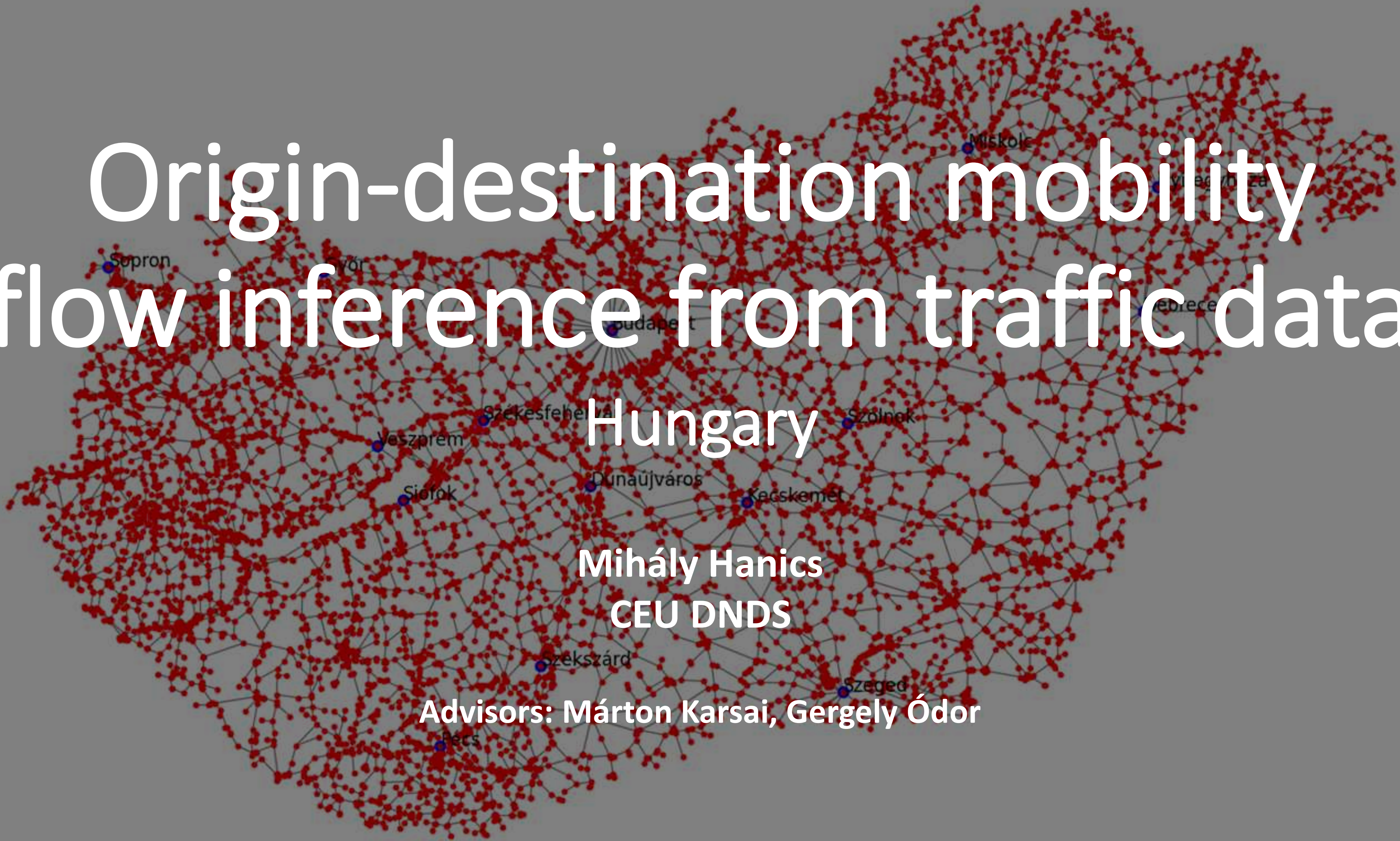
Origin-destination mobility flow inference from traffic data

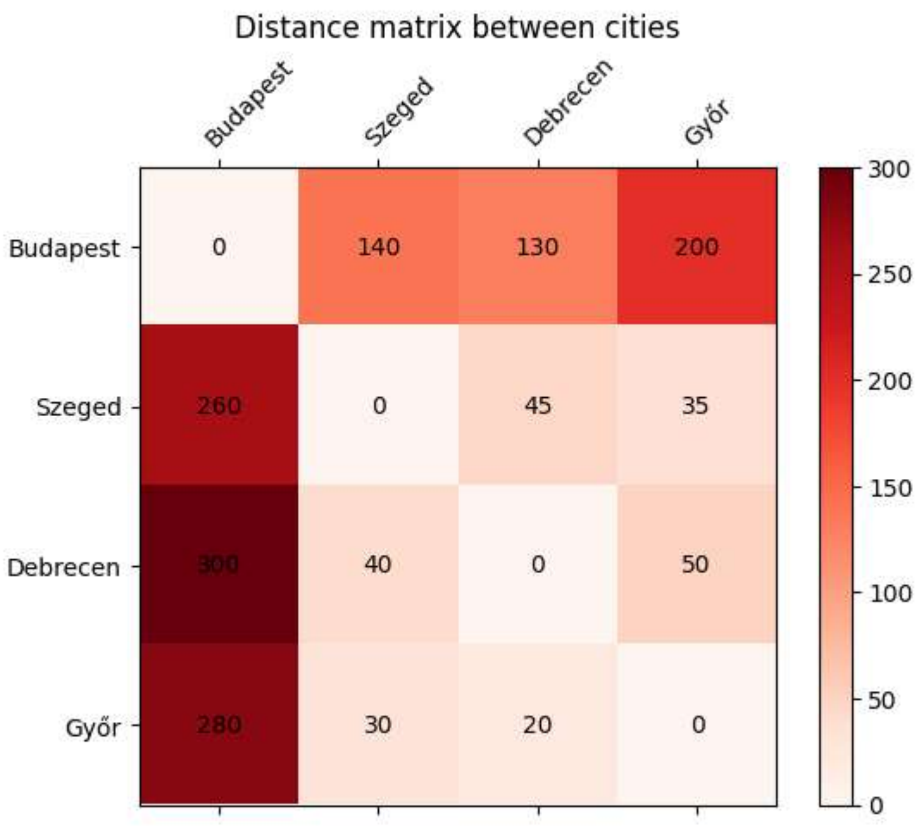
Hungary

Mihály Hanics
CEU DNDS

Advisors: Márton Karsai, Gergely Ódor



Inter-settlement origin-destination matrix



An example for an O-D matrix: rows represent the origin locations, columns the destinations. The stored value would equal the commuters inbetween

Problem: We would like to know how many people commute from a Hungarian city/town/settlement to others regularly.

Origin – destination pairs: For each location pair, the number of people travelling (commuting) between the two places over a time period (e.g. per day) is represented.

Usecases: Epidemic spreading modelling (our main interest), human mobility, traffic and traffic congestion prediction, human behaviour studies.

Challenges: Directly measured data is rare, inferring from certain types of data can raise other concerns (e.g. privacy) Census data is only collected once over multiple years.

Goal: An alternative way to create Hungary’s O-D matrix with high accuracy and great resolution.

Ground truth data from census

Commuting map in Hungary, from self-reported data [Ódor 2022]

Hungarian Central Statistical Office (KSH) 2016:

- Data 1: Settlement statistics: population, latitude-longitude
- Data 2: Commuting from one settlement to another:
  - includes commuting to school, and to work

Both datasets include each settlement / pair of settlements where there is at least one commuter.

~3200 settlements, ~93000 pairs (non-zero commution)

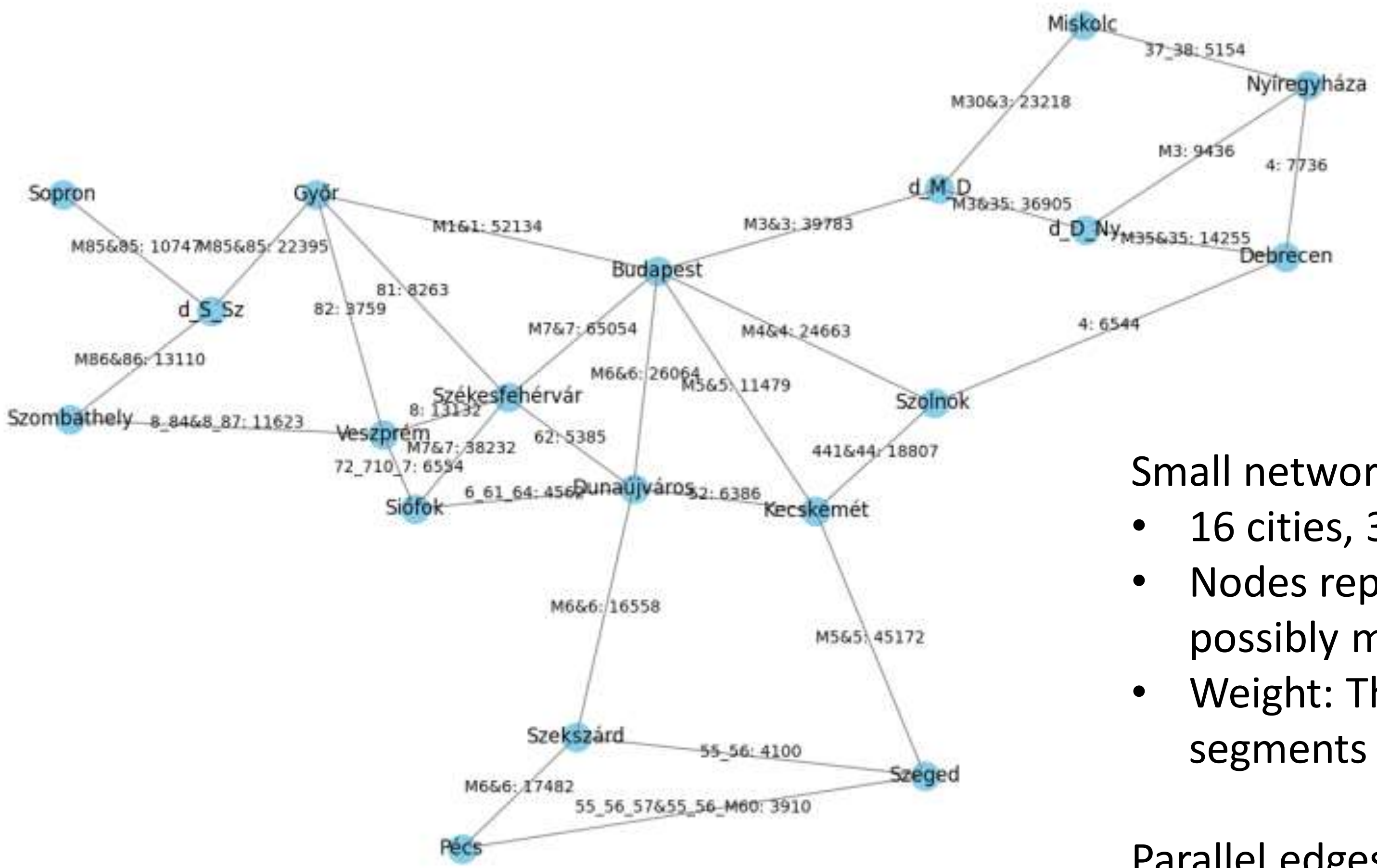
Observational data: Road traffic volume

Hungarian Public Roads 2022:

- Nation-wide daily public road traffic measurements, average over a year
- Vehicle, heavy vehicle, bicycle counts 14000 road segments, e.g. highways.
- Geographical data
- Some extra information on measurement

Közút száma	Útkategória	Vármegye	A számlálóállomás															
			szelvénye	érvényességi szakaszának					fek-vése	forgalom-jellege	típusa	forgalmi sávok száma	utolsó számlálás éve	adat forrása	számlált napok száma	pontosság	kódja
				határszelvényei		OKA csomópontjai		hossza									
						kezdő	vég										
			[km+m]	[km+m]	[km+m]			[km]								[+/- %]	
		vármegyei átlagérték						81,436									
3	I. rendű főút	Borsod-Abaúj-Zemplén vármegye	135+ 949	134 + 576	136 + 243	M000348	C050901	1,667	K	d 2	M1	2	2017	felszorzott		20,0%	**3024**
3	I. rendű főút	Borsod-Abaúj-Zemplén vármegye	138+ 855	136 + 243	142 + 379	C050901	C050102	6,135	K	c 2	M1	2	2017	felszorzott		20,0%	**5545**
3	I. rendű főút	Borsod-Abaúj-Zemplén vármegye	142+ 832	142 + 379	146 + 054	C050102	C050103	3,675	K	c 2	M1	2	2015	felszorzott		30,0%	**4462**
3	I. rendű főút	Borsod-Abaúj-Zemplén vármegye	147+ 269	146 + 054	148 + 857	C050103	C050104	2,803	K	c 2	M1	2	2017	felszorzott		20,0%	**1925**
3	I. rendű főút	Borsod-Abaúj-Zemplén vármegye	152+ 897	148 + 857	157 + 755	C050104	C050107	8,898	L	c 2	FCS+J	2	2022	mért	279	1,0%	**13534**
3	I. rendű főút	Borsod-Abaúj-Zemplén vármegye	162+ 843	157 + 755	165 + 064	C050107	C050110	7,309	K	c 2	M1	2	2017	felszorzott		20,0%	**7691**

Test network construction



„Proof of concept"

In accordance with the traffic data.

Small network:

- 16 cities, 3 highway division points
- Nodes represent the locations, edges are roads, possibly multiple roads
- Weight: The minimum traffic volume on the road segments represented by the edge

Parallel edges may appear between two nodes (e.g. road 4 and highway M4), these are combined into one (sum)

Models - methods

Gravity [Parsons 1960]

$$T_{ij} = K \frac{m_i^{\alpha} m_j^{\beta}}{{d_{ij}}^{k}}$$

where:

- T_{ij}: number of trips from zone i to zone j
- m_i , m_j relate to the number of trips leaving location i / attracted by location j
- d_{ij}: the distance between zones *i* and *j*
- deterrence: decreasing number of trips as the distance increases, by power-law
- K: constant of proportionality

For us: β, k estimated, $\alpha = 1$, K comes from the total trips sum

Bell [Bell 1983]

Other methods optimize an objective function subject to constraints:

$$\max f(\boldsymbol{t}) \qquad \text{assuming } \boldsymbol{v} = \boldsymbol{P}\boldsymbol{t}$$

where $\boldsymbol{t}$ contains the O-D pairs (J-size column vector) (**We infer it**)

- $\boldsymbol{v}$: Traffic volumes – **We measure it, input data** (I-size column vector)
– $\boldsymbol{P}$**-matrix**: When travelling from A to B, how likely you are to pass through a certain road – **This we pre-determine** ($\mathrm{I} \times J$ matrix)

v_i: road i traffic, p_{ij}: probability of going through road i given the O-D pair j

Bell-model: maximizes the probability of observing $\boldsymbol{t}$ given initial probabilities

We use the scikit-mobility models to estimate an O-D matrix [Pappalardo 2022]

Models - methods

Gravity [Parsons 1960]

$$T_{ij} = K \frac{m_i^{\alpha} m_j^{\beta}}{{d_{ij}}^{k}}$$

where:

- T_{ij}: number of trips from zone i to zone j
- m_i , m_j relate to the number of trips leaving location i / attracted by location j
- d_{ij}: the distance between zones *i* and *j*
- deterrence: decreasing number of trips as the distance increases, by power-law
- K: constant of proportionality

For us: β, k estimated, $\alpha = 1$, K comes from the total trips sum

Bell [Bell 1983]

Other methods optimize an objective function subject to constraints:

$$\max f(\boldsymbol{t}) \qquad \text{assuming } \boldsymbol{v} = \boldsymbol{Pt}$$

where $\boldsymbol{t}$ contains the O-D pairs (J-size column vector) (**We infer it**)

- $\boldsymbol{v}$: Traffic volumes – **We measure it, input data** (I-size column vector)
- $\boldsymbol{P}$**-matrix**: When travelling from A to B, how likely you are to pass through a certain road – **This we pre-determine** (I $\times$ J matrix)

Bell-model: maximizes the probability of observing $\boldsymbol{t}$ given probabilities $\boldsymbol{q}$

$$f(\boldsymbol{t}) = \frac{\left(\sum_j t_j\right)!}{\prod_j (t_j!)} \cdot \prod_j {q_j}^{t_j} \qquad \text{(multinomial distribution)}$$

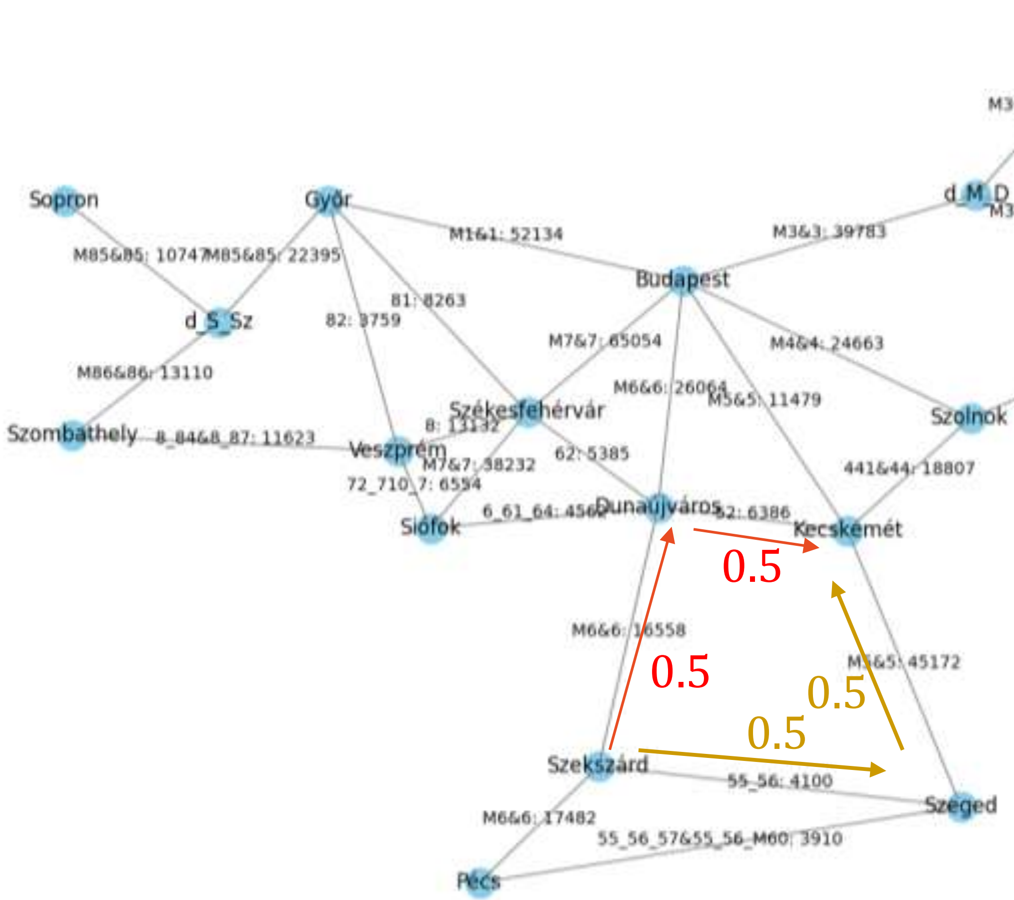
We use the scikit-mobility models to estimate an O-D matrix [Pappalardo 2022]

Bell-model in practice

Optimization: $\max f(\boldsymbol{t})$ assuming $\boldsymbol{v} = \boldsymbol{P}\boldsymbol{t}$

Bell: $\max \mathrm{P}(\boldsymbol{t} \mid \boldsymbol{q}) \rightarrow f(\boldsymbol{t}) = \frac{(\sum_j t_j)!}{\prod_j (t_j!)} \cdot \prod_j {q_j}^{t_j}$

assuming multinomial distributions



$$\begin{pmatrix} \cdots & 0 & \cdots \\ & \vdots & \\ \cdots & 0.5 & \cdots \\ & \vdots & \\ \cdots & 0.5 & \cdots \\ \iddots & \vdots & \ddots \end{pmatrix}$$

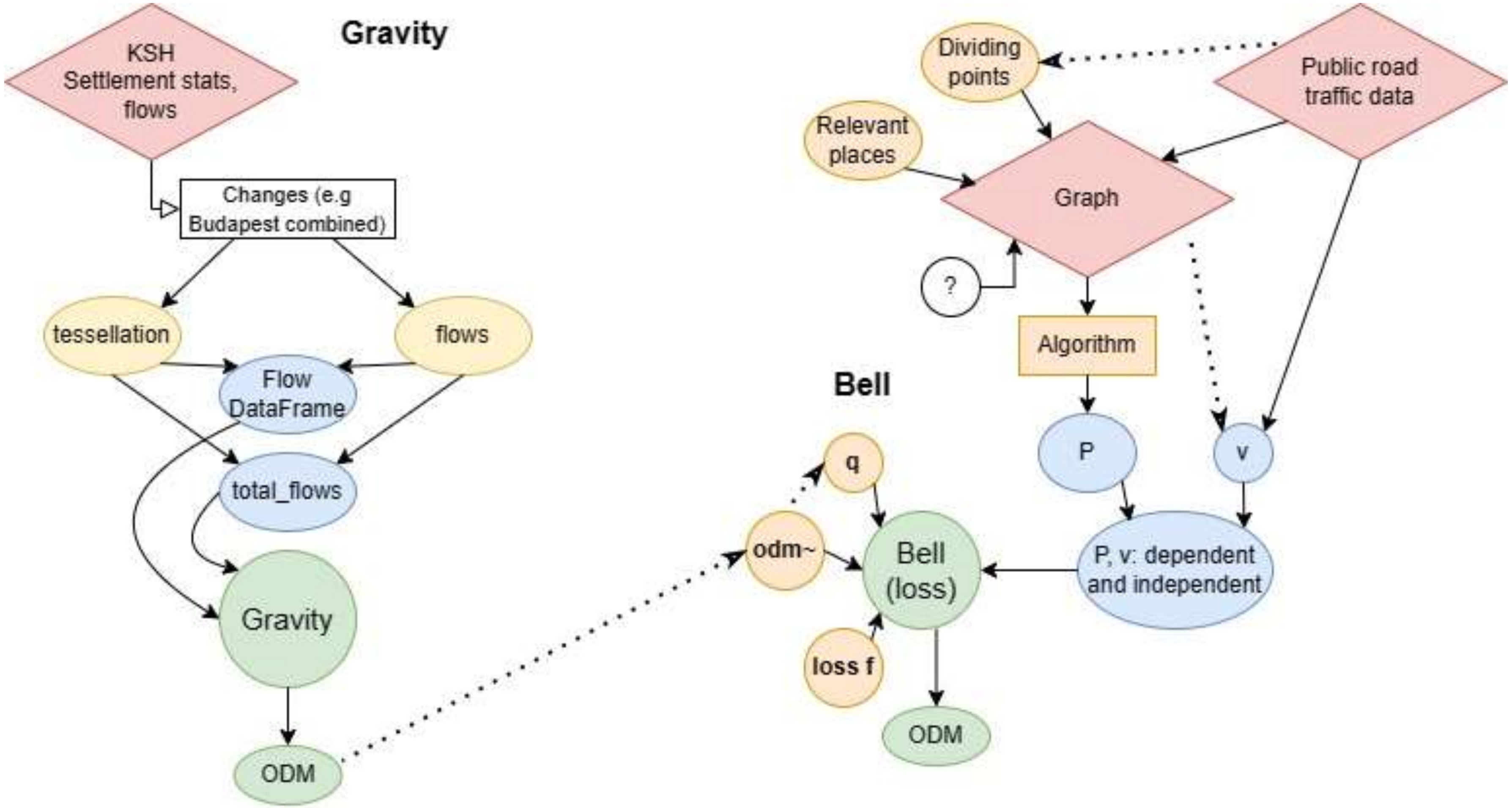
$\boldsymbol{q}$: Estimated by normalizing an initial guess: $q_j = \frac{\tilde{t_j}}{\sum_i \tilde{t_i}}$

$\boldsymbol{P}$: Shortest paths - Between all pairs of locations, all shortest paths are computed.

For each of the *N* shortest paths between two locations, $\frac{1}{N}$ is added to a path's edges

Some constraints in practice conflict -> use them as loss (L1 mostly)

Gravity, Bell models pipeline



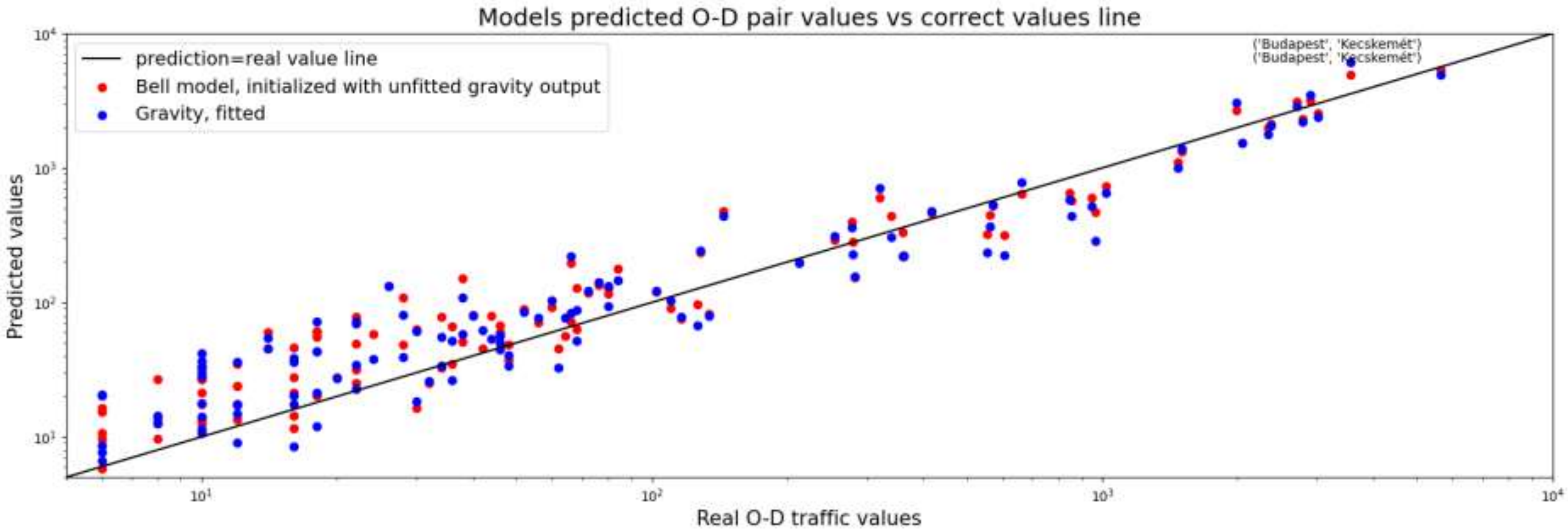
Possible "fusion" of ODMs

Results on the small graph

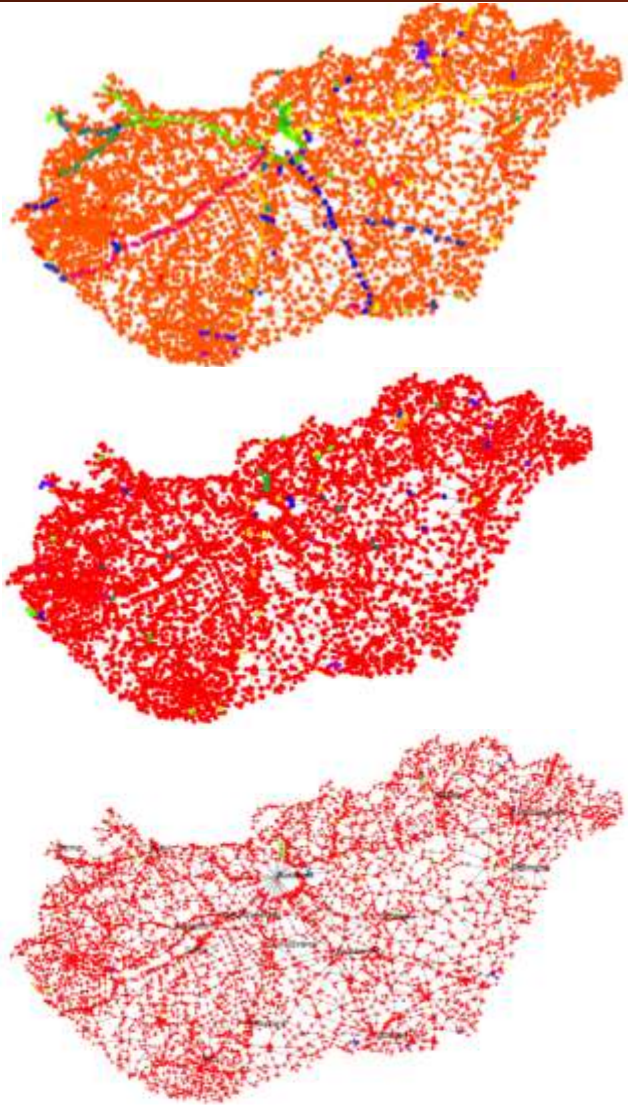
	Correlation	MSE	MAE	Relative diff^2	Better prediction
Gravity	95.8%	62479	~124	1103	49
Bell from Gravity	97.4%	37298	~94	1138	71

Bell model improves on the initial model

This is true for the gravity model, but other tested models: entropy min/max. Superior especially in large O-D values.



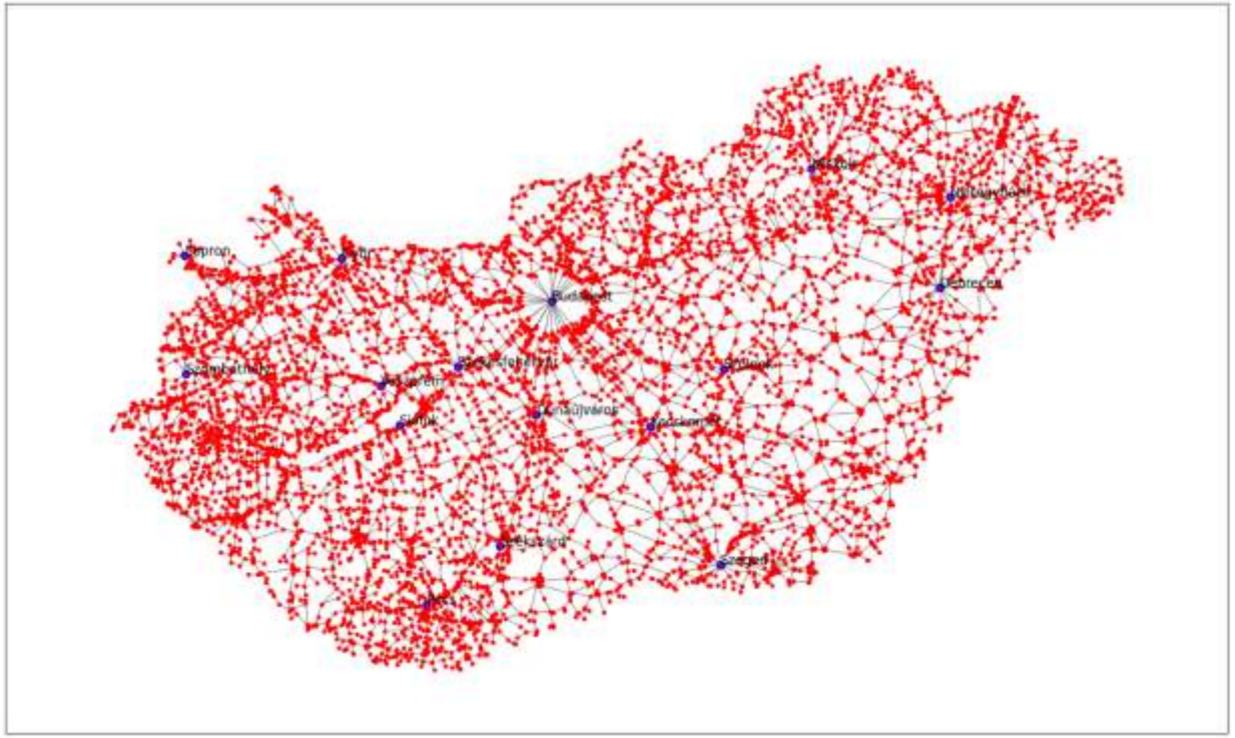
Next steps, future directions



Scale up the network to as large as we can: all roads, more settlements

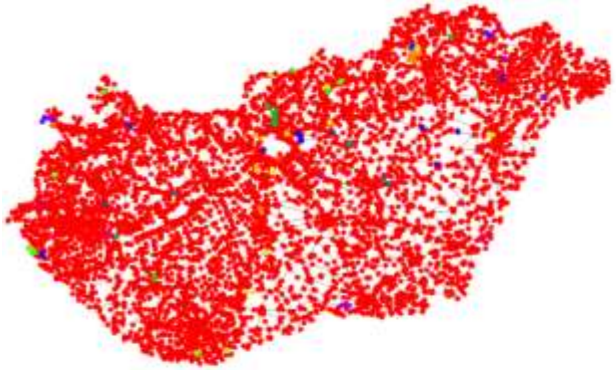
Issues:

- Technical issues e.g. different components (see figures on left)



The connected network from adding new edges and combining nodes.

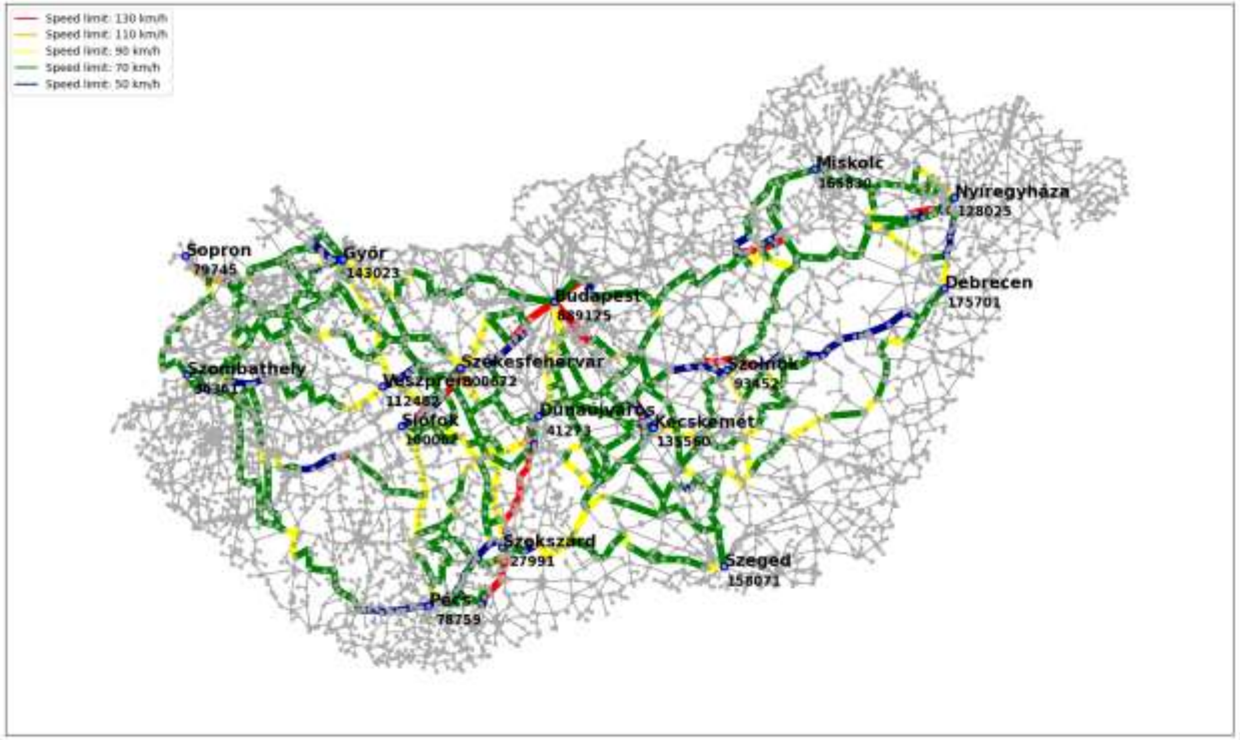
Next steps, future directions



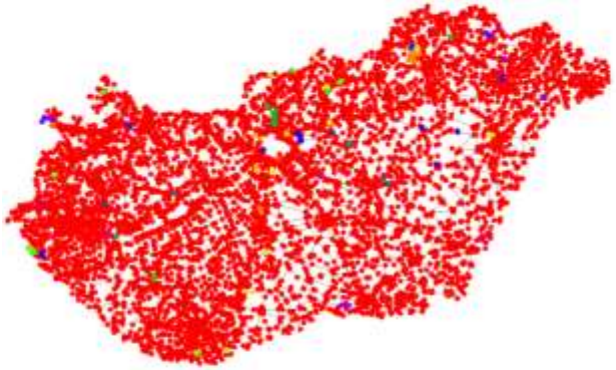
Scale up the network to as large as we can: all roads, more settlements

Issues:

– P-matrix creation: shortest path works weakly + we need multiple found paths



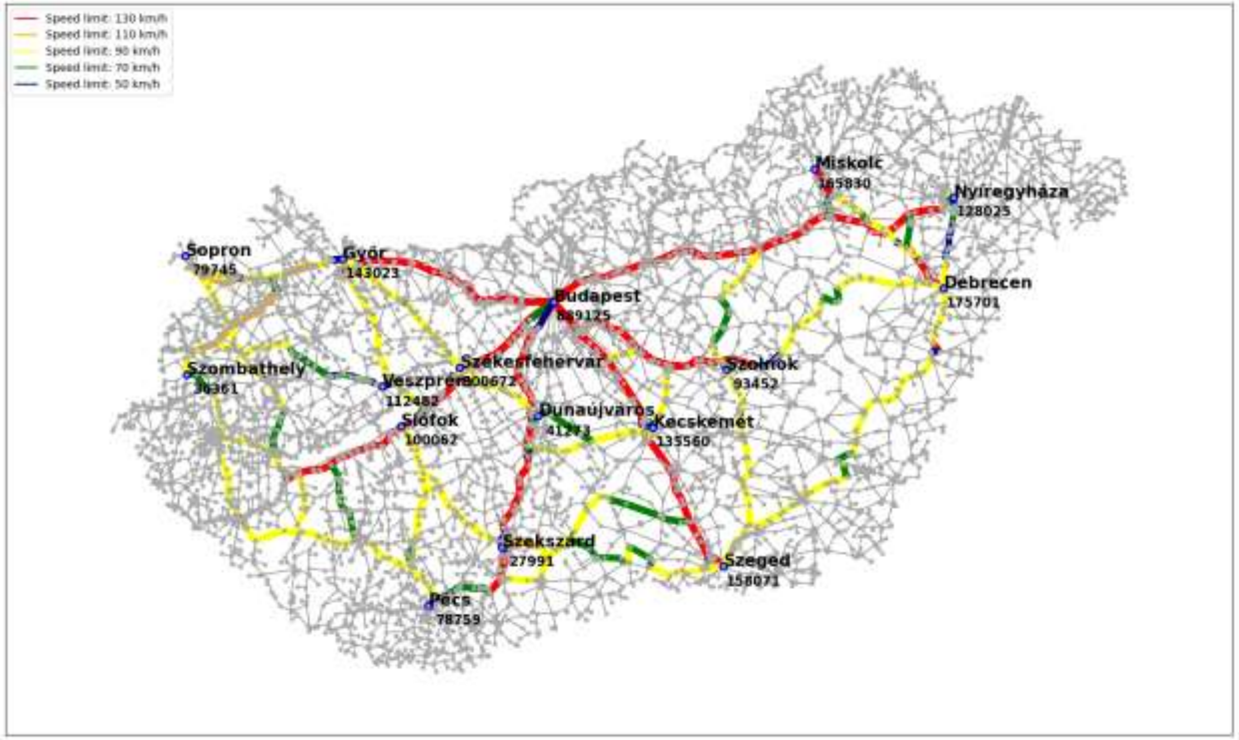
Next steps, future directions



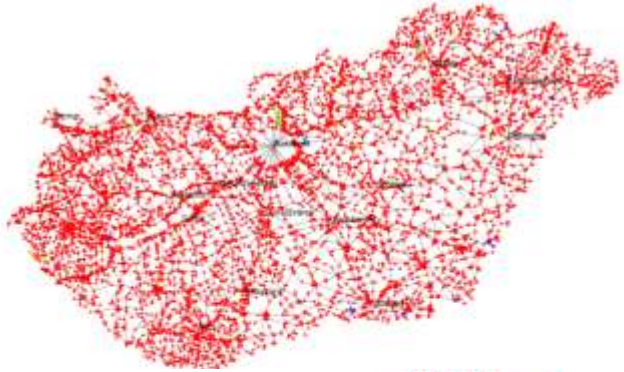
UPDATE:

Shortest time routes seem to work well.

One issue that came up is that now the used roads are „too crowded”-> low O-D value predictions



Next steps, future directions



Scale up the network to as large as we can: all roads, more settlements

Issues:

- Computation time: optimization ~46 sec for 16 locations, grows fast with more locations
  - Network creation: finding intersections took 25 mins, overall procedure 30+ mins

Run the analysis for monthly traffic data between 2016 and 2022

References

[Ódor 2021] Ódor G, Czifra D, Komjáthy J, Lovász L, Karsai M. Switchover phenomenon induced by epidemic seeding on geometric networks. Proceedings of the National Academy of Sciences. 2021;118(41):e2112607118. pmid:34620714

[Pappalardo 2022] Pappalardo, L., Simini, F., Barlacchi, G., & Pellungrini, R. (2022). scikit-mobility: A Python Library for the Analysis, Generation, and Risk Assessment of Mobility Data. Journal of Statistical Software, 103(1), 1–38.

[Parsons 1960] Parsons, Christopher R. 'Do Migrants Really Foster Trade'. *The Trade-Migration Nexus, a Panel Approach* 2000 (1960).

[Bell 1983] Bell, M. G. (1983). The estimation of an origin-destination matrix from traffic counts. Transportation Science, 17(2), 198-217.

Data:

Nemzeti Adathozzáférési Pont – Magyar Közút Nonprofit Zrt.: Éves keresztmetszeti forgalmi adatok (2022)
https://napportal.kozut.hu/
Központi Statisztikai Hivatal - https://www.ksh.hu/