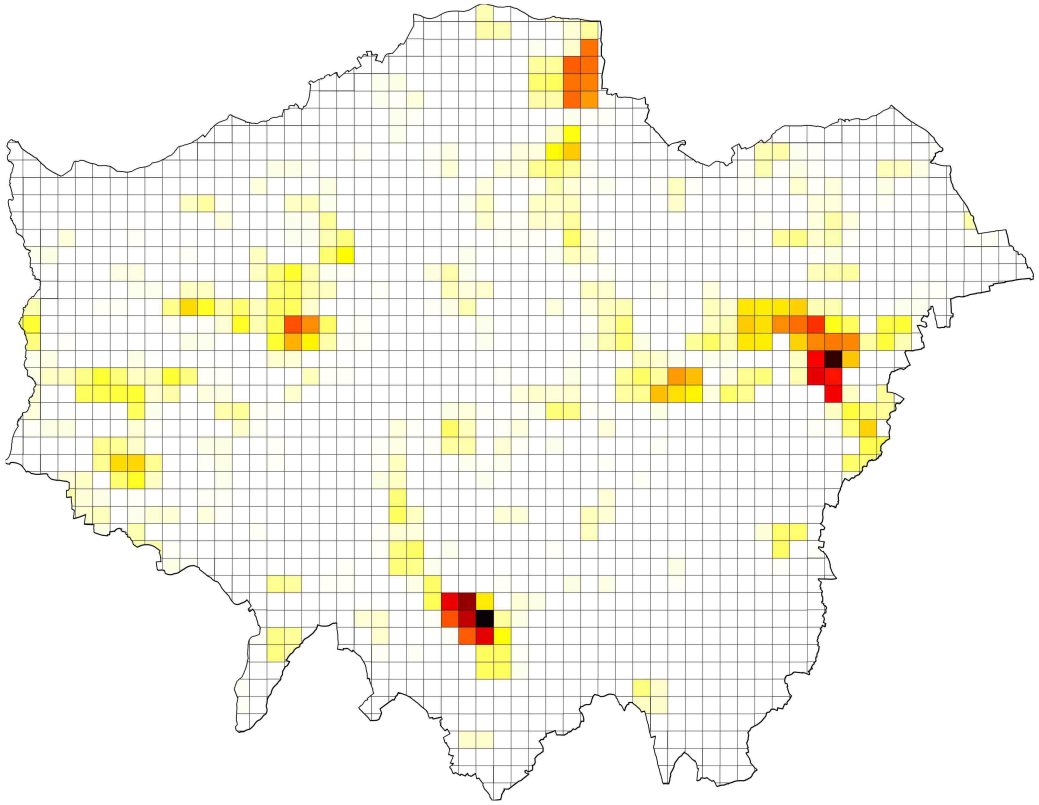
Urban attributes

London average Airbnb prices per cell:



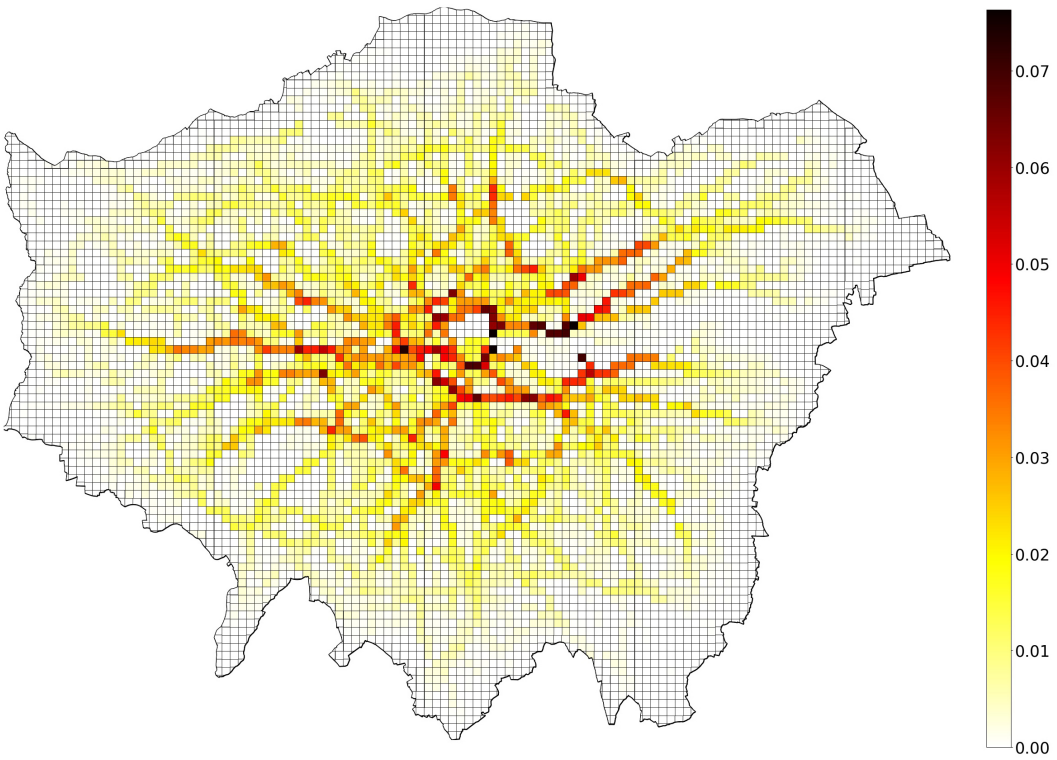
Urban attributes

London industrial facility area per cell:



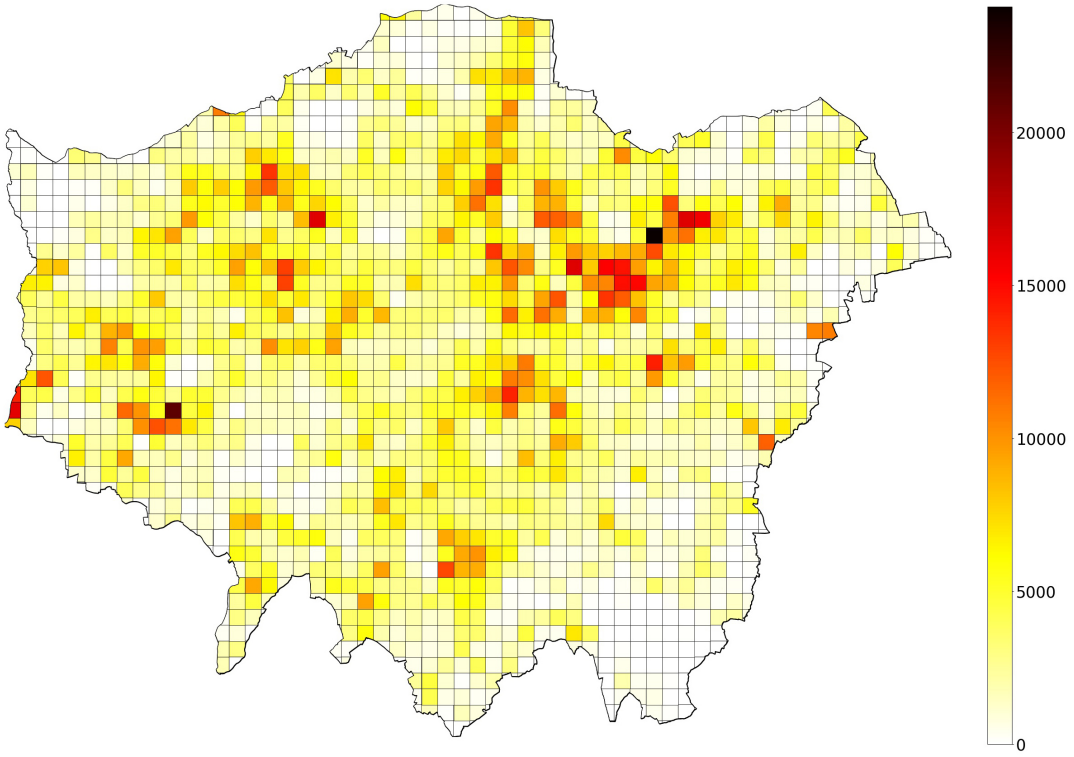
Urban attributes

London average street betweenness centrality per cell:



Urban attributes

London total car in-flow per cell:



Adapted PageRank Algorithm (APA)

Augmenting the Google PageRank algorithm with urban socio-economic attributes

Adapted PageRank Algorithm[1]

The urban attribute data matrix as the jump probability in the random walk.

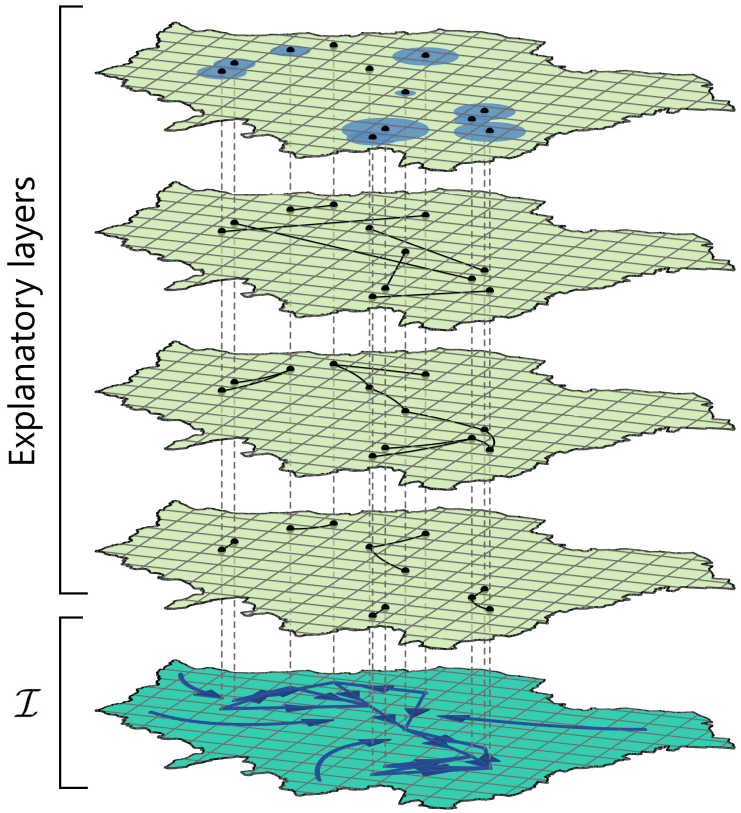
Let $G = (V(G), E(G))$ be a graph representing an urban network. The APA algorithm is as follows.

1. obtain the transition matrix P.
2. obtain the data urban attribute data matrix D associated with the nodes.
3. select the vector $\mathbf{v_0}$, depending on the importance of the attributes studied.
4. obtain the vector $\mathbf{v} = D \cdot \mathbf{v_0}$.
5. normalize $\mathbf{v}$ vector, $\mathbf{v} \longrightarrow \mathbf{v}^*$.
6. construct V matrix as $V = \mathbf{v}^* e^T$.
7. obtain M_{APA} matrix as $M_{APA} = (1-\alpha)P + \alpha V$.
8. The eigenvector $\mathbf{x}$ associated with the dominant eigenvalue $\lambda_1 = 1$ of M_{APA} is the ranking.

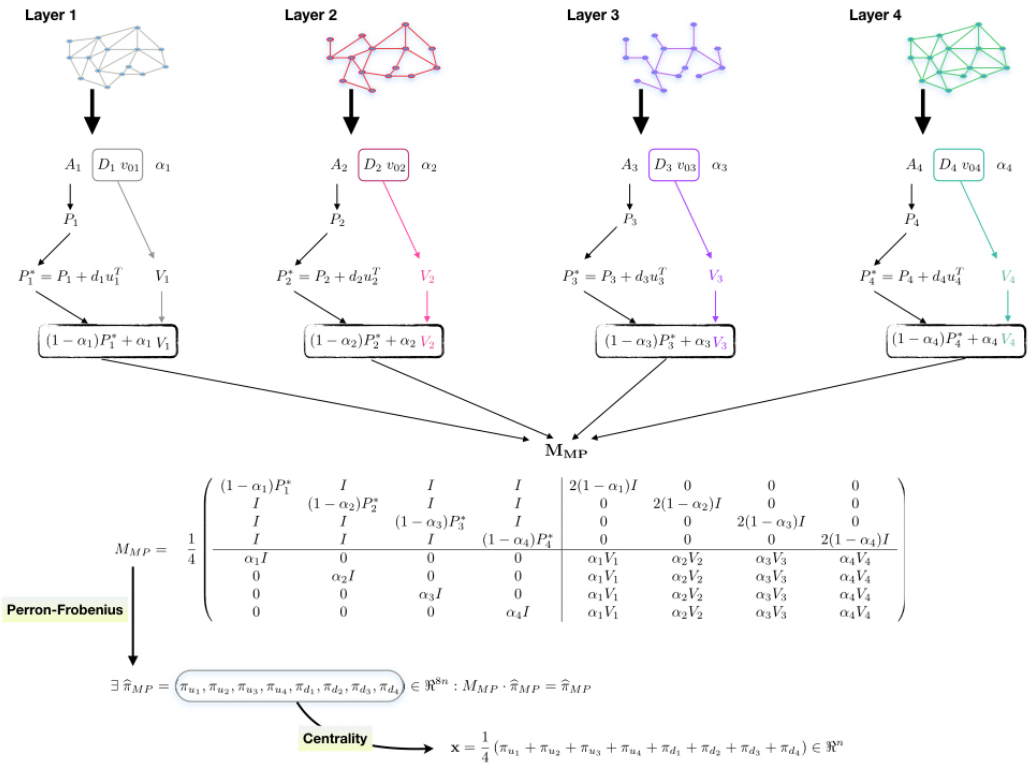
[1] G. Yeghikyan, et al., Analysis and comparison of centrality measures applied to urban networks with data. *Journal of Computational Science*, page 101127, 2020

Multilayer attributed urban mobility network

Adding attribute layers describing **relations** between city locations:



Multilayer APA centrality[1,2]

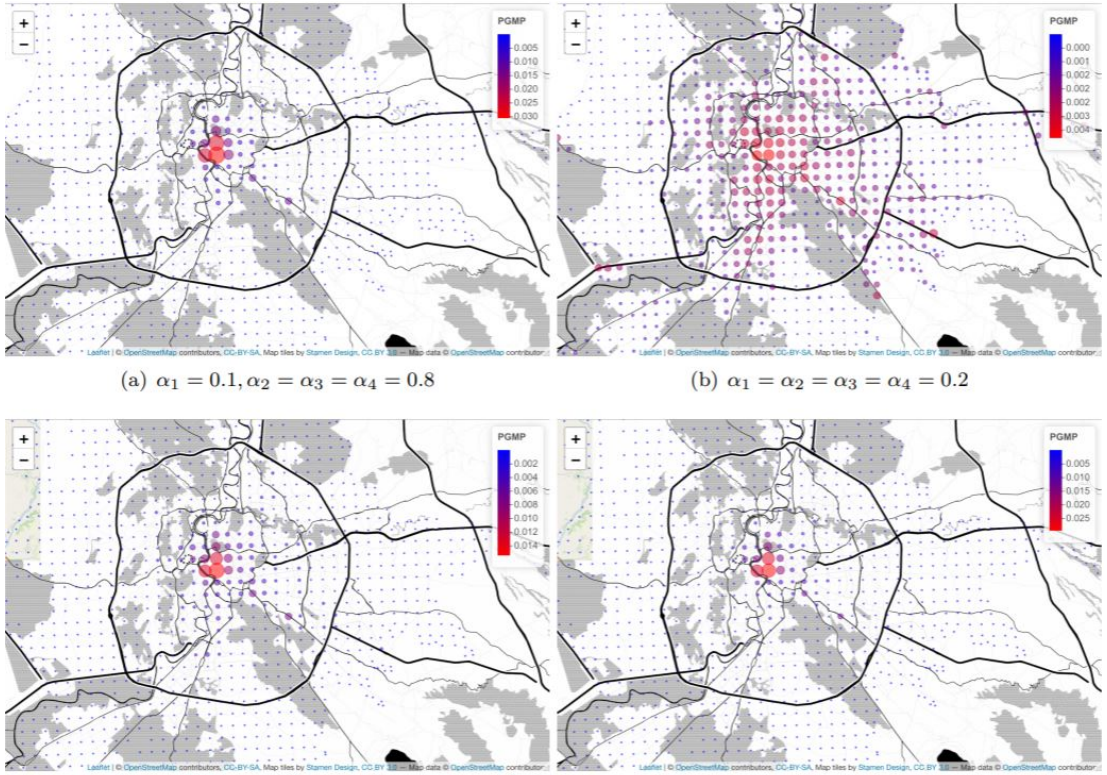


[1] G. Yeghikyan, et al., Understanding mobility in Rome by means of a multiplex network with data. *Applied Mathematics and Computation*, *under review*

[2] G. Yeghikyan, et al., Analysis and comparison of centrality measures applied to urban networks with data. *Journal of Computational Science*, *under review*

Multilayer APA centrality

Rome Multilayer APA centrality with different importance given to node attribute data in different layers:



(a) $\alpha_1 = 0.1, \alpha_2 = \alpha_3 = \alpha_4 = 0.8$ (b) $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0.2$

(c) $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0.5$ (d) $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0.7$

Spatio-temporal network centrality

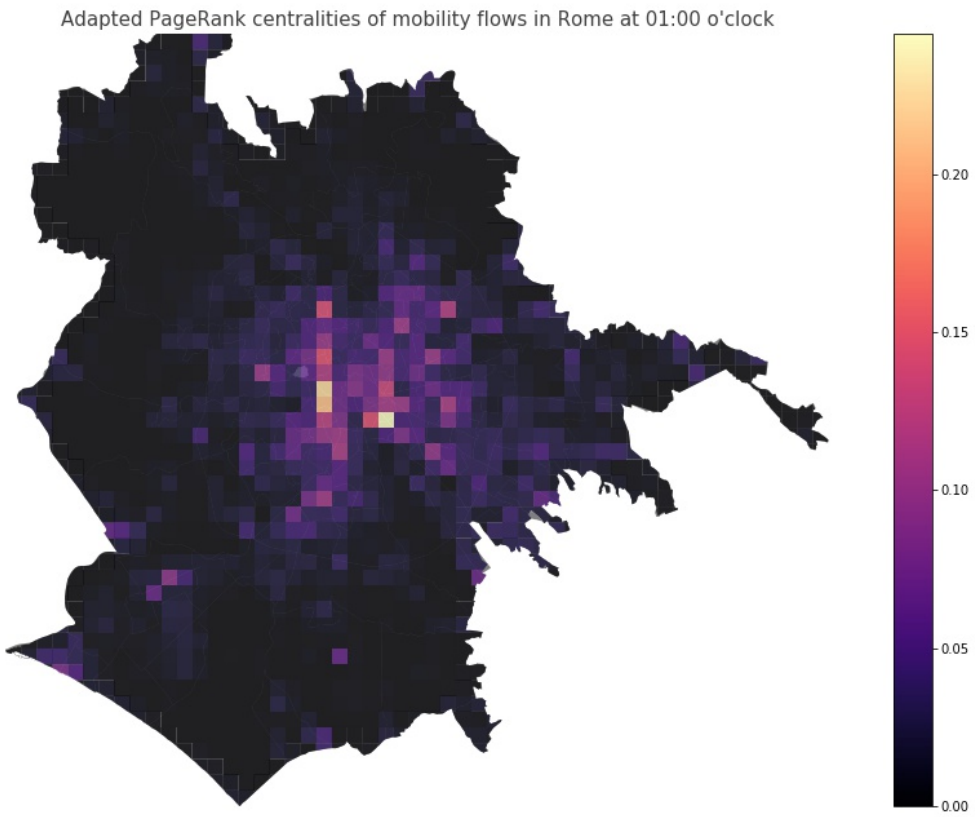
Ranking places in attributed temporal urban mobility networks[1]

[1] G. Yeghikyan, et al., Ranking places in attributed temporal urban mobility networks. PloS one, 2020, *under review*

Rome mobility network over time

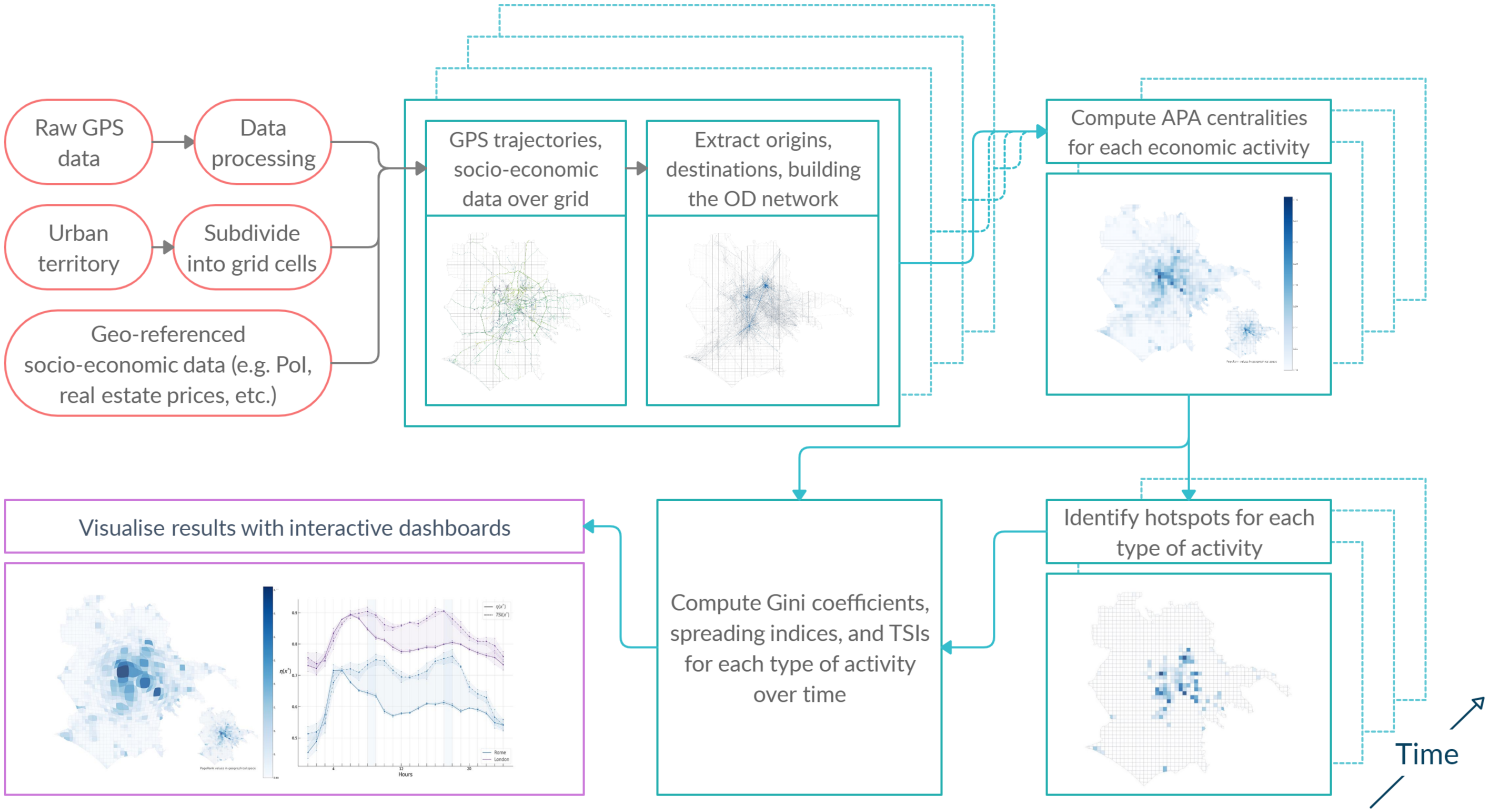
Visualization of Rome temporal mobility network

Rome APA centralities over time



APA centralities over time[1]

Temporal APA centrality workflow



APA centralities over time

Spreading index:

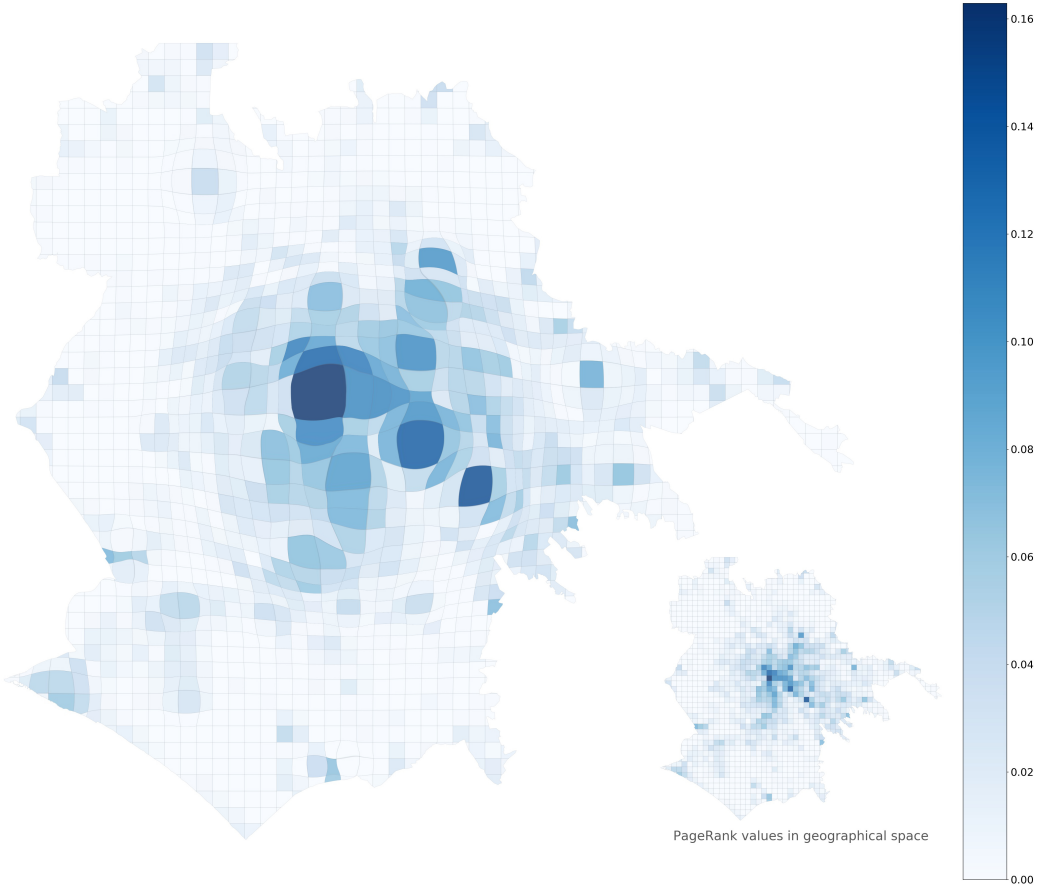
$$\eta\left(\mathbf{x}^*\right) = \frac{\frac{1}{N(\mathbf{x}^*)}\sum_{i,j} d(i,j)\mathbf{1}_{(\mathbf{x}_i>\mathbf{x}^*)}\mathbf{1}_{(\mathbf{x}_j>\mathbf{x}^*)}}{\frac{1}{N}\sum_{i,j} d(i,j)}$$

Time-space spreading index:

$$TSI\left(\mathbf{x}^*\right) = \frac{\frac{1}{N(\mathbf{x}^*)}\sum_{i,j} t(i,j)\mathbf{1}_{(\mathbf{x}_i>\mathbf{x}^*)}\mathbf{1}_{(\mathbf{x}_j>\mathbf{x}^*)}}{\frac{1}{N}\sum_{i,j} t(i,j)}$$

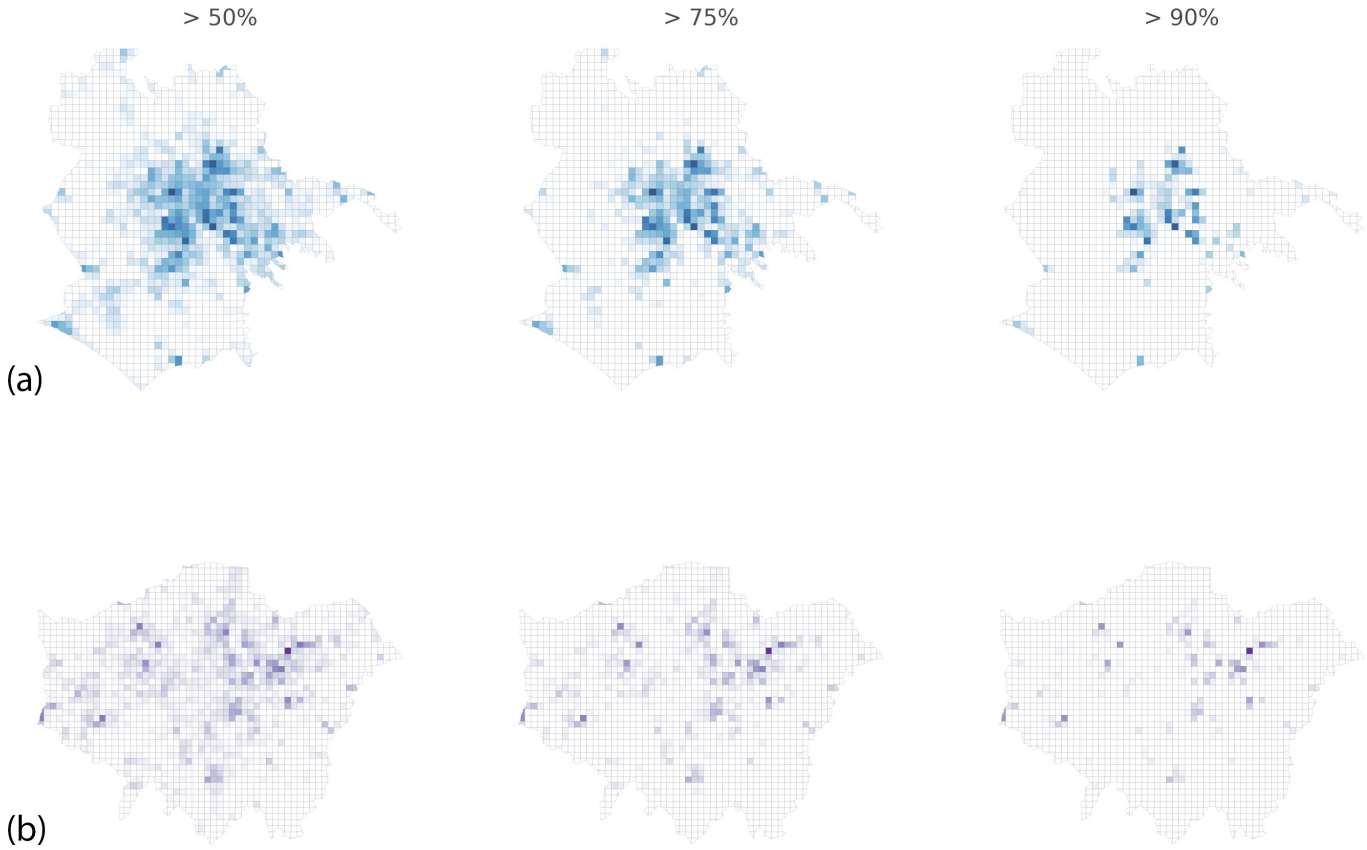
APA centralities over time

Multidimensional Scaling representation with travel time as the distance metric:



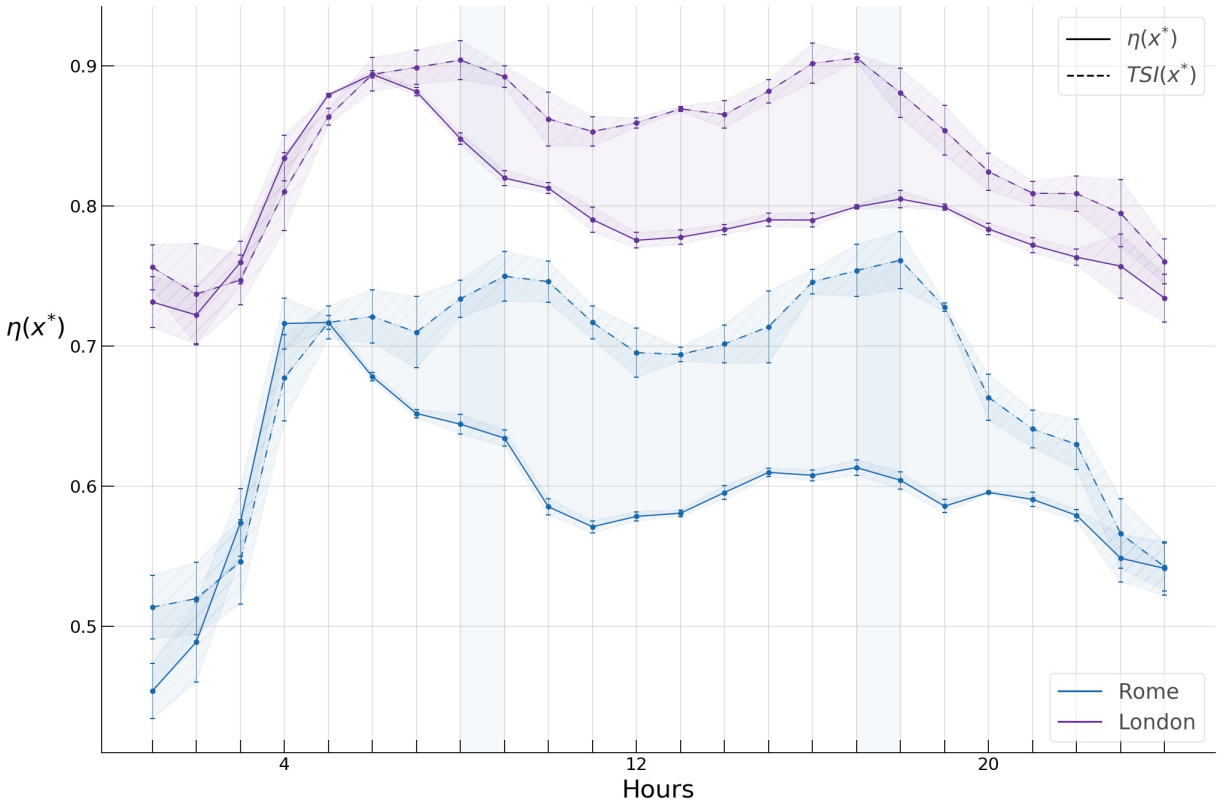
APA centralities over time

APA centralities above specified percentile in Rome and London:



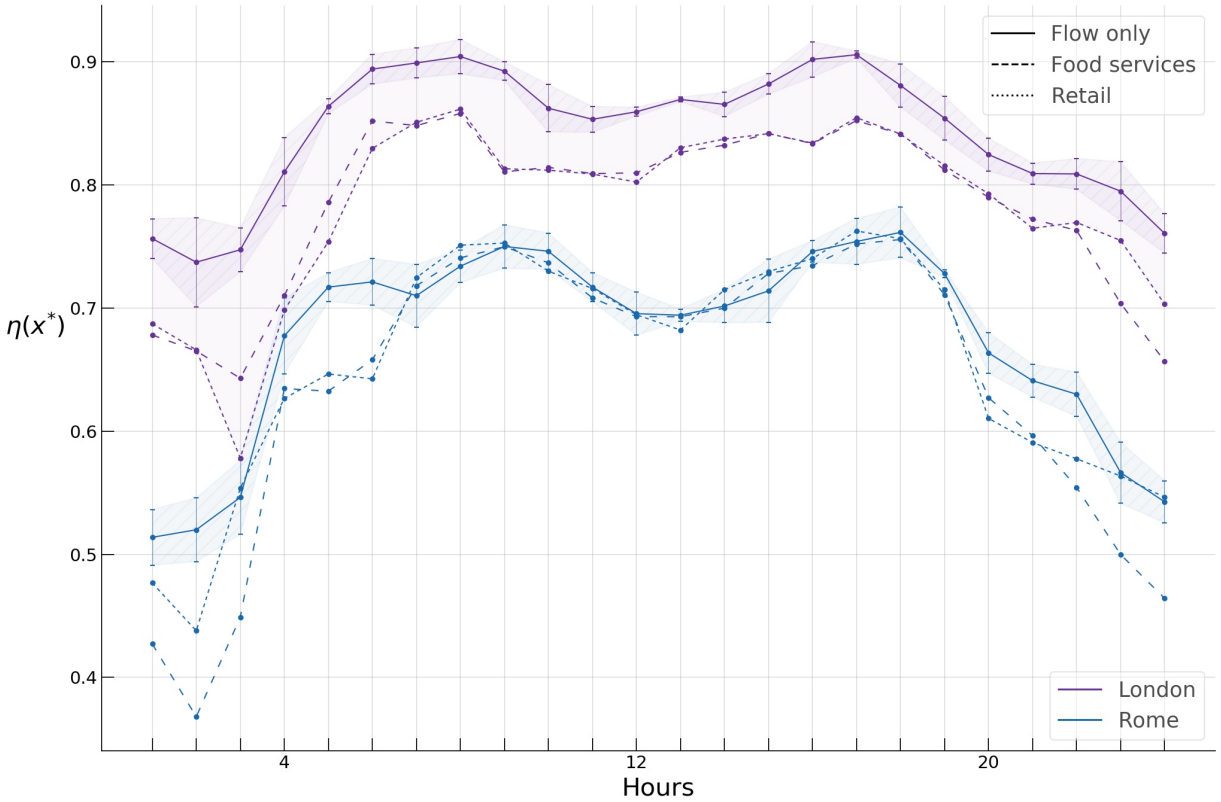
APA centralities over time

Spreading index vs. TSI in Rome and London:



APA centralities over time

Temporal behaviour of spreading indices of flow only, retail and food APA in Rome and London:

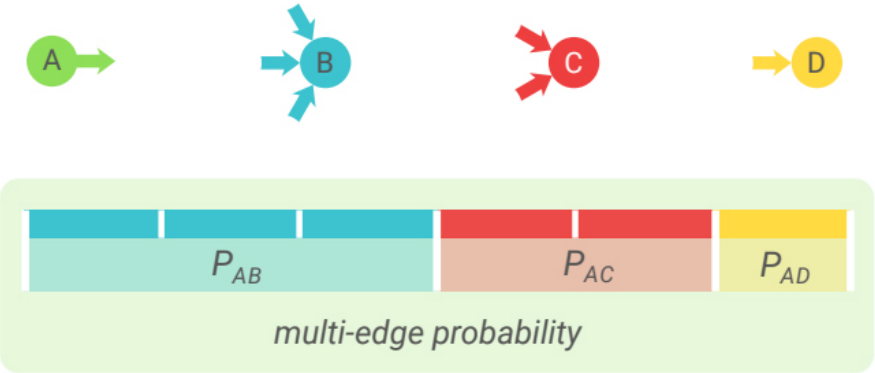


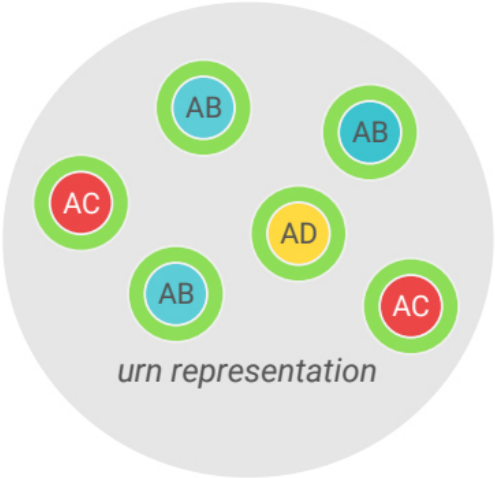
Multilayer Network Regression

Explaining urban mobility from urban socio-economic attributes

Hypergeometric graph ensembles

The configuration model represented (left) as a conventional edge rewiring process and (right) as an urn problem. In the former case, once the out-stub (A,-) has been sampled for rewiring, then one in-stub is sampled uniformly at random from those available, to draw a new multi-edge[2]:

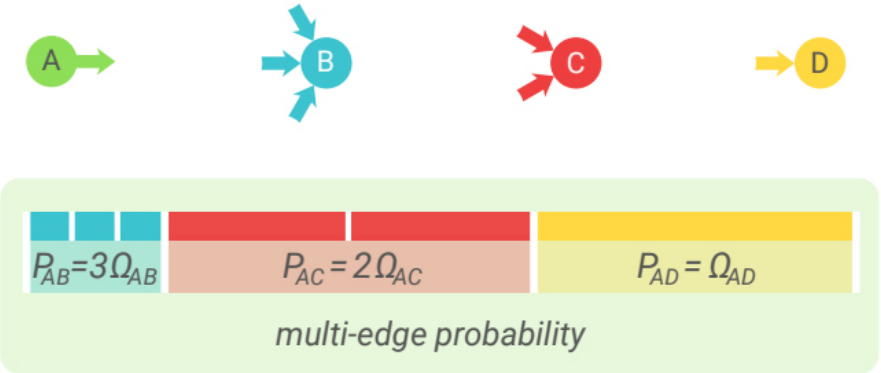


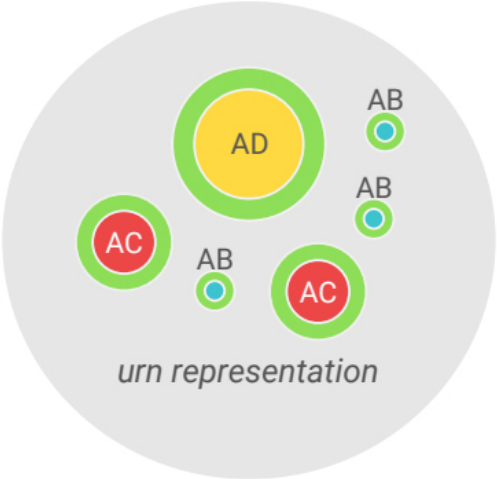


[2] Casiraghi, G.; Nanumyan, V.; Scholtes, I.; Schweitzer, F. (2016). Generalized Hypergeometric Ensembles: Statistical Hypothesis Testing in Complex Networks. arXiv preprint: 1607.02441

Hypergeometric graph ensembles

The effect of edge propensities on the configuration model. Differently from the standard configuration model, here the stubs are not sampled uniformly at random[3]:

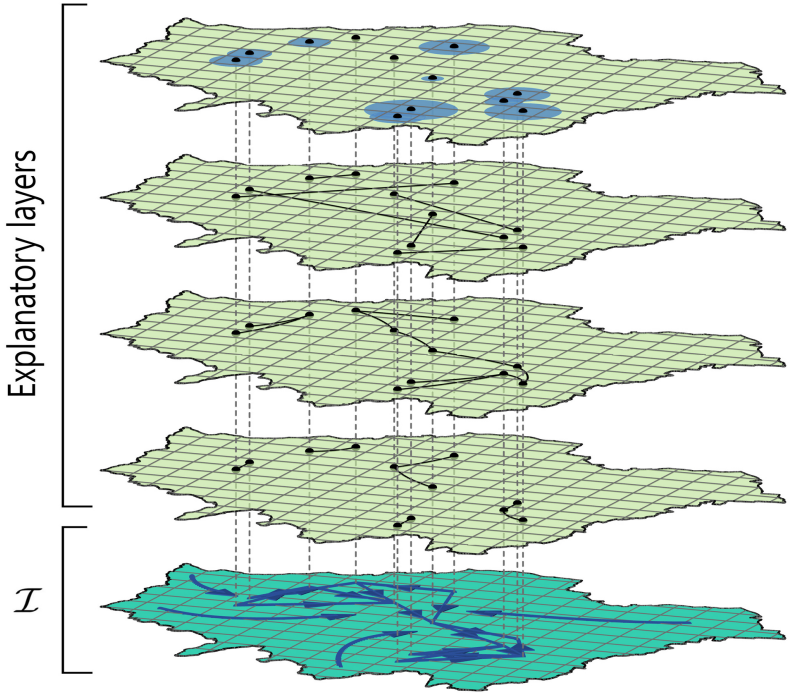




[3] Casiraghi, G.; Nanumyan, V.; Scholtes, I.; Schweitzer, F. (2016). Generalized Hypergeometric Ensembles: Statistical Hypothesis Testing in Complex Networks. arXiv preprint: 1607.02441

Hypergeometric graph ensembles

Multilayer network representation of urban relations and MLE via Wallenius non-central hypergeometric distribution



$$L(\beta|\mathcal{I}) = \left[\prod_{i,j}\binom{\Xi_{ij}}{A_{ij}}\right]\int_0^1 \prod_{i,j}\left(1 - z^{\frac{\prod_{l=1}^{p} R_{l,ij}^{\beta_l}}{S_\beta}}\right)^{A_{ij}} dz,$$

with $S_\beta = \sum_{i,j}\prod_{l=1}^{r} R_{l,ij}^{\beta_l}(\Xi_{ij} - A_{ij})$.

Mobility models' performance on London data

Comparison of gHypE multilayer regression performance against baseline methods

Model	**Adjusted R^2**	**AIC**	**SRMSE**	**SSI**
gravity	0.4698	1.2124e+07	16.4691	0.3789
origin-constrained gravity	0.4881	1.1704e+07	16.3256	0.3841
destination-constrained gravity	0.4879	1.1709e+07	16.3283	0.3884
doubly-constrained gravity	0.4924	1.1606e+07	15.8827	0.3937
Poisson log-linear regression	0.6291	8.2458e+05	13.2451	0.4615
Negative Binomial regression	0.5258	2.1854e+06	13.5257	0.5382
Spatially adjusted Poisson	0.6571	3.1108e+05	12.9981	0.5321
Spatially adjusted NB	0.5869	1.0201e+06	12.3859	0.5819
gHypE multilayer regression	**0.7228**	**8.2209e+03**	**7.4491**	**0.6194**

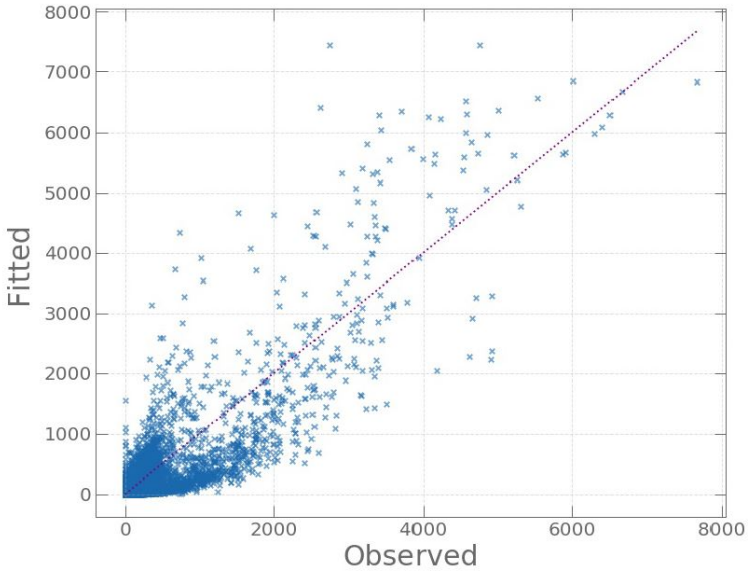
gHypE Multilayer network regression on London data

MLE coefficients of the gHypE multilayer regression model in London

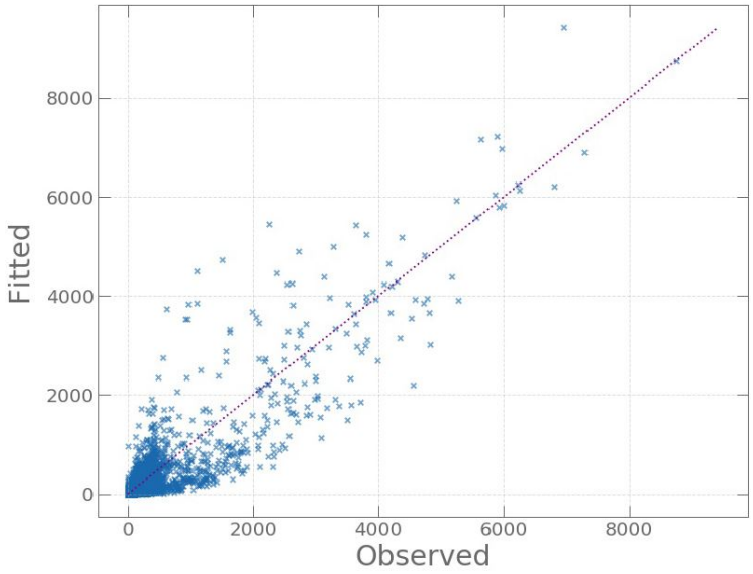
Coefficients	Estimate	Std. Err	p-value
betweenness	0.132	5.286e-02	1.995e-06
network distance	-1.256	1.149e-01	<1e-16
route factor	0.312	1.769e-01	2.124e-02
airbnb	-0.82	2.488e-02	4.321e-02
time	-0.631	2.717e-02	<1e-16
speed	-0.152	5.494e-02	1.112e-03
bus	0.209	3.361e-02	3.921e-02
subway	-0.121	3.481e-02	3.199e-02
densities	0.228	8.396e-03	4.289e-03
residential to rest	0.208	5.824e-02	1.119e-02
highway	0.062	1.098e-01	2.819e-03
correlation	-0.479	1.692e-02	<1e-16
APA flow only	0.670	3.611e-02	<1e-16
APA food	0.607	2.001e-02	<1e-16
APA retail	0.634	1.452e-02	<1e-16

gHypE Multilayer network regression

gHypE network regression fitted prediction values for (a) London (b) Rome



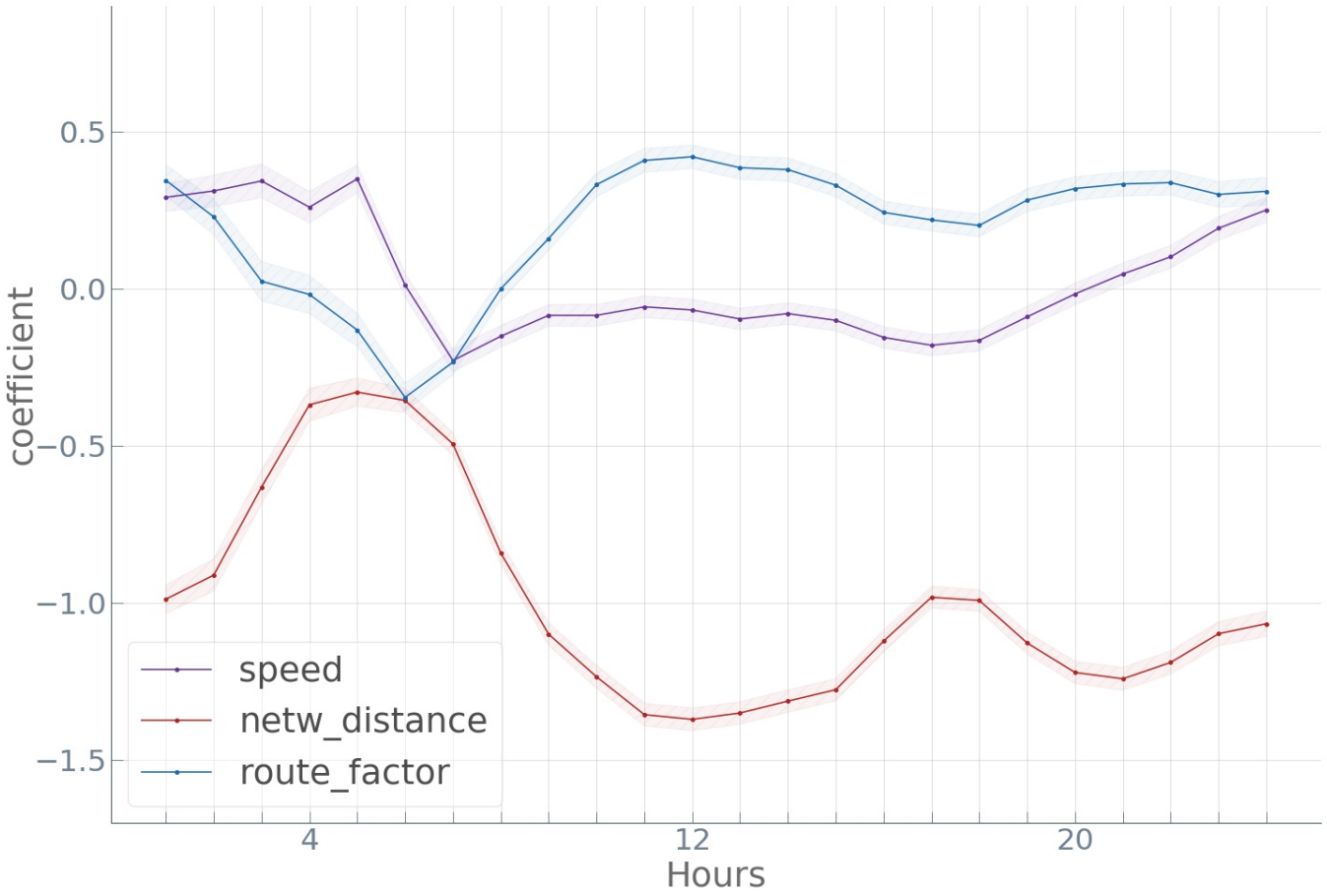
(a)



(b)

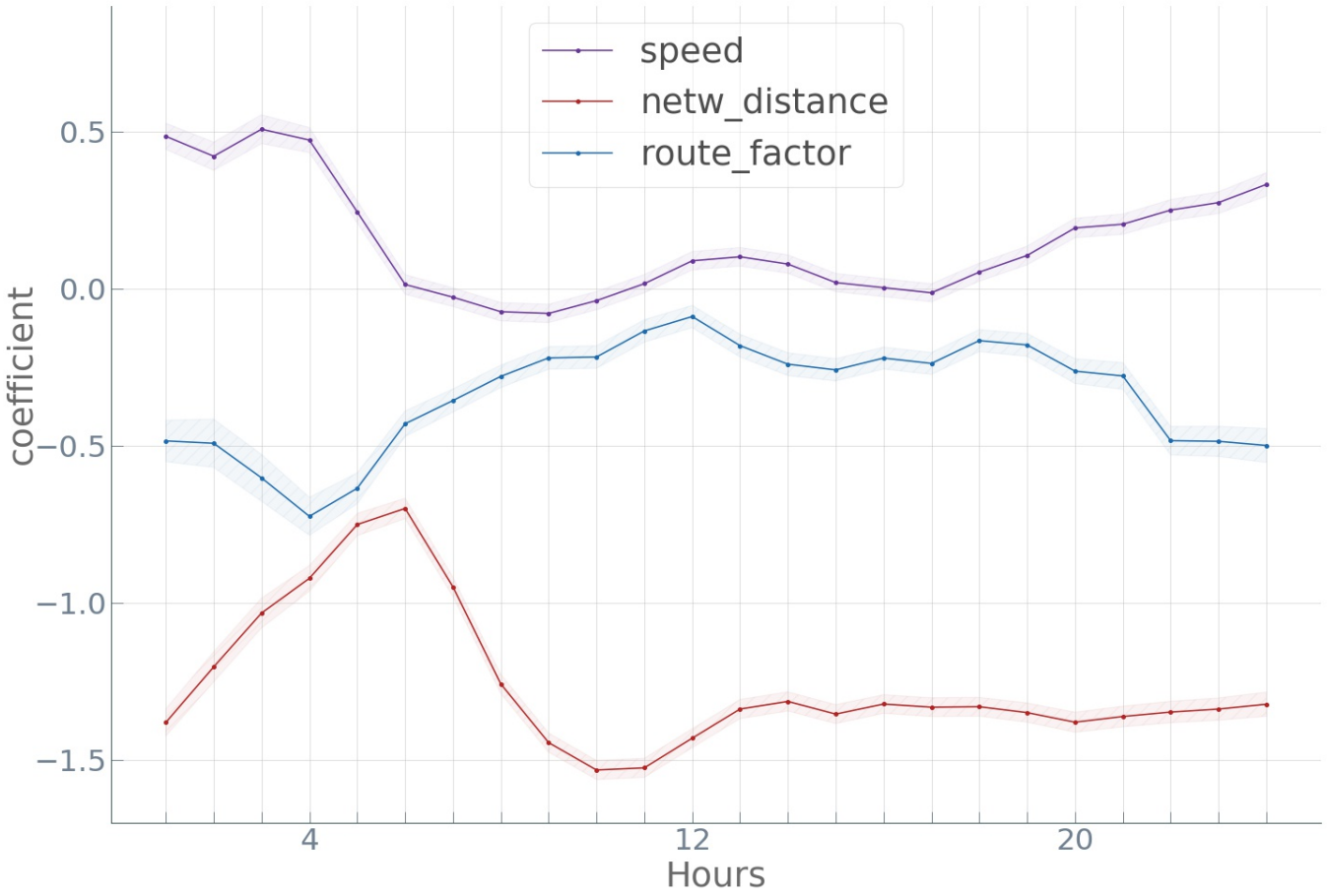
Hourly temporal regression results

London network distance, route factor, and speed coefficients over time



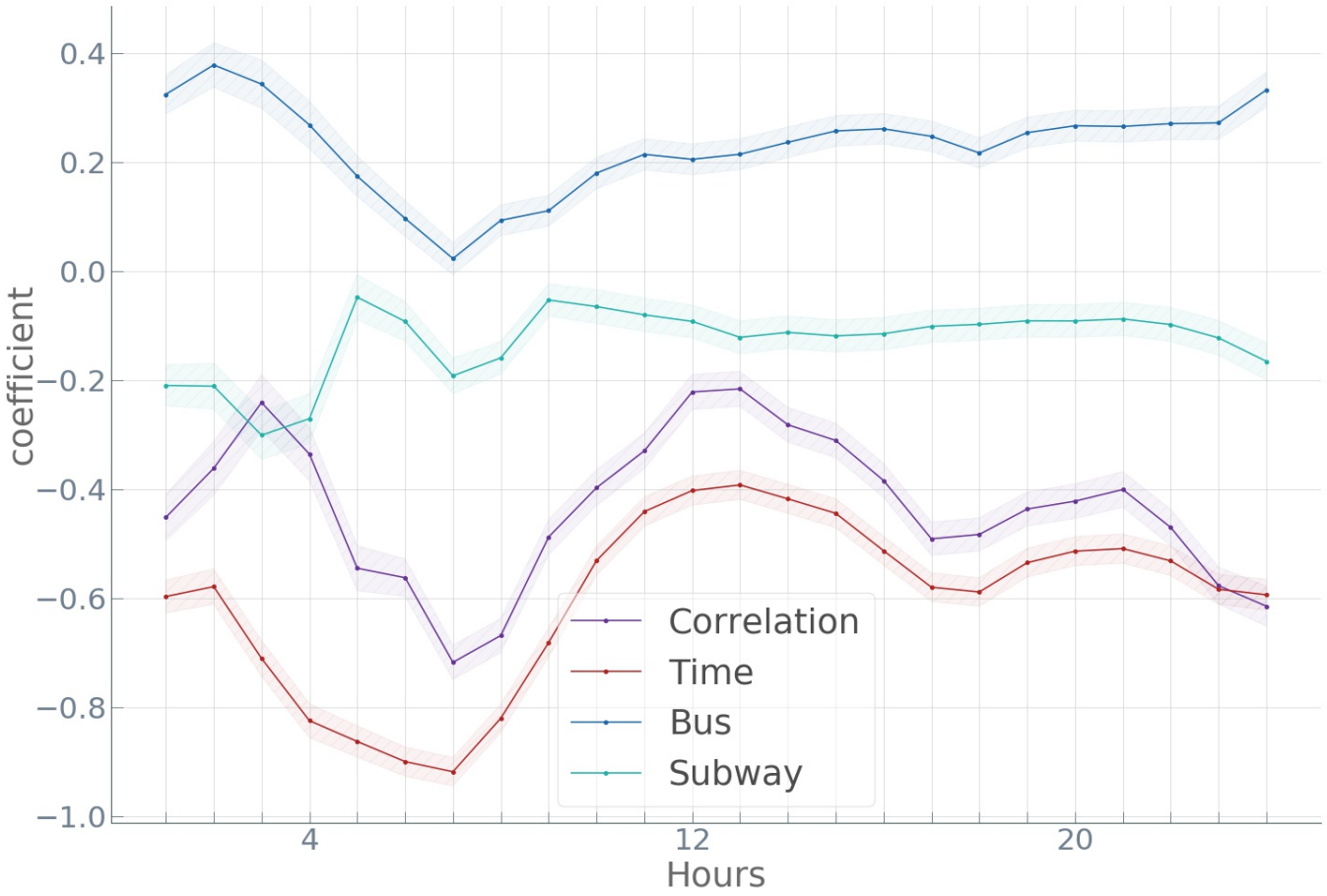
Hourly temporal regression results

Rome network distance, route factor, and speed coefficients over time



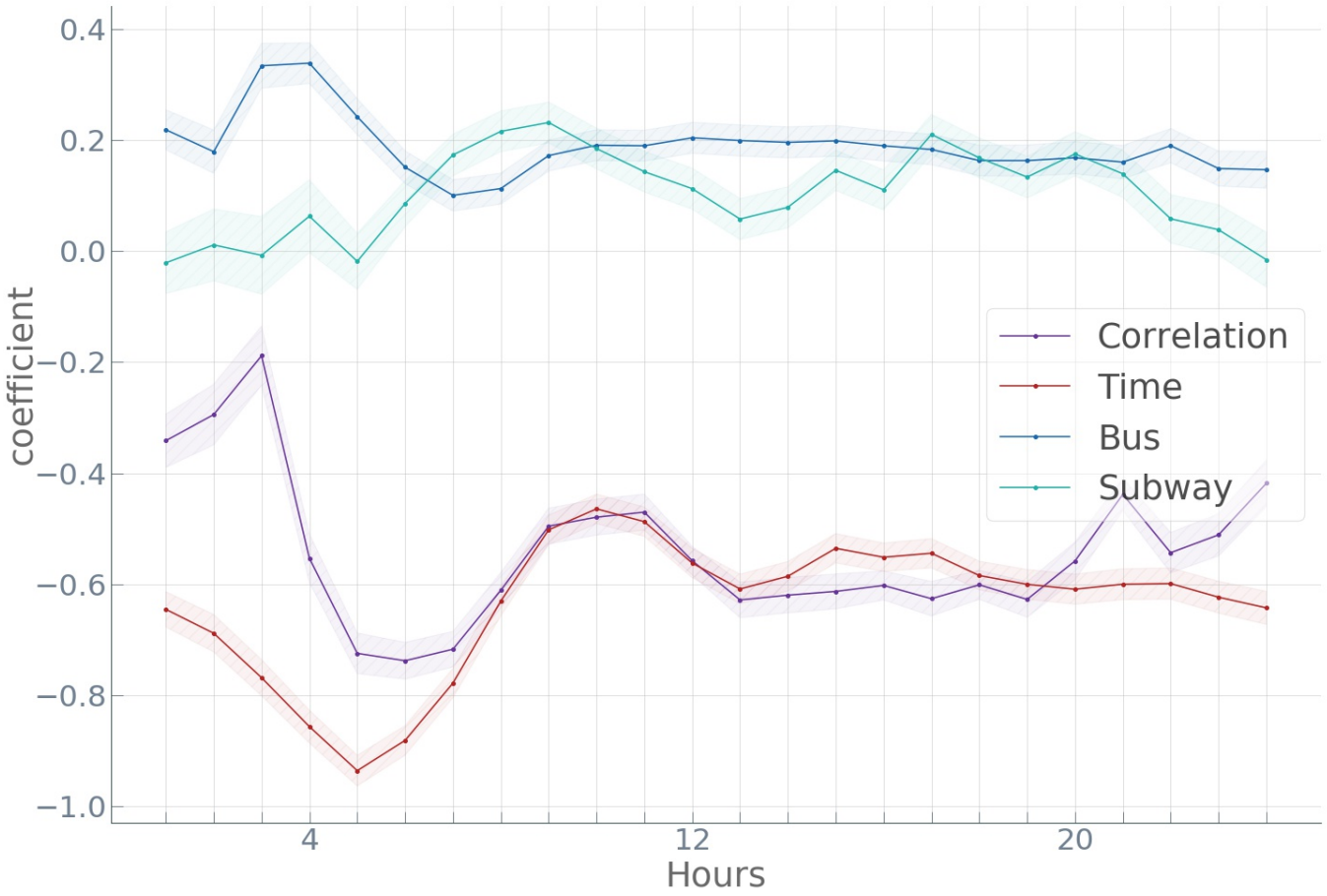
Hourly temporal regression results

London travel time, correlation, subway, and bus coefficients over time



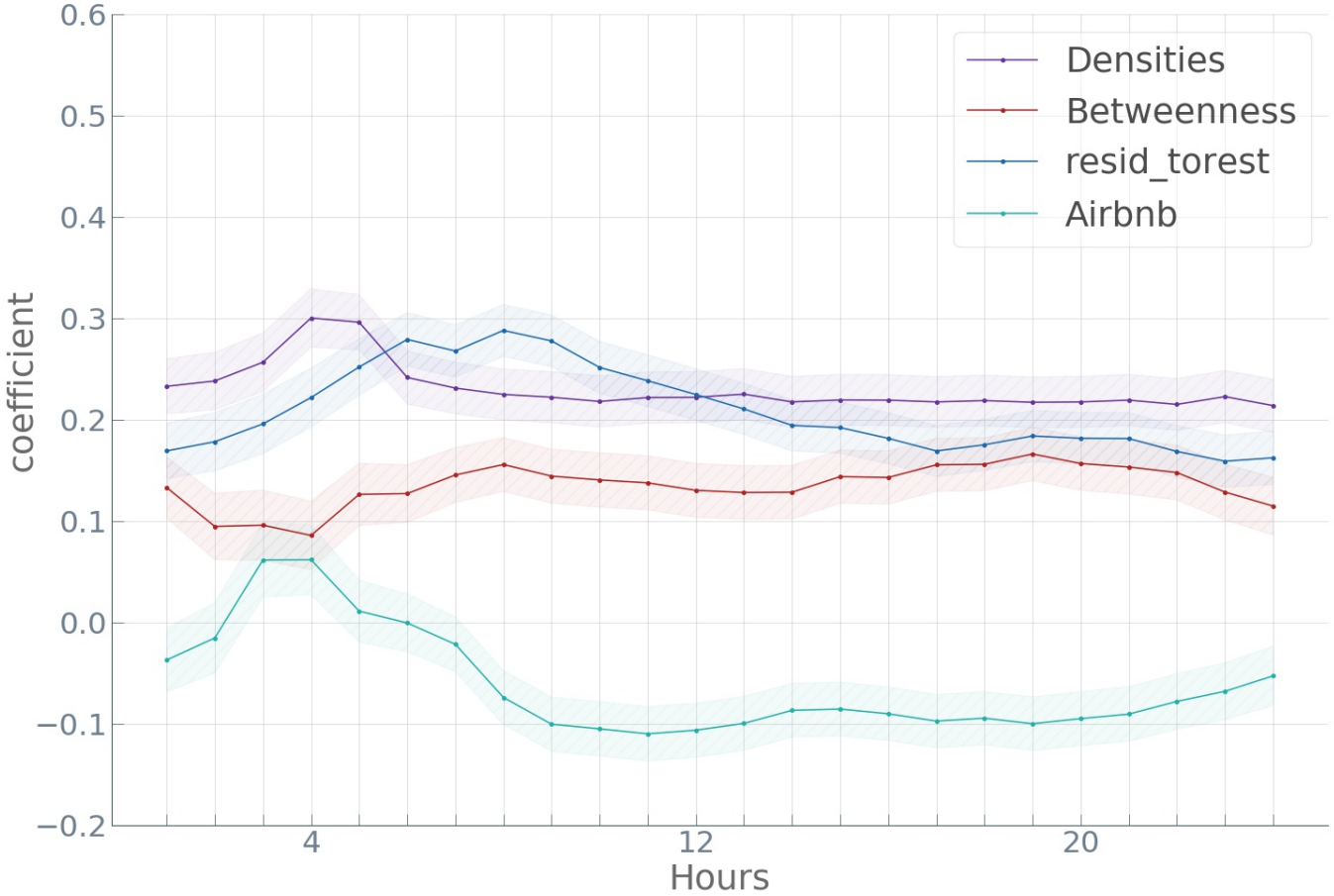
Hourly temporal regression results

Rome travel time, correlation, subway, and bus coefficients over time



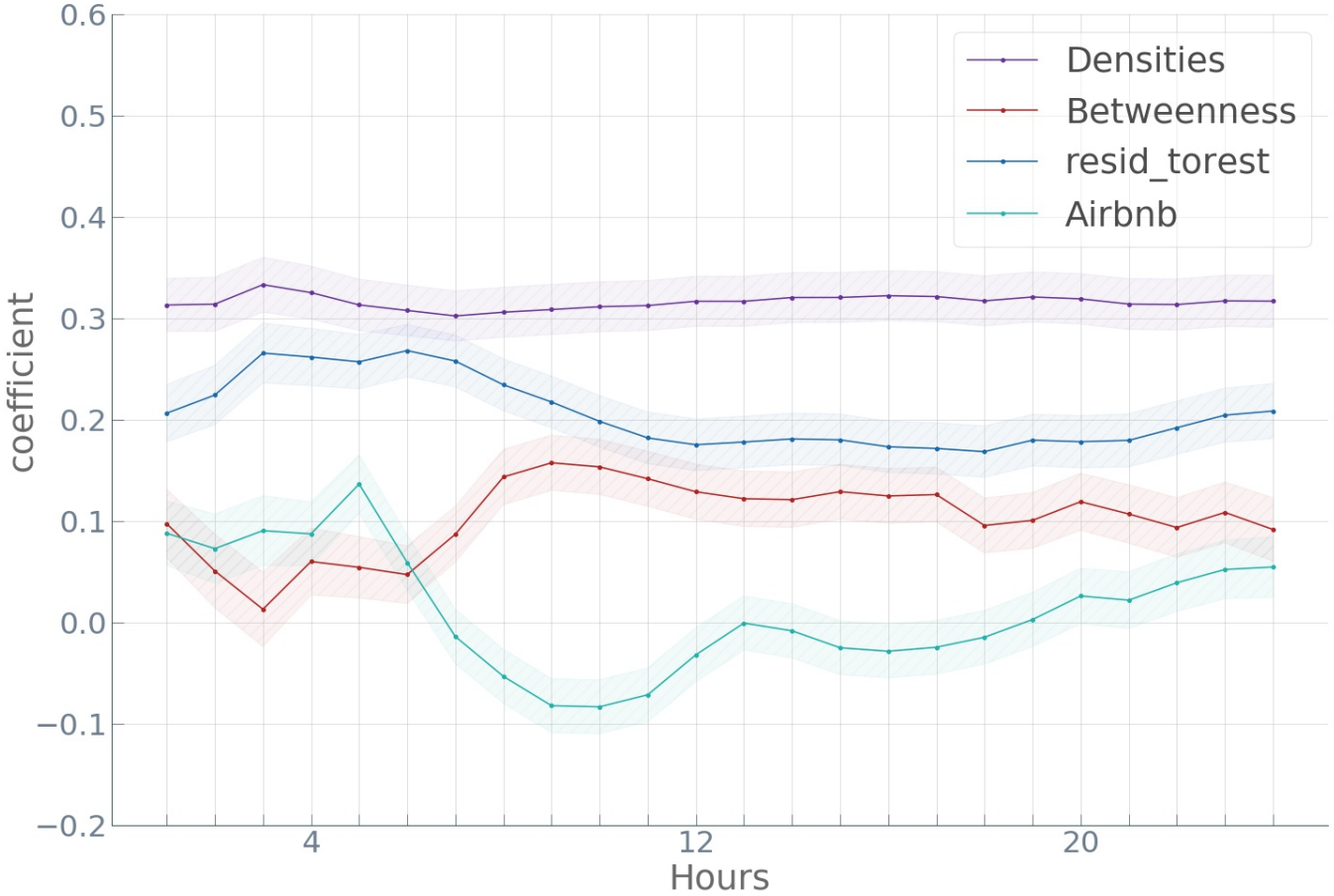
Hourly temporal regression results

London population densities, betweenness centrality, residential-to-other, and Airbnb price coefficients over time



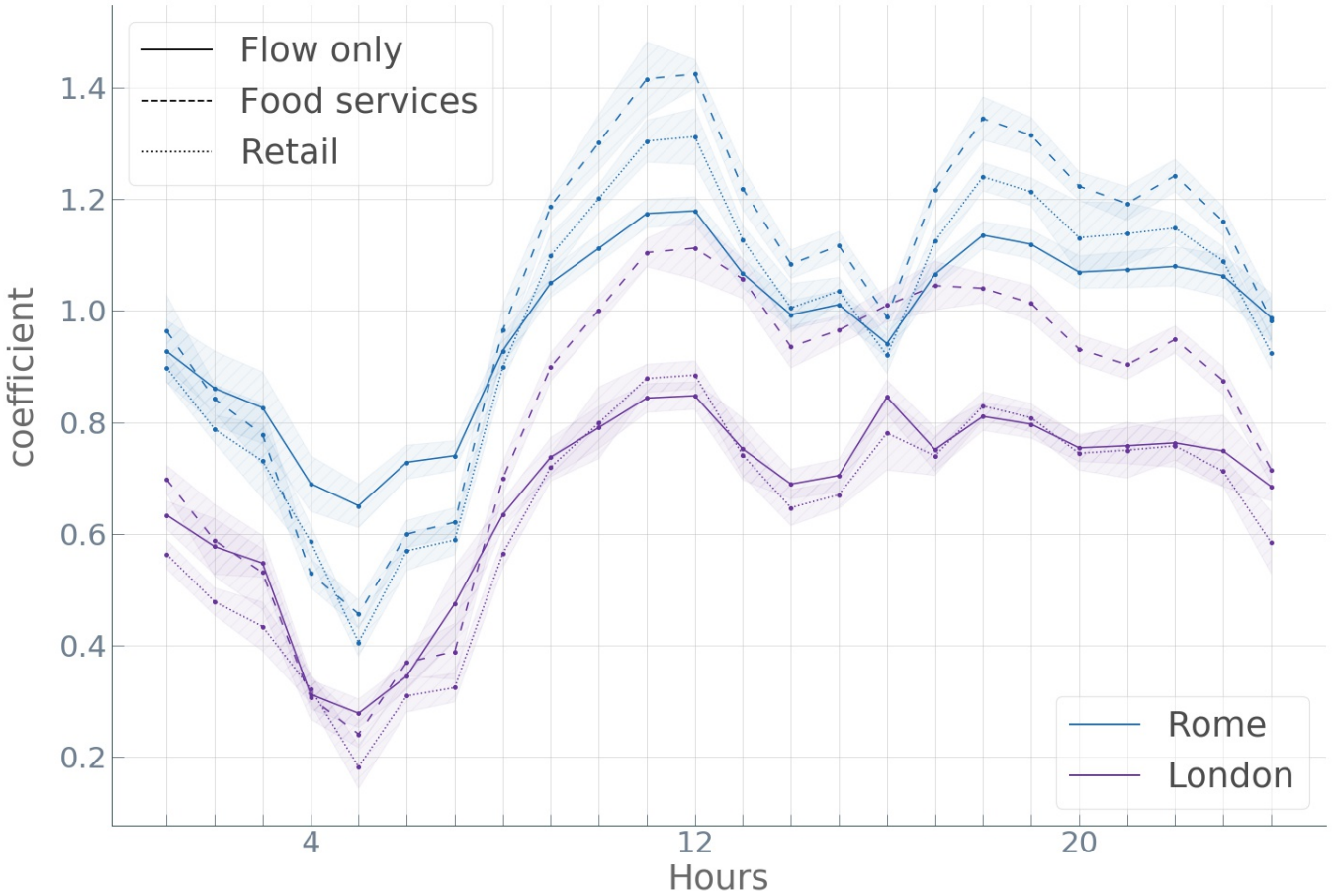
Hourly temporal regression results

Rome population densities, betweenness centrality, residential-to-other, and Airbnb price coefficients over time



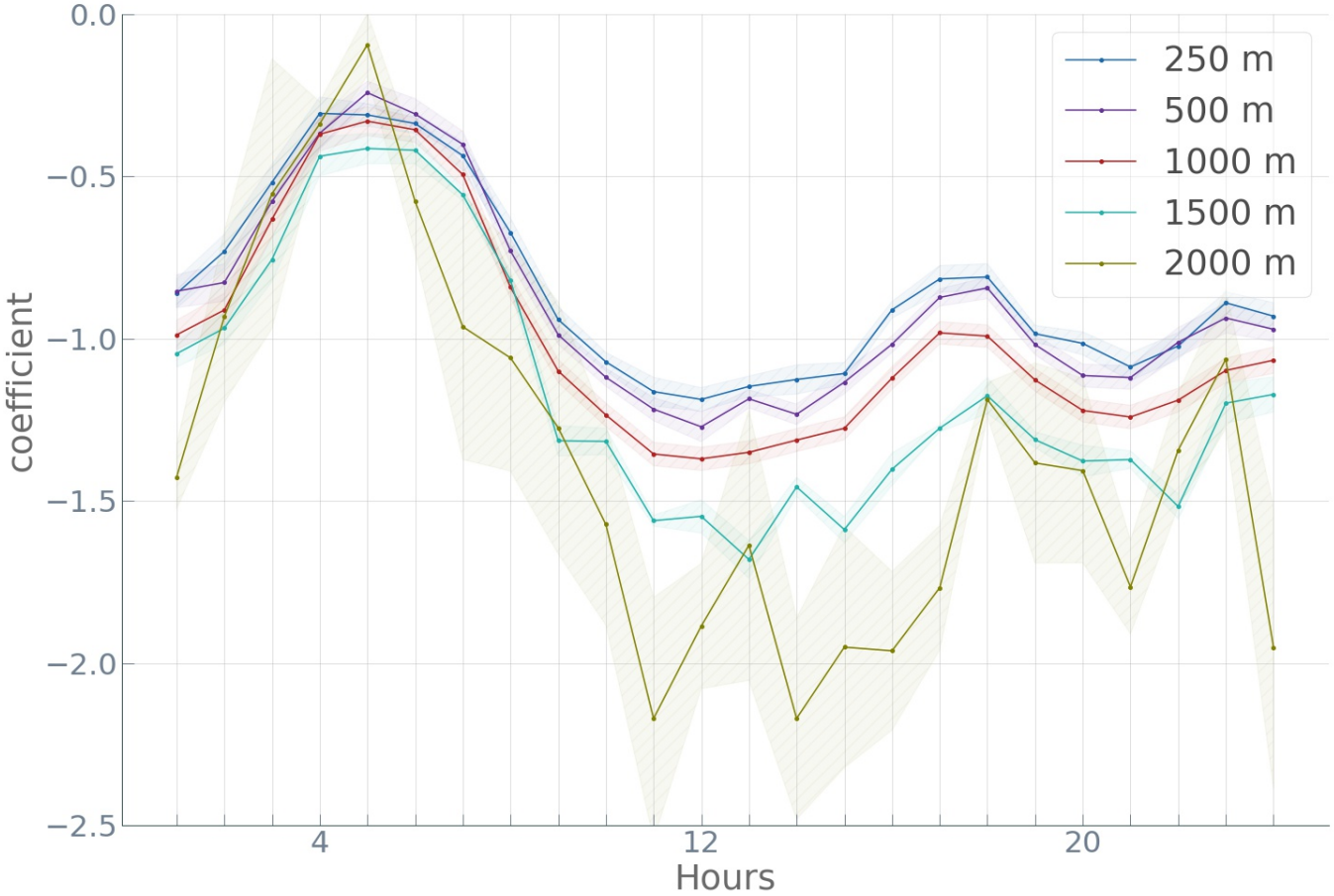
Hourly temporal regression results

Flow only, food services, and retail APA centrality coefficients over time in London and Rome



Hourly temporal regression results

The network distance regression parameter over time under different spatial grid resolutions in London



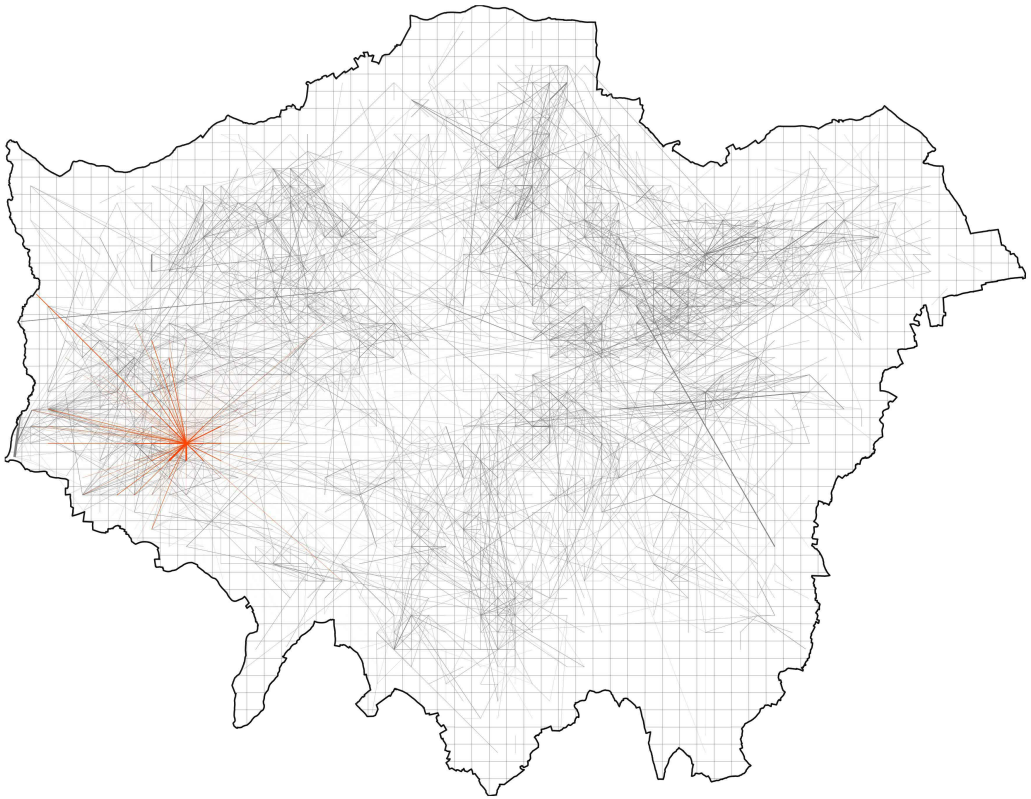
Urban Graph Neural Networks

Predicting urban mobility flows from urban socio-economic attributes and the urban OD network structure[1]

[1] G. Yeghikyan et al., Learning mobility flows from urban features with spatial interaction models and neural networks. arXiv preprint arXiv:2004.11924, 2020 (to appear in the Proceedings of IEEE International Conference on Smart Computing (SMARTCOMP 2020)

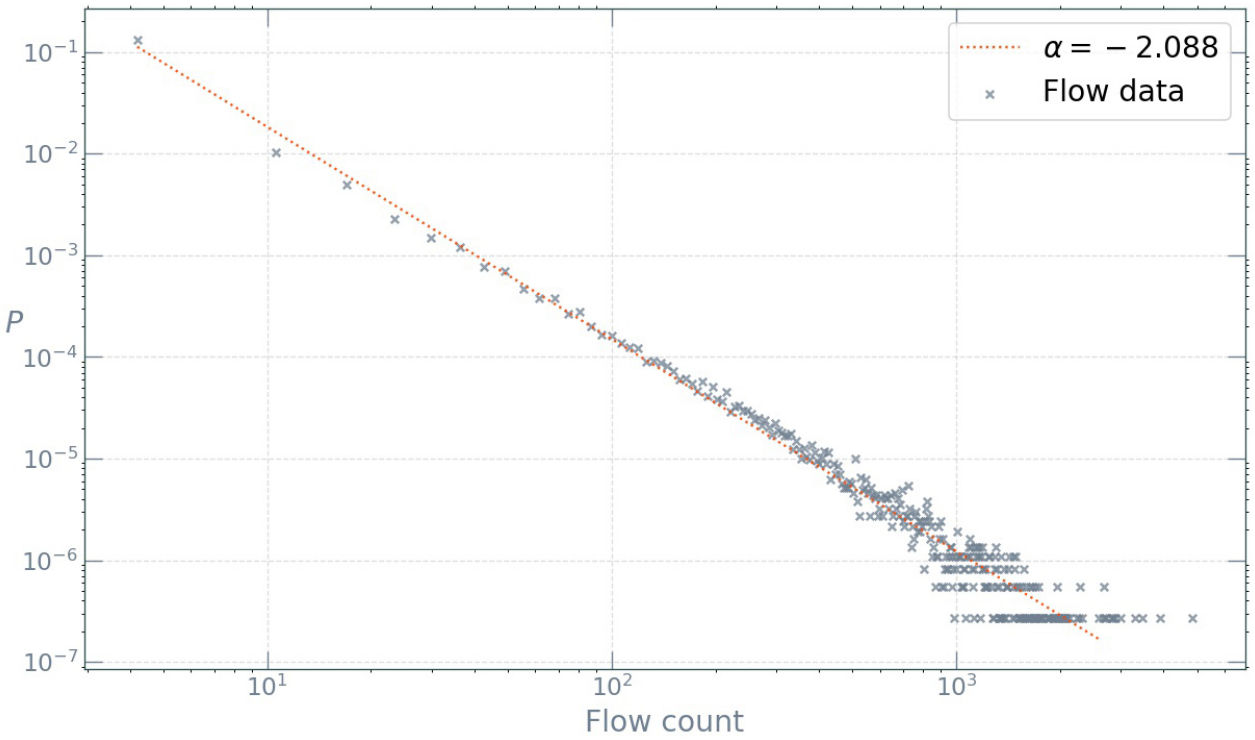
Urban Graph Neural Networks

Target flows between a node of interest and every other node in the London OD network:



Urban Graph Neural Networks

Log-log plot of the probability distribution of the London OD flows fitted with a power-law distribution $p(x) \propto x^{-\alpha}$ with exponent $\alpha = -2.088$



Goodness-of-fit measures

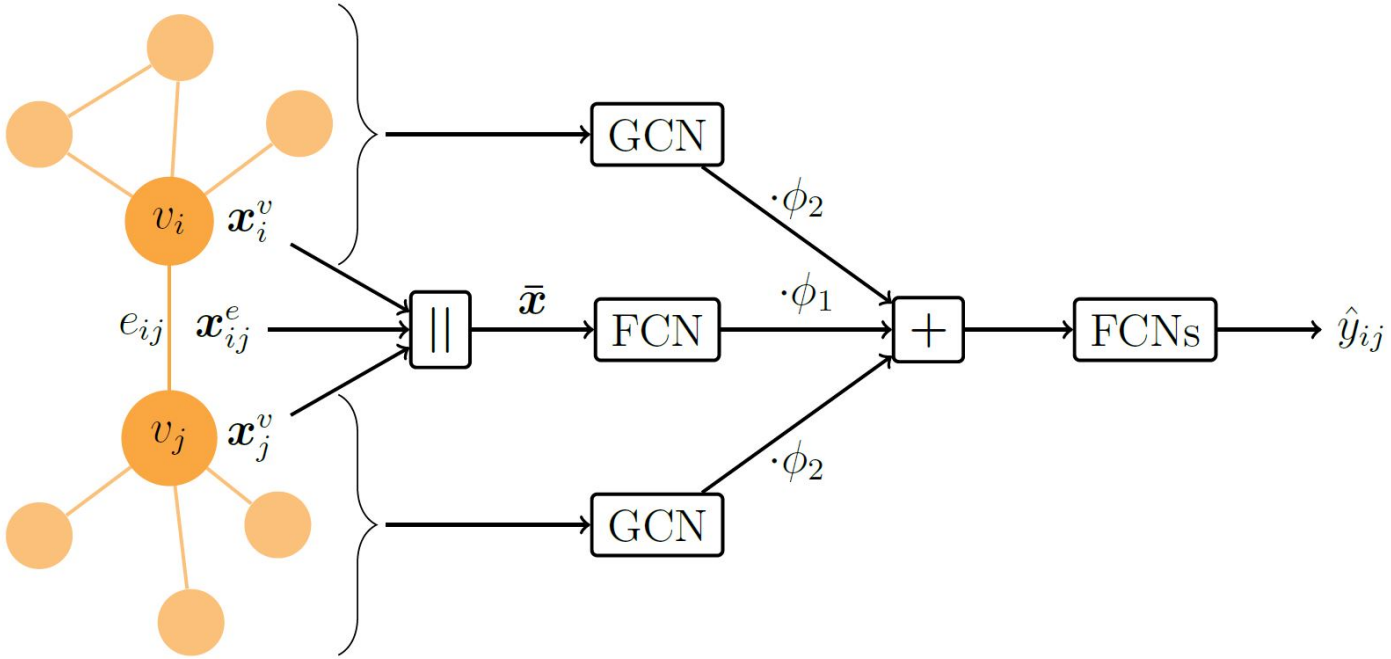
Mean Absolute Error (MAE):

$$\mathrm{MAE} = \frac{1}{|\mathcal{E}|} \sum_i \sum_j |y_{ij} - \hat{y}_{ij}|$$

4 bins with the following boundaries: $0 \leq 10.0 \leq 100.0 \leq 1000.0 \leq 10000.0$, corresponding to MAE_0, MAE_1, MAE_2, MAE_3, respectively.

$$\text{Bin mean MAE} = \frac{\mathrm{MAE}_0 + \mathrm{MAE}_1 + \mathrm{MAE}_2 + \mathrm{MAE}_3}{4}$$

Urban Graph Neural Network architecture:

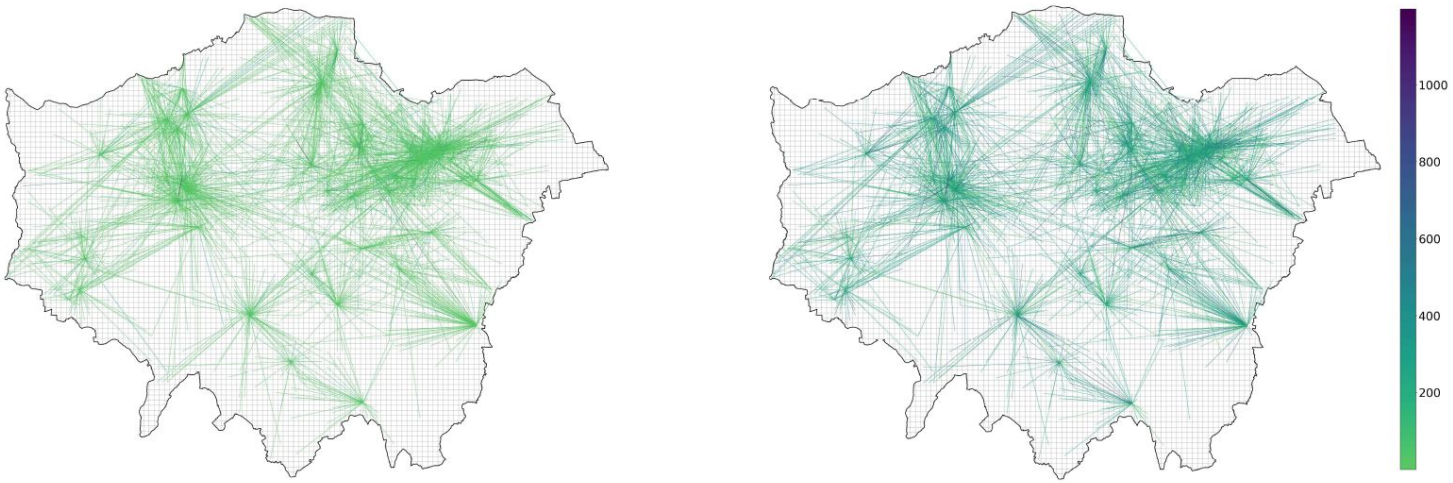


Urban Graph Neural Networks

Comparison of model performance in terms of MAE grouped by flow magnitude:

MAE	**Total**	$[0; 10)$	$[10; 10^2)$	$[10^2; 10^3)$	$[10^3; 10^4)$	**bin mean**
DC-GM	167.58	64.88	170.45	881.98	2176.35	823.42
Huff	122.89	48.21	99.86	511.41	1476.72	534.05
Poisson	106.74	40.69	88.56	475.23	1261.41	466.47
NB	92.62	33.02	76.96	431.44	1087.12	407.14
SAM	75.09	19.31	61.53	395.01	989.30	366.29
gHypE	58.11	9.02	53.10	346.96	832.26	310.34
XGBoost	31.59 ± 5.88	2.61 ± 0.89	45.12 ± 11.06	228.96 ± 39.96	549.83 ± 84.79	206.63 ± 34.18
FCNN	12.55 ± 0.91	0.33 ± 0.08	28.97 ± 4.93	161.12 ± 22.36	408.88 ± 36.59	149.82 ± 13.65
GNN-geo	13.34 ± 2.51	0.52 ± 0.40	31.63 ± 9.68	161.32 ± 9.09	422.04 ± 25.70	153.88 ± 9.74
GNN-flow	15.35 ± 4.23	0.63 ± 0.62	38.66 ± 16.65	170.06 ± 17.41	458.05 ± 64.56	166.85 ± 16.39
GNN-APA	9.51 ± 0.43	0.24 ± 0.05	20.26 ± 3.75	152.09 ± 7.14	399.90 ± 19.63	143.12 ± 6.88
GNN-APA-mpx	10.05 ± 0.51	0.28 ± 0.06	23.31 ± 4.30	155.89 ± 8.04	406.01 ± 21.12	146.37 ± 8.04

MAE residuals of flows associated with test nodes (a) GNN-geo. (b) XGBoost:



(a) (b)

THANK YOU!