

Supplementary Information

A Link model approach to identify congestion hotspots

Aleix Bassolas, Sergio Gómez and Alex Arenas

List of Figures

S1	Evolution of the order parameter η in DT+MST graphs as a function of the injection rate ρ for each of the models and several values of R_{DT}	3
S2	Analysis of congestion hotspots in Amsterdam.	4
S3	Analysis of congestion hotspots in Brussels.	4
S4	Analysis of congestion hotspots in Madrid.	5
S5	Analysis of congestion hotspots in Miami.	5
S6	Analysis of congestion hotspots in Mumbai.	6
S7	Analysis of congestion hotspots in Paris.	6
S8	Analysis of congestion hotspots in Pittsburgh.	7
S9	Analysis of congestion hotspots in Sao Paulo.	7
S10	Analysis of congestion hotspots in Seattle.	8
S11	Analysis of congestion hotspots in Taipei.	8
S12	Analysis of congestion hotspots in Toronto.	9
S13	Analysis of congestion hotspots in Washington.	9
S14	Correlation between the real and modelled delays in the node dynamics when destinations are distributed according to the community POIs.	10
S15	Correlation between the real and modelled delays in the link dynamics when destinations are distributed according to the shopping POIs.	11
S16	Correlation between the real and modelled delays in the node dynamics when destinations are distributed according to the shopping POIs.	11
S17	Correlation between the real and modelled delays in the link dynamics when destinations are distributed according to the food POIs.	12
S18	Correlation between the real and modelled delays in the node dynamics when destinations are distributed according to the food POIs.	12
S19	Correlation between the real and modelled delays in the link dynamics when destinations are distributed according to the entertainment POIs.	13
S20	Correlation between the real and modelled delays in the node dynamics when destinations are distributed according to the entertainment POIs.	13
S21	Analysis of the normalized mean squared error for the Correlation the delays in the node dynamics when destinations are distributed according to the community POIs.	14
S22	Analysis of the normalized mean squared error for the Correlation the delays in the node dynamics when destinations are distributed according to the community POIs.	14
S23	Analysis of the normalized mean squared error for the Correlation the delays in the link dynamics when destinations are distributed according to the shopping POIs.	15
S24	Analysis of the normalized mean squared error for the Correlation the delays in the node dynamics when destinations are distributed according to the shopping POIs.	15
S25	Residual analysis for the regression between the travel times in the link dynamics when destinations are distributed according to the community POIs.	16

S26	Residual analysis for the regression between the delays in the link dynamics when destinations are distributed according to the community POIs.	16
S27	Residual analysis for the regression between the travel times in the node dynamics when destinations are distributed according to the community POIs. .	16
S28	Residual analysis for the regression between the delays in the node dynamics when destinations are distributed according to the community POIs.	17
S29	Distribution of residuals for the regression between the travel times in the link dynamics when destinations are distributed according to the community POIs.	17
S30	Distribution of residuals for the regression between the delays in the link dynamics when destinations are distributed according to the community POIs.	17
S31	Distribution of residuals for the regression between the travel times in the node dynamics when destinations are distributed according to the community POIs.	18
S32	Distribution of residuals for the regression between the delays in the node dynamics when destinations are distributed according to the community POIs.	18

List of Tables

S1	Pearson correlation between the traffic counts in Madrid and the flow of vehicles according to our model. The analysis has been performed including the 10% of links with highest flow.	10
----	---	----

1 Congestion phenomena in the DT+MST model

In this section we assess the congestion in the DT+MST model developed in [1, 2] that mimics the structure of real cities with a more densely connected center and a sparse periphery. A set of points is homogeneously distributed in a 2D space of size $L \times L$ and are connected according to the the Delaunay triangulation (DT) [3]. Within a distance smaller than R_{DT} from the center of the domain the network is fully preserved while for the region with radius greater than R_{DT} most of the links are removed in order to keep the maximum spanning tree (MST) that maximizes the betweenness centrality. The value of R_{DT} determines the underlying structure of the spatial graph, low values leading to network dominated by the MST and vice-versa for high values.

In Fig. S1 we report η as a function of ρ for different values of R_{DT} in networks of 500 nodes distributed in a space of 80×80 . To better compare the networks we plot them as a function of \tilde{R}_{DT} which is calculated as $\frac{R_{DT}}{L/2}$. The overall trend seems to indicate that congestion decreases the DT region increases, likely because that there are wider paths alternatives as compared to the MST. In the context of the results for the cost-driven networks shown in Fig. 4 of the main text, congestion seems to increase in a more progressive way, likely as a consequence of the multiple regimes of betweenness centrality already found in [1, 2].

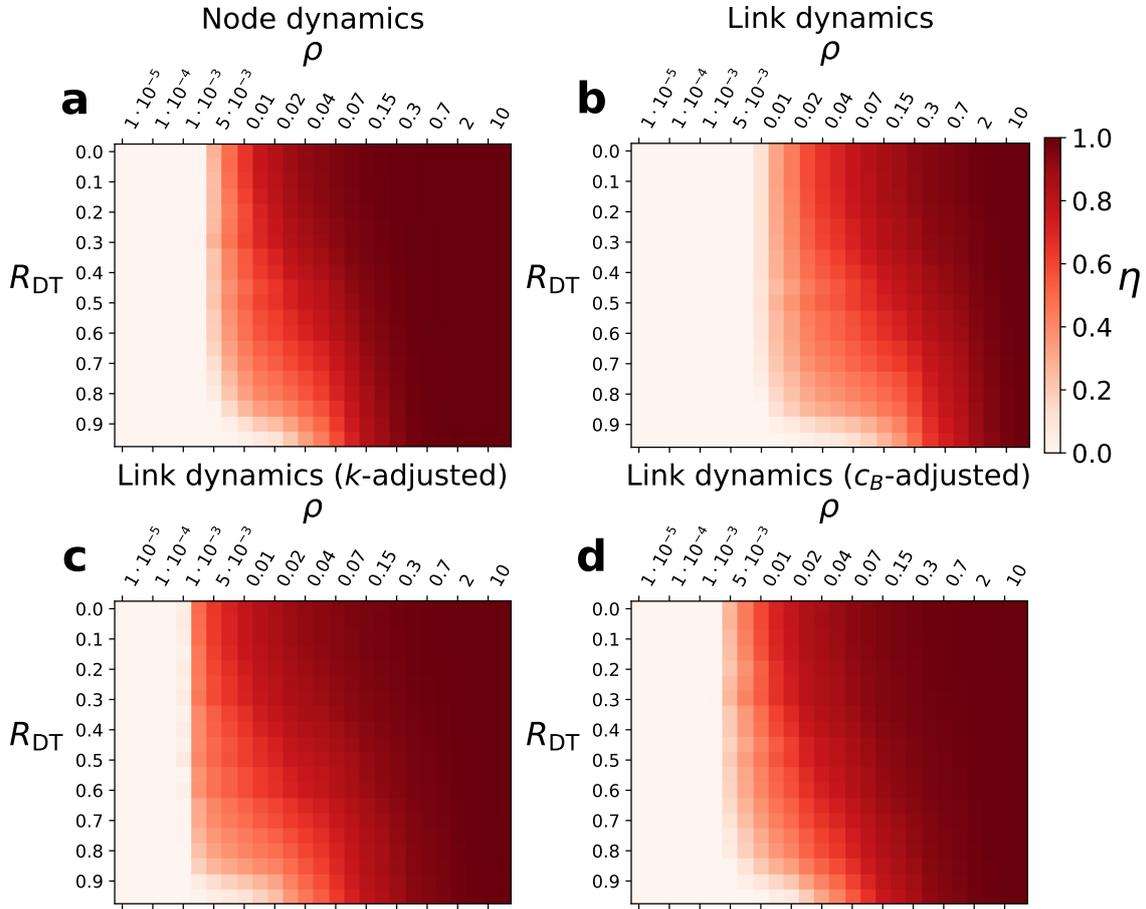


Figure S1: **Evolution of the order parameter η in DT+MST graphs as a function of the injection rate ρ for each of the models and several values of R_{DT} .** Evolution of $\eta(\rho)$ for **a** the node model, **b** the link model without capacity normalization ($\tau_{ij} = \tau_i$), **c** the link model with k -adjusted normalization ($\tilde{\tau}_{ij}^k$) and **d** the link model with c_B -adjusted normalization ($\tilde{\tau}_{ij}^{c_B}$). The parameter R_{DT} is normalized by $l/2$ where l is the side size of the squared domain $l \times l$.

2 Spatial distribution of congestion hotspots

We report here the congestion hotspots for the rest of cities analyzed in the case of the node model and the link model with a capacity adjusted for node degree. More concretely we have respectively in Figs. S2-S13 the results for Amsterdam, Brussels, Madrid, Miami, Mumbai, Paris, Pittsburgh, Sao Paulo, Seattle, Taipei, Toronto and Washington.

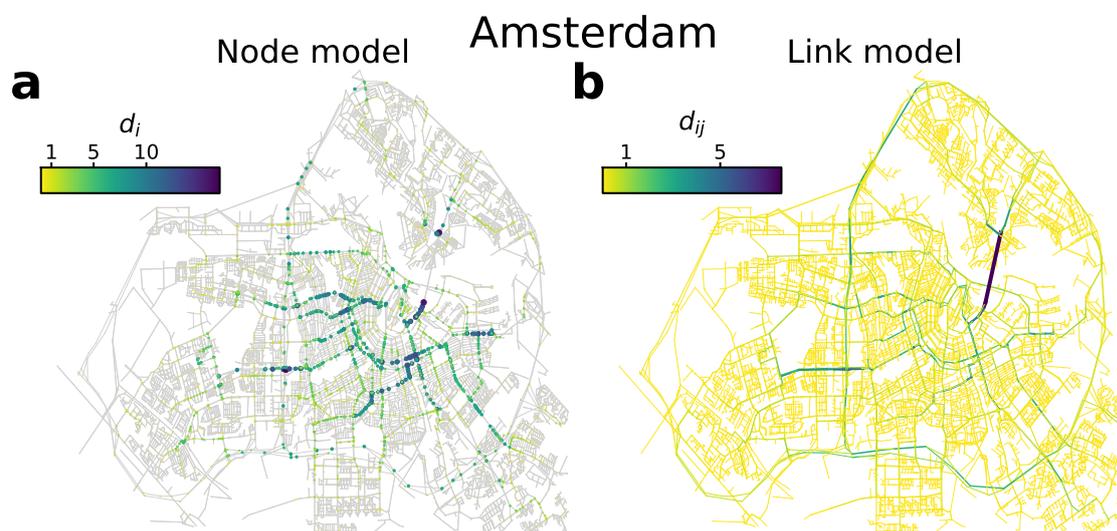


Figure S2: **Analysis of congestion hotspots in Amsterdam.** Congestion hotspots observed in Amsterdam for **a** the node model and **b** the link model with destinations distributed according to community venues. Both maps were generated with $\eta_{\text{data}} = 0.26$.

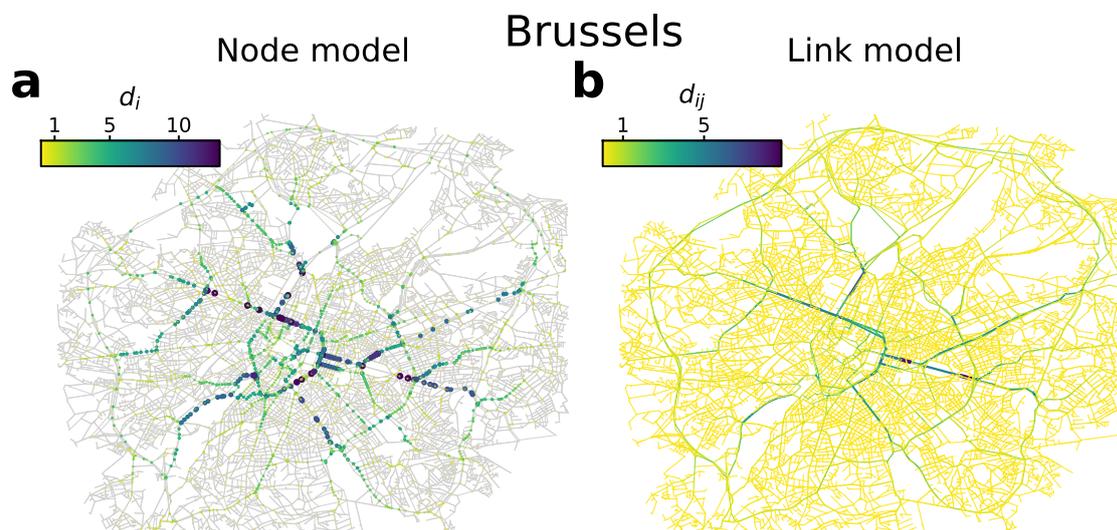


Figure S3: **Analysis of congestion hotspots in Brussels.** Congestion hotspots observed in Brussels for **a** the node model and **b** the link model with destinations distributed according to community venues. Both maps were generated with $\eta_{\text{data}} = 0.38$.

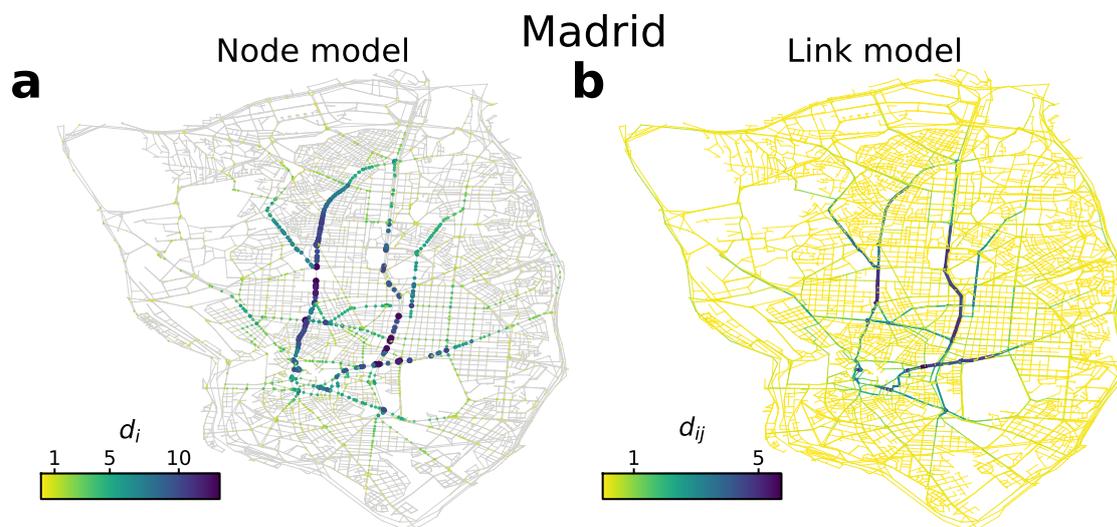


Figure S4: **Analysis of congestion hotspots in Madrid.** Congestion hotspots observed in Madrid for **a** the node model and **b** the link model with destinations distributed according to community venues. Both maps were generated with $\eta_{\text{data}} = 0.23$.

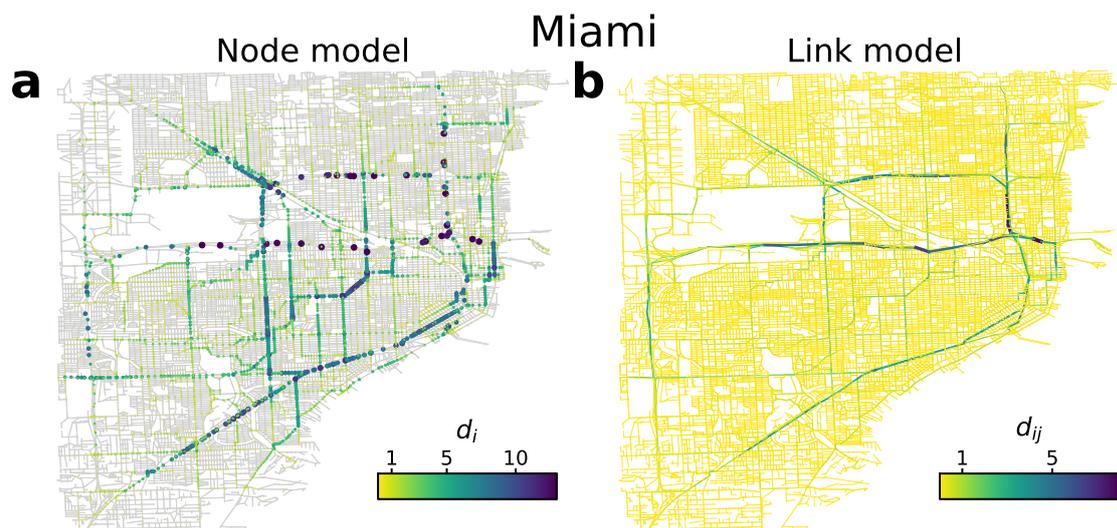


Figure S5: **Analysis of congestion hotspots in Miami.** Congestion hotspots observed in Miami for **a** the node model and **b** the link model with destinations distributed according to community venues. Both maps were generated with $\eta_{\text{data}} = 0.31$.

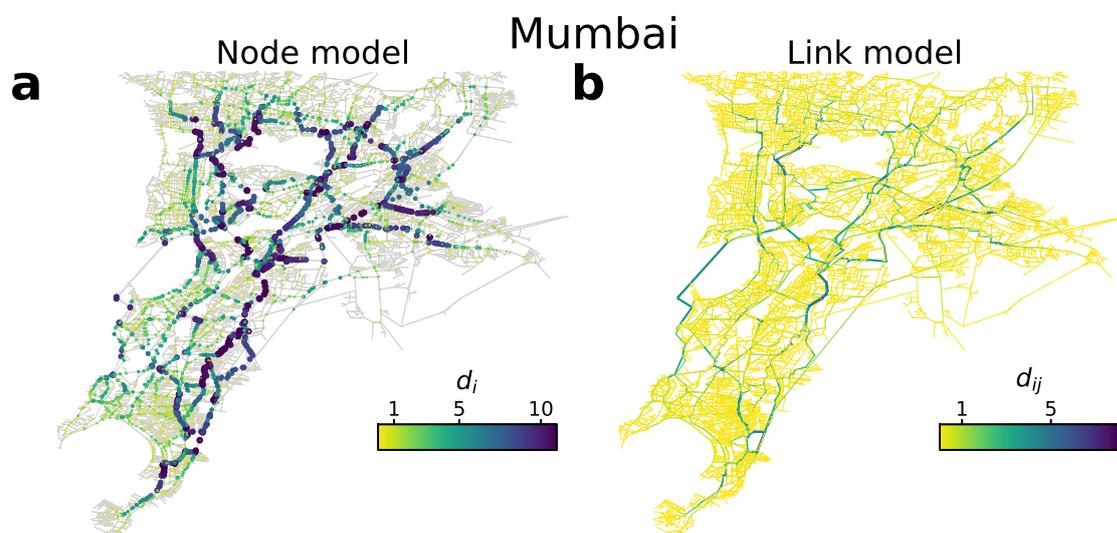


Figure S6: **Analysis of congestion hotspots in Mumbai.** Congestion hotspots observed in Mumbai for **a** the node model and **b** the link model with destinations distributed according to community venues. Both maps were generated with $\eta_{\text{data}} = 0.65$.

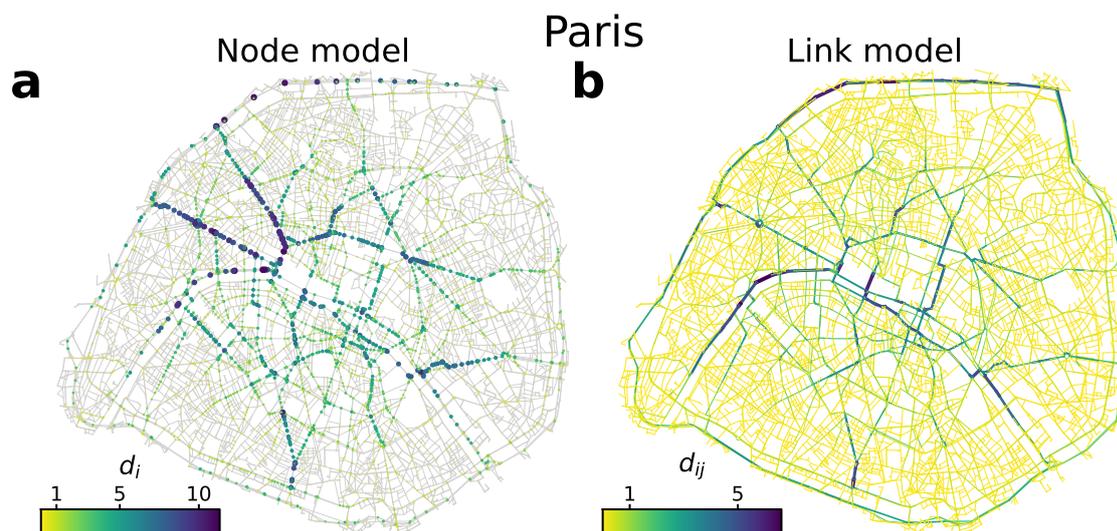


Figure S7: **Analysis of congestion hotspots in Paris.** Congestion hotspots observed in Paris for **a** the node model and **b** the link model with destinations distributed according to community venues. Both maps were generated with $\eta_{\text{data}} = 0.39$.

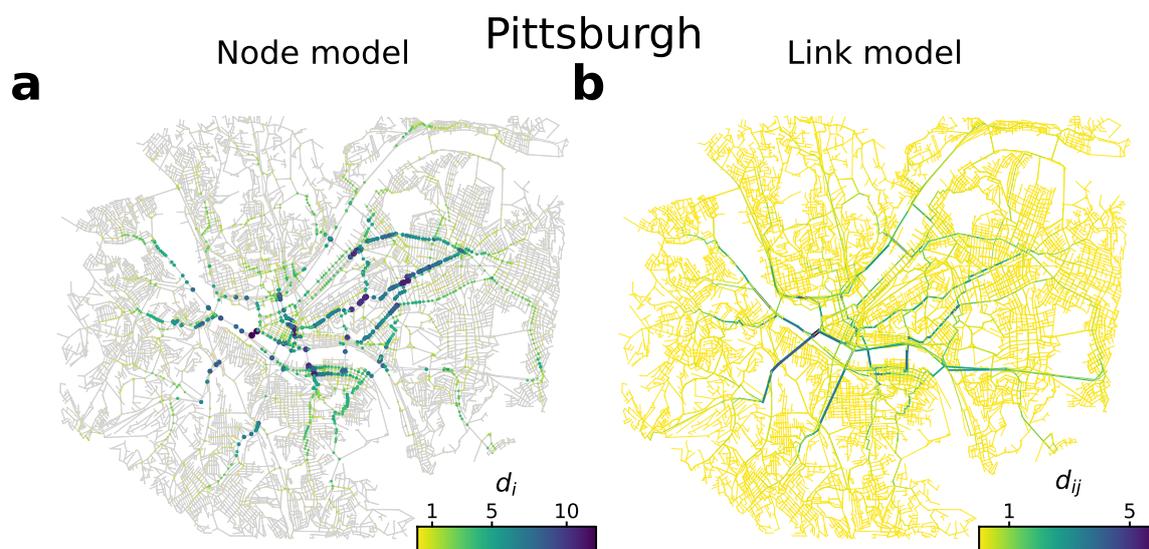


Figure S8: **Analysis of congestion hotspots in Pittsburgh.** Congestion hotspots observed in Pittsburgh for **a** the node model and **b** the link model with destinations distributed according to community venues. Both maps were generated with $\eta_{\text{data}} = 0.21$.

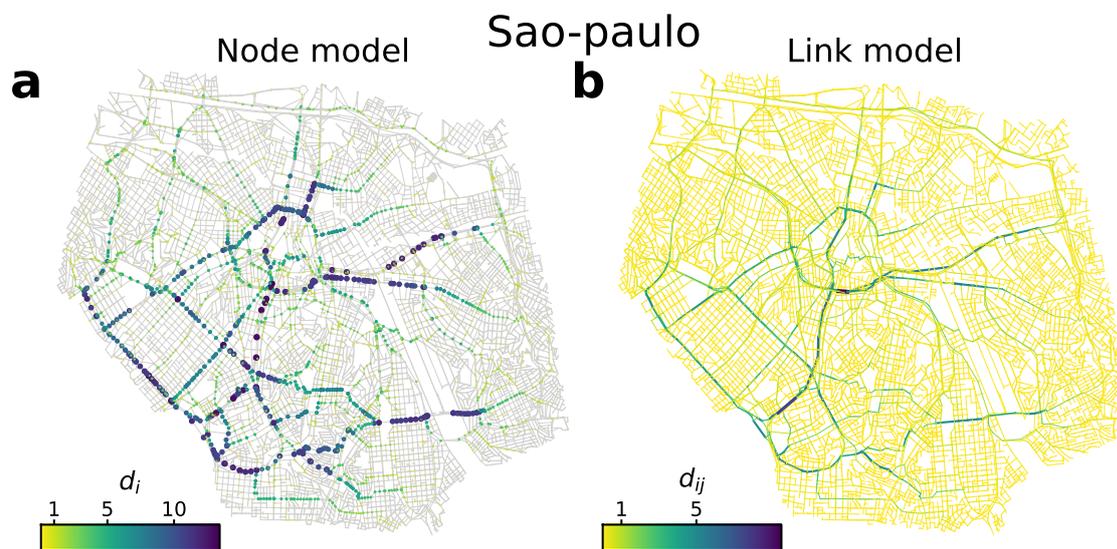


Figure S9: **Analysis of congestion hotspots in Sao Paulo.** Congestion hotspots observed in Sao Paulo for **a** the node model and **b** the link model with destinations distributed according to community venues. Both maps were generated with $\eta_{\text{data}} = 0.45$.

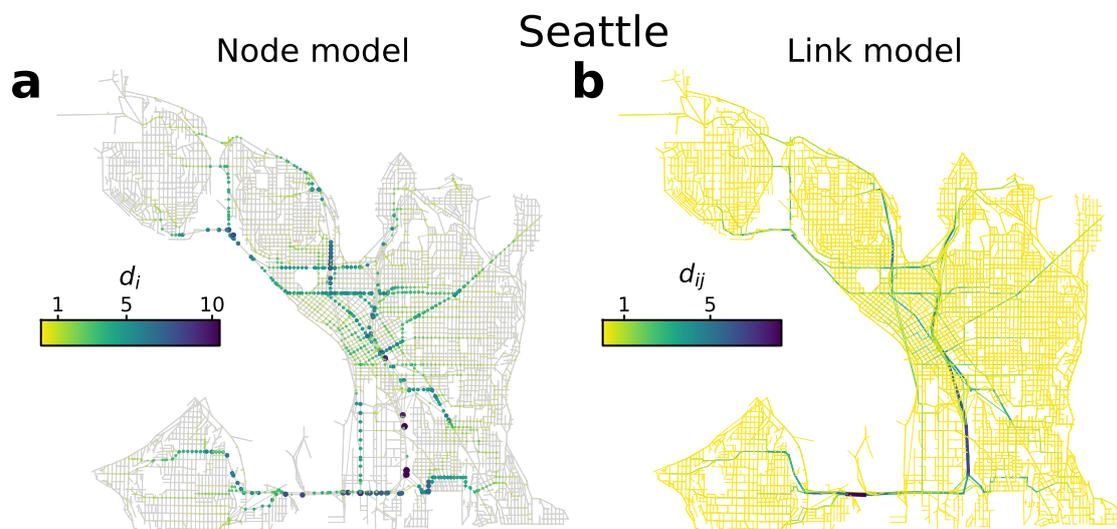


Figure S10: **Analysis of congestion hotspots in Seattle.** Congestion hotspots observed in Seattle for **a** the node model and **b** the link model with destinations distributed according to community venues. Both maps were generated with $\eta_{\text{data}} = 0.31$.

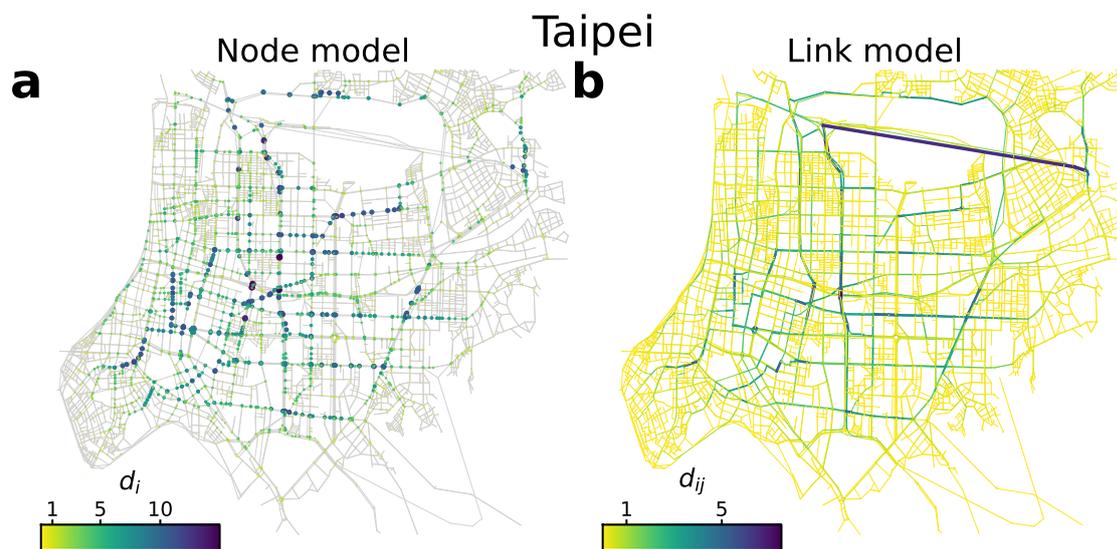


Figure S11: **Analysis of congestion hotspots in Taipei.** Congestion hotspots observed in Taipei for **a** the node model and **b** the link model with destinations distributed according to community venues. Both maps were generated with $\eta_{\text{data}} = 0.35$.

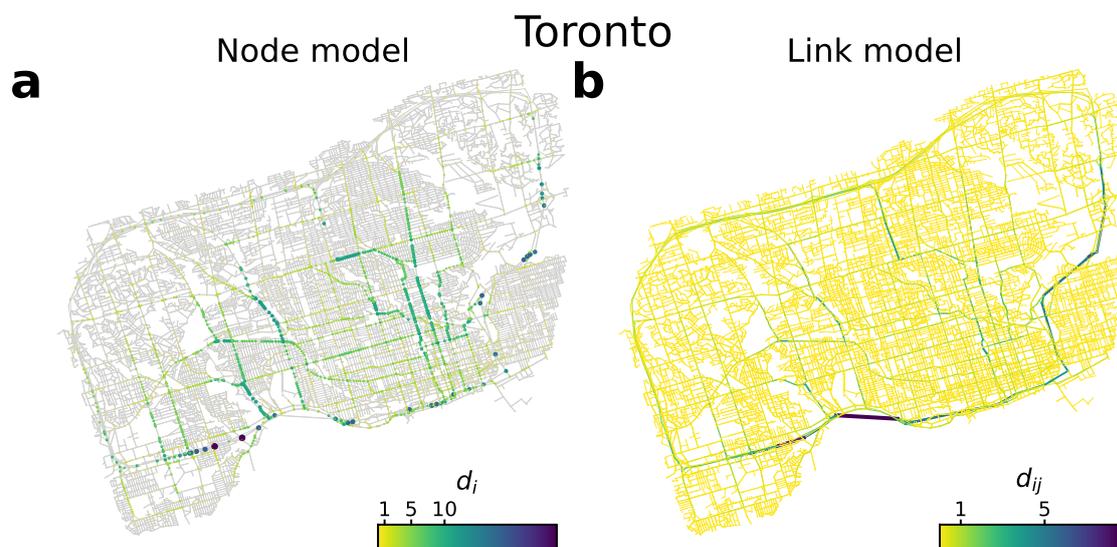


Figure S12: **Analysis of congestion hotspots in Toronto.** Congestion hotspots observed in Toronto for **a** the node model and **b** the link model with destinations distributed according to community venues. Both maps were generated with $\eta_{\text{data}} = 0.33$.

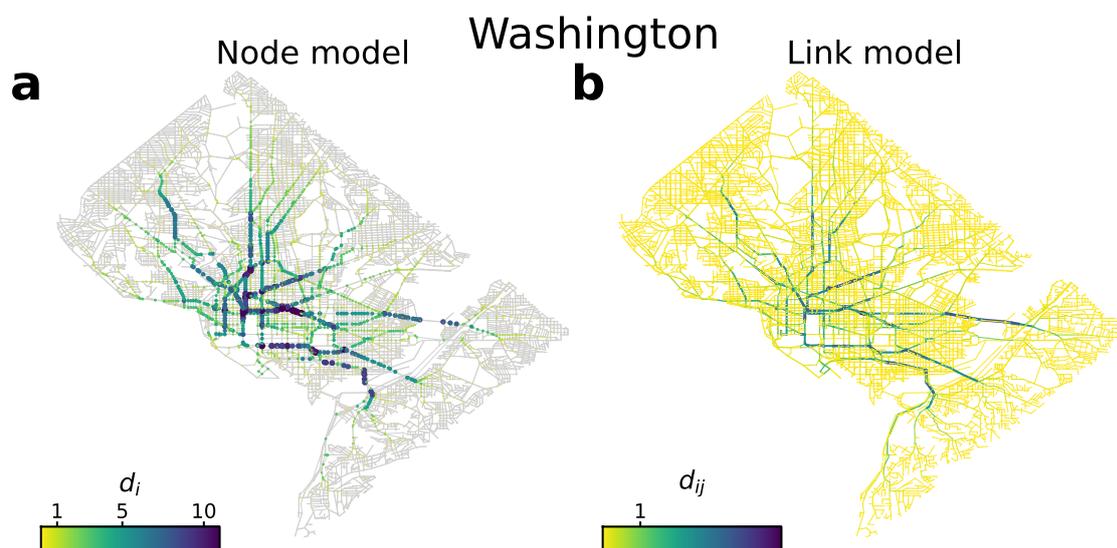


Figure S13: **Analysis of congestion hotspots in Washington.** Congestion hotspots observed in Washington for **a** the node model and **b** the link model with destinations distributed according to community venues. Both maps were generated with $\eta_{\text{data}} = 0.29$.

Table S1: Pearson correlation between the traffic counts in Madrid and the flow of vehicles according to our model. The analysis has been performed including the 10% of links with highest flow.

Venue	r_P
Community	0.49***
Outdoors	0.30***
Nightlife	0.39***
Shopping	0.41***
Food	0.40***
Travel	0.44***
Entertainment	0.43***

3 Correlation with traffic counts and observed delays

We analyze here the correlations obtained for the expected delay with a different distribution of destinations and other models. In Fig. S15 we provide the results for the link model when destinations are distributed according to the shopping venues where we observe that there is also an increase on the prediction power from its non-delayed counterpart. By focusing only the delay, there is also a significant correlation similar to the results in Fig. 7 of the main paper. For comparison we display in Figs. S14 and S16 the same analysis for the node model. As it is shown there, although the correlations are still present, they are lower than for the link model, specially if we focus only on the delay. We have also analyzed the case of food venues (Figs. S17 and S18) and entertainment venues (Figs. S19 and S20). Additional results regarding the normalized mean squared error is shown in Figs. S21, S22, S23 and S24 We also provide in Figs. S30, S29, S26, S25 the analysis of residuals for the travel times and delay regression in the case of community venues for the link model and in Figs. S32, S31, S28, S27 for the node model.

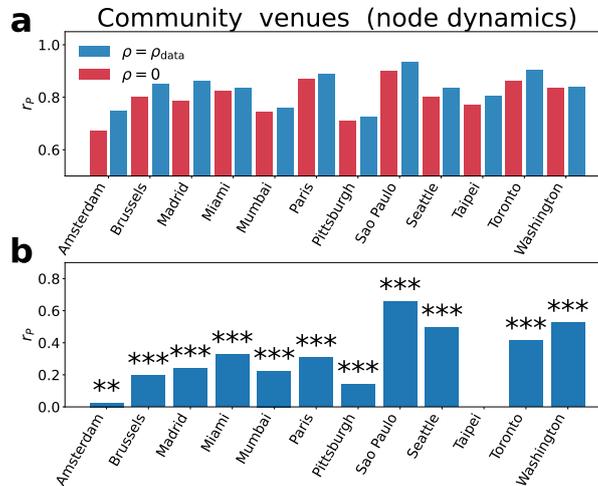


Figure S14: **Correlation between the real and modelled delays in the node dynamics when destinations are distributed according to the community POIs.** (a) Comparison between the Pearson correlation coefficient obtained between the travel times from Uber Data [4] during the morning peak (8 – 10am) in a set of cities and the travel times obtained for $\rho = 0$ (red) and $\rho = \rho_{\text{data}}$ (blue). (b) Pearson correlation coefficient between the delay observed in the data and in the model. Asterisks indicate the level of significance (* p-value < 0.05, ** p-value < 0.01, *** p-value < 0.001). The injection rate for each city ρ_{data} is set to match η with the percentage of delay observed in the Tom Tom traffic index data [5].

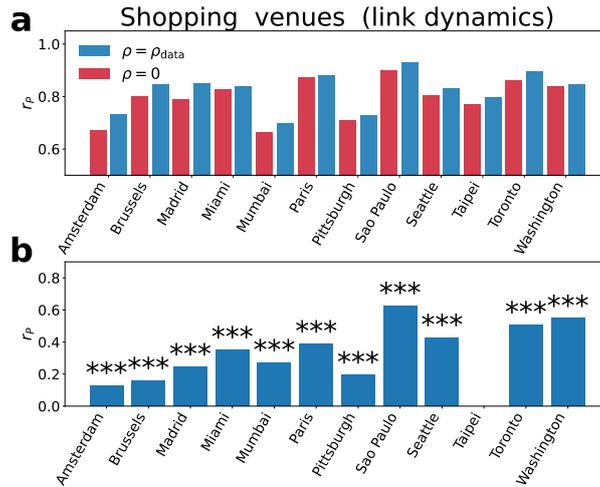


Figure S15: **Correlation between the real and modelled delays in the link dynamics when destinations are distributed according to the shopping POIs.** (a) Comparison between the Pearson correlation coefficient obtained between the travel times from Uber Data [4] during the morning peak (8 – 10am) in a set of cities and the travel times obtained for $\rho = 0$ (red) and $\rho = \rho_{\text{data}}$ (blue). (b) Pearson correlation coefficient between the delay observed in the data and in the model. Asterisks indicate the level of significance (* p-value < 0.05, ** p-value < 0.01, *** p-value < 0.001). The injection rate for each city ρ_{data} is set to match η with the percentage of delay observed in the Tom Tom traffic index data [5].

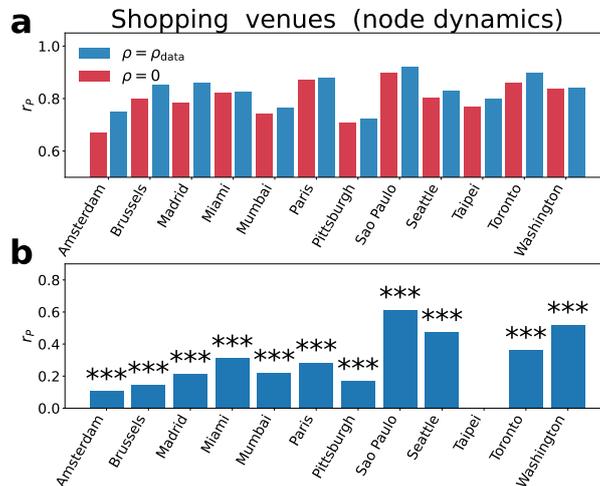


Figure S16: **Correlation between the real and modelled delays in the node dynamics when destinations are distributed according to the shopping POIs.** (a) Comparison between the Pearson correlation coefficient obtained between the travel times from Uber Data [4] during the morning peak (8 – 10am) in a set of cities and the travel times obtained for $\rho = 0$ (red) and $\rho = \rho_{\text{data}}$ (blue). (b) Pearson correlation coefficient between the delay observed in the data and in the model. Asterisks indicate the level of significance (* p-value < 0.05, ** p-value < 0.01, *** p-value < 0.001). The injection rate for each city ρ_{data} is set to match η with the percentage of delay observed in the Tom Tom traffic index data [5].

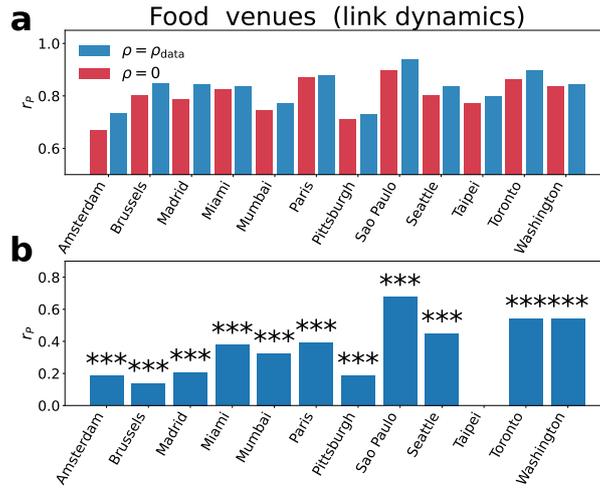


Figure S17: **Correlation between the real and modelled delays in the link dynamics when destinations are distributed according to the food POIs.** (a) Comparison between the Pearson correlation coefficient obtained between the travel times from Uber Data [4] during the morning peak (8–10am) in a set of cities and the travel times obtained for $\rho = 0$ (red) and $\rho = \rho_{\text{data}}$ (blue). (b) Pearson correlation coefficient between the delay observed in the data and in the model. Asterisks indicate the level of significance (* p-value < 0.05, ** p-value < 0.01, *** p-value < 0.001). The injection rate for each city ρ_{data} is set to match η with the percentage of delay observed in the Tom Tom traffic index data [5].

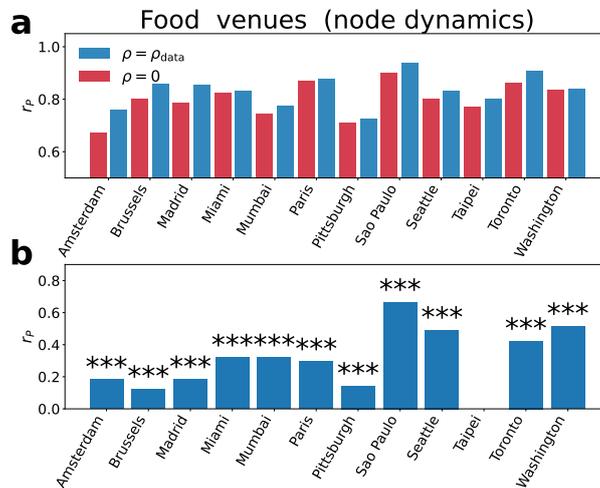


Figure S18: **Correlation between the real and modelled delays in the node dynamics when destinations are distributed according to the food POIs.** (a) Comparison between the Pearson correlation coefficient obtained between the travel times from Uber Data [4] during the morning peak (8–10am) in a set of cities and the travel times obtained for $\rho = 0$ (red) and $\rho = \rho_{\text{data}}$ (blue) as detailed in Eq. XX. (b) Pearson correlation coefficient between the delay observed in the data and in the model. Asterisks indicate the level of significance (* p-value < 0.05, ** p-value < 0.01, *** p-value < 0.001). The injection rate for each city ρ_{data} is set to match η with the percentage of delay observed in the Tom Tom traffic index data [5].

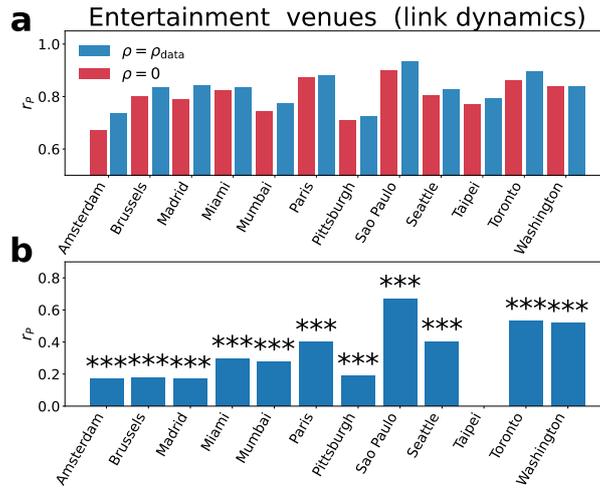


Figure S19: **Correlation between the real and modelled delays in the link dynamics when destinations are distributed according to the entertainment POIs.** (a) Comparison between the Pearson correlation coefficient obtained between the travel times from Uber Data [4] during the morning peak (8 – 10am) in a set of cities and the travel times obtained for $\rho = 0$ (red) and $\rho = \rho_{data}$ (blue). (b) Pearson correlation coefficient between the delay observed in the data and in the model. Asterisks indicate the level of significance (* p-value < 0.05, ** p-value < 0.01, *** p-value < 0.001). The injection rate for each city ρ_{data} is set to match η with the percentage of delay observed in the Tom Tom traffic index data [5].

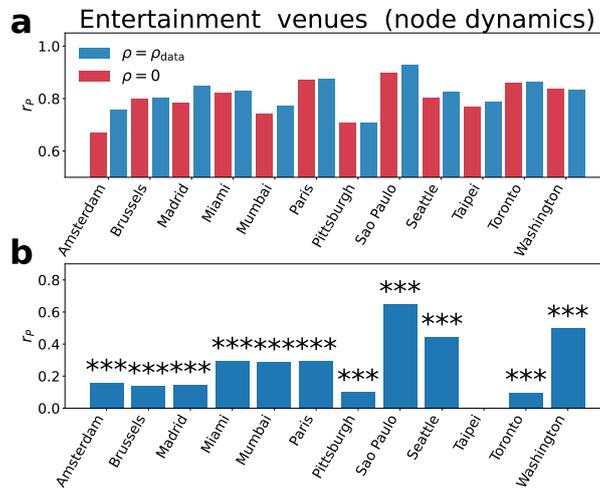


Figure S20: **Correlation between the real and modelled delays in the node dynamics when destinations are distributed according to the entertainment POIs.** (a) Comparison between the Pearson correlation coefficient obtained between the travel times from Uber Data [4] during the morning peak (8 – 10am) in a set of cities and the travel times obtained for $\rho = 0$ (red) and $\rho = \rho_{data}$ (blue). (b) Pearson correlation coefficient between the delay observed in the data and in the model. Asterisks indicate the level of significance (* p-value < 0.05, ** p-value < 0.01, *** p-value < 0.001). The injection rate for each city ρ_{data} is set to match η with the percentage of delay observed in the Tom Tom traffic index data [5].

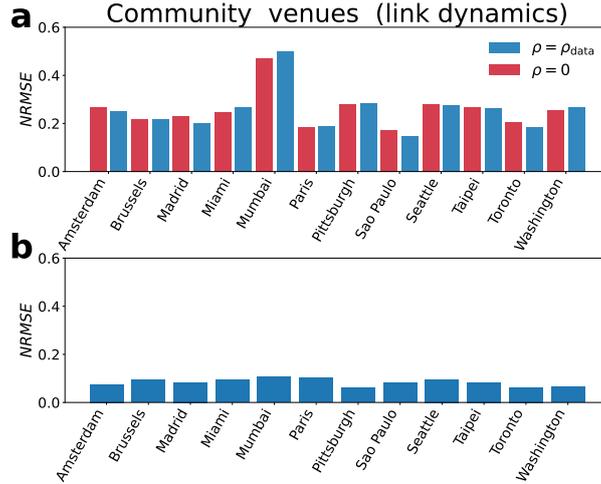


Figure S21: **Correlation between the real and modelled delays in the node dynamics when destinations are distributed according to the community POIs.** (a) Normalized root mean squared error (NRMSE) obtained by dividing the standard deviation of the residuals by the sample mean for the regression between the travel times from Uber Data [4] during the morning peak (8 – 10am) in a set of cities and the travel times obtained for $\rho = 0$ (red) and $\rho = \rho_{data}$ (blue). (b) Normalized root mean squared error (NRMSE) obtained by dividing the standard deviation of the residuals by the sample mean for the regression between the delay observed in the data and in the model. Asterisks indicate the level of significance (* p-value < 0.05, ** p-value < 0.01, *** p-value < 0.001. The injection rate for each city ρ_{data} is set to match η with the percentage of delay observed in the Tom Tom traffic index data [5].

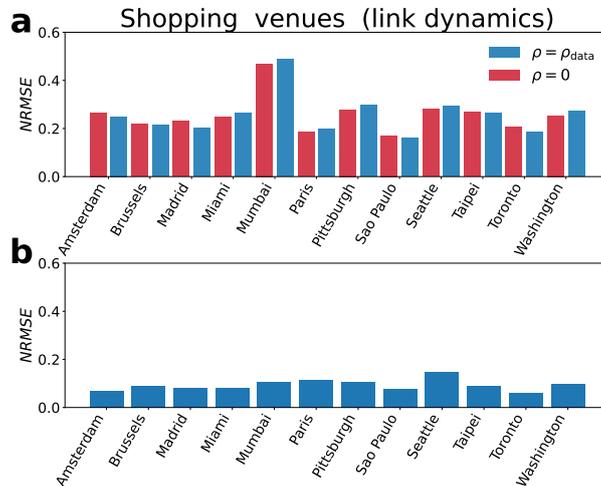


Figure S22: **Correlation between the real and modelled delays in the node dynamics when destinations are distributed according to the community POIs.** (a) Normalized root mean squared error (NRMSE) obtained by dividing the standard deviation of the residuals by the sample mean for the regression between the travel times from Uber Data [4] during the morning peak (8 – 10am) in a set of cities and the travel times obtained for $\rho = 0$ (red) and $\rho = \rho_{data}$ (blue). (b) Normalized root mean squared error (NRMSE) obtained by dividing the standard deviation of the residuals by the sample mean for the regression between the delay observed in the data and in the model. Asterisks indicate the level of significance (* p-value < 0.05, ** p-value < 0.01, *** p-value < 0.001. The injection rate for each city ρ_{data} is set to match η with the percentage of delay observed in the Tom Tom traffic index data [5].

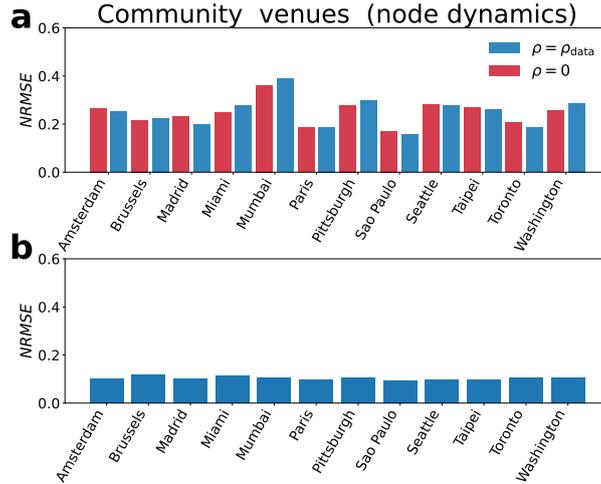


Figure S23: **Correlation between the real and modelled delays in the link dynamics when destinations are distributed according to the shopping POIs.** (a) Normalized root mean squared error (NRMSE) obtained by dividing the standard deviation of the residuals by the sample mean for the regression between the travel times from Uber Data [4] during the morning peak (8 – 10am) in a set of cities and the travel times obtained for $\rho = 0$ (red) and $\rho = \rho_{\text{data}}$ (blue) as detailed in Eq. XX. (b) Normalized root mean squared error (NRMSE) obtained by dividing the standard deviation of the residuals by the sample mean for the regression between the delay observed in the data and in the model. Asterisks indicate the level of significance (* p-value < 0.05, ** p-value < 0.01, *** p-value < 0.001). The injection rate for each city ρ_{data} is set to match η with the percentage of delay observed in the Tom Tom traffic index data [5].

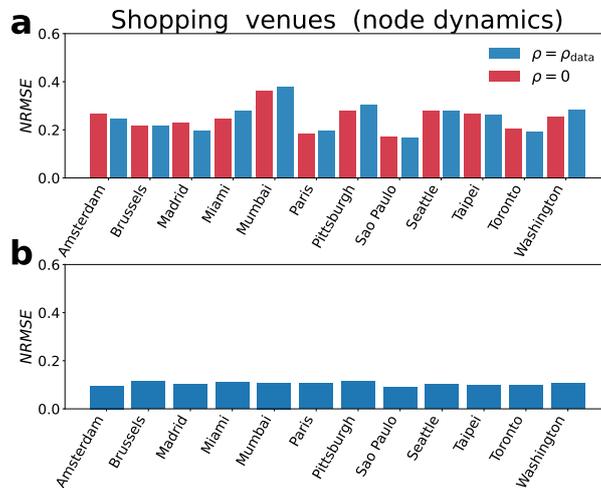


Figure S24: **Correlation between the real and modelled delays in the node dynamics when destinations are distributed according to the shopping POIs.** (a) Normalized root mean squared error (NRMSE) obtained by dividing the standard deviation of the residuals by the sample mean for the regression between the travel times from Uber Data [4] during the morning peak (8 – 10am) in a set of cities and the travel times obtained for $\rho = 0$ (red) and $\rho = \rho_{\text{data}}$ (blue). (b) Normalized root mean squared error (NRMSE) obtained by dividing the standard deviation of the residuals by the sample mean for the regression between the delay observed in the data and in the model. Asterisks indicate the level of significance (* p-value < 0.05, ** p-value < 0.01, *** p-value < 0.001). The injection rate for each city ρ_{data} is set to match η with the percentage of delay observed in the Tom Tom traffic index data [5].

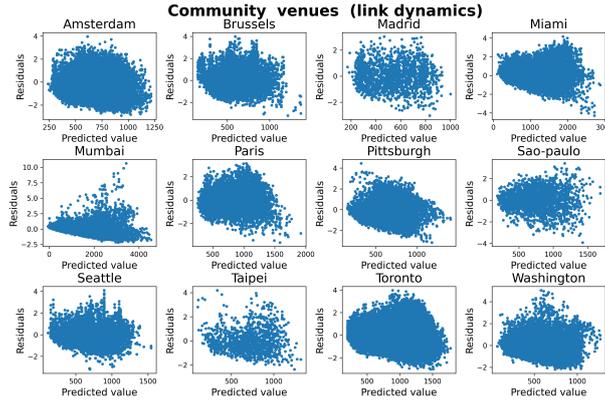


Figure S25: Residual analysis for the regression between the travel times in the link dynamics when destinations are distributed according to the community POIs. (a) Residual analysis for the regression between the travel times from Uber Data [4] during the morning peak (8 – 10am) in a set of cities and the travel times obtained for $\rho = \rho_{\text{data}}$.

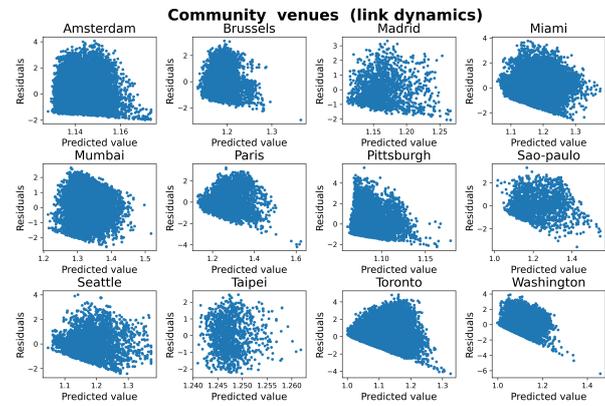


Figure S26: Residual analysis for the regression between the delays in the link dynamics when destinations are distributed according to the community POIs. (a) Residual analysis for the regression between the travel times from Uber Data [4] during the morning peak (8 – 10am) in a set of cities and the travel times obtained for $\rho = \rho_{\text{data}}$.

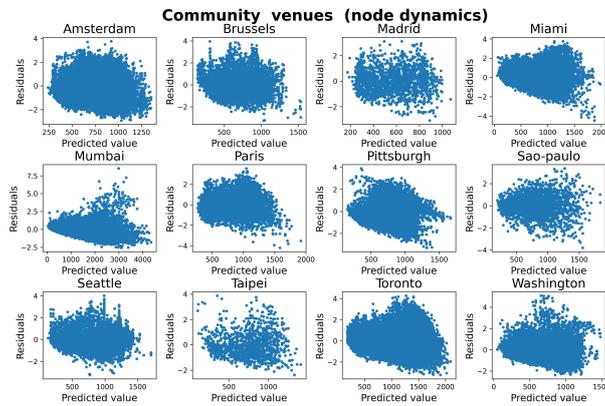


Figure S27: Residual analysis for the regression between the travel times in the node dynamics when destinations are distributed according to the community POIs. (a) Residual analysis for the regression between the travel times from Uber Data [4] during the morning peak (8 – 10am) in a set of cities and the travel times obtained for $\rho = \rho_{\text{data}}$.

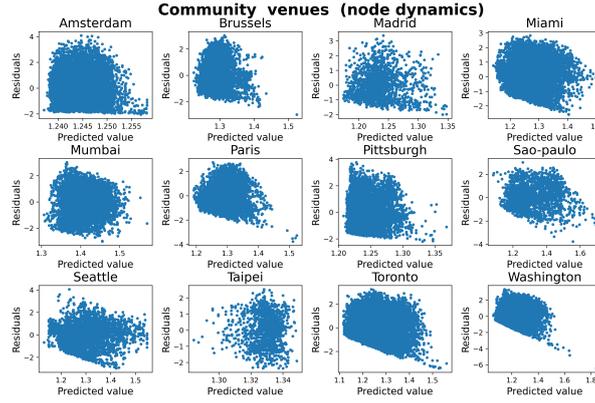


Figure S28: **Residual analysis for the regression between the delays in the node dynamics when destinations are distributed according to the community POIs.** (a) Residual analysis for the regression between the delay observed in the Uber Data [4] during the morning peak (8 – 10am) and in the model for $\rho = \rho_{\text{data}}$.

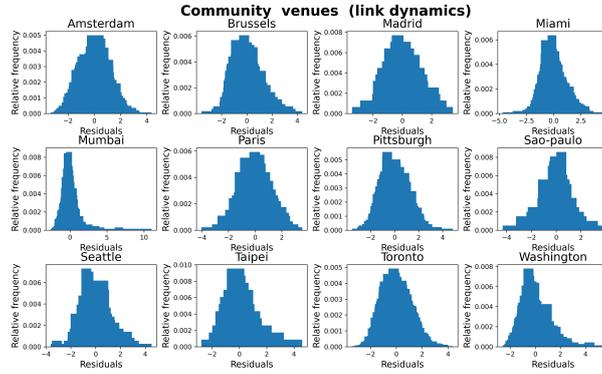


Figure S29: **Distribution of residuals for the regression between the travel times in the link dynamics when destinations are distributed according to the community POIs.** (a) Distribution of residuals for the regression between the travel times from Uber Data [4] during the morning peak (8 – 10am) in a set of cities and the travel times obtained for $\rho = \rho_{\text{data}}$.

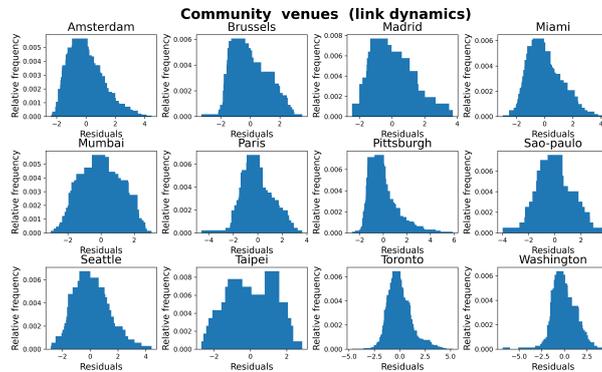


Figure S30: **Distribution of residuals for the regression between the delays in the link dynamics when destinations are distributed according to the community POIs.** ((a) Distribution of residuals for the regression between the delay observed in the Uber Data [4] during the morning peak (8 – 10am) and in the model for $\rho = \rho_{\text{data}}$.

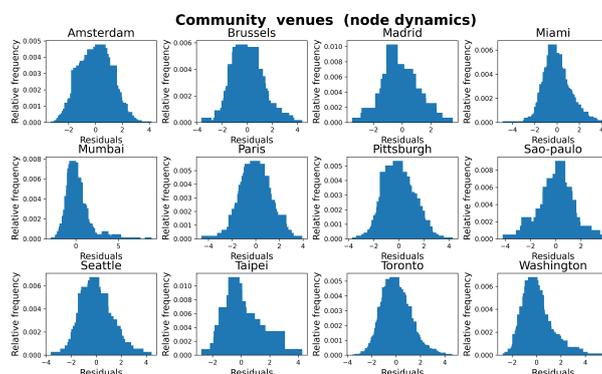


Figure S31: **Distribution of residuals for the regression between the travel times in the node dynamics when destinations are distributed according to the community POIs.** (a) Distribution of residuals for the regression between the travel times from Uber Data [4] during the morning peak (8 – 10am) in a set of cities and the travel times obtained for $\rho = \rho_{\text{data}}$.

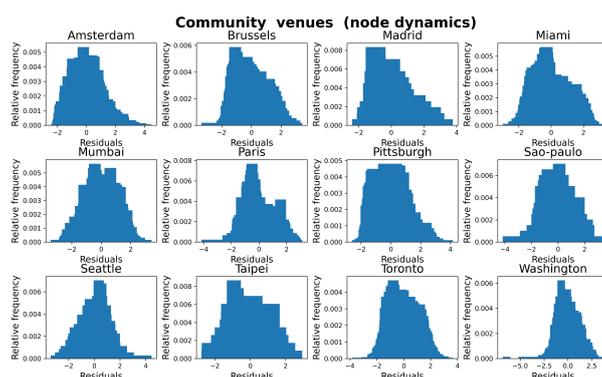


Figure S32: **Distribution of residuals for the regression between the delays in the node dynamics when destinations are distributed according to the community POIs.** (a) Distribution of residuals for the regression between the delay observed in the Uber Data [4] during the morning peak (8 – 10am) and in the model for $\rho = \rho_{\text{data}}$.

Supplementary References

- [1] Lampo, A., Borge-Holthoefer, J., Gómez, S., and Solé-Ribalta, A. Multiple abrupt phase transitions in urban transport congestion. *Physical Review Research* **3**(1), 013267 (2021).
- [2] Lampo, A., Borge-Holthoefer, J., Gómez, S., and Solé-Ribalta, A. Emergence of spatial transitions in urban congestion dynamics. *Applied Network Science* **6**(1), 1–16 (2021).
- [3] Lee, D.-T. and Schachter, B. J. Two algorithms for constructing a delaunay triangulation. *International Journal of Computer & Information Sciences* **9**(3), 219–242 (1980).
- [4] Uber Movement: Let's find smarter ways forward, together. <https://movement.uber.com/?lang=es-ES>, (2016). [Online; accessed 2021-10-01].
- [5] TomTom Traffic Index. "https://www.tomtom.com/en_gb/traffic-index/", (2021). [Online; accessed 2022-03-01].