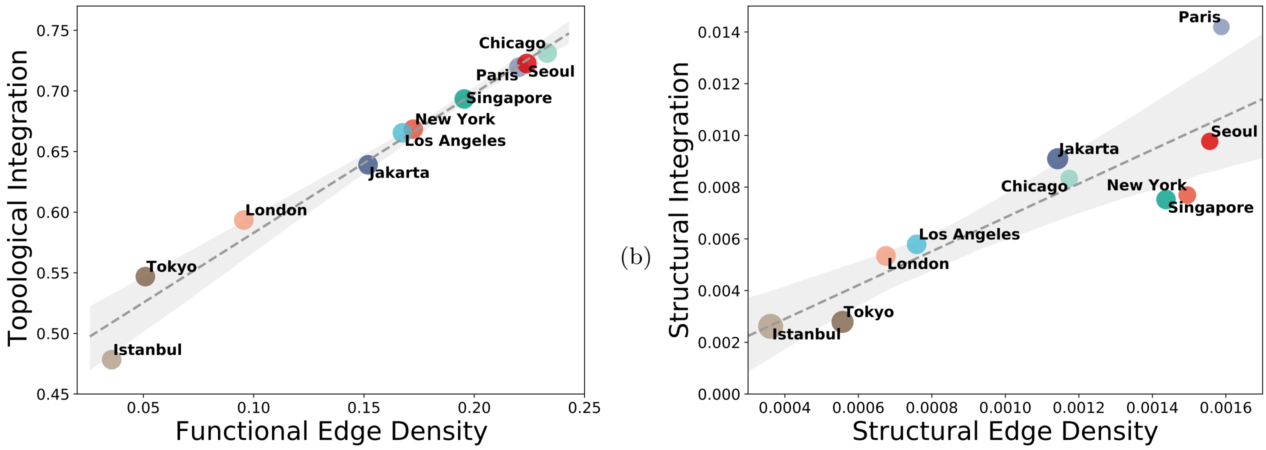
Supplementary Information for: Unraveling the hidden organisation of urban systems and their mobility flows

Riccardo Gallotti[1,*] Giulia Bertagnolli[1,2], and Manlio De Domenico[1†]

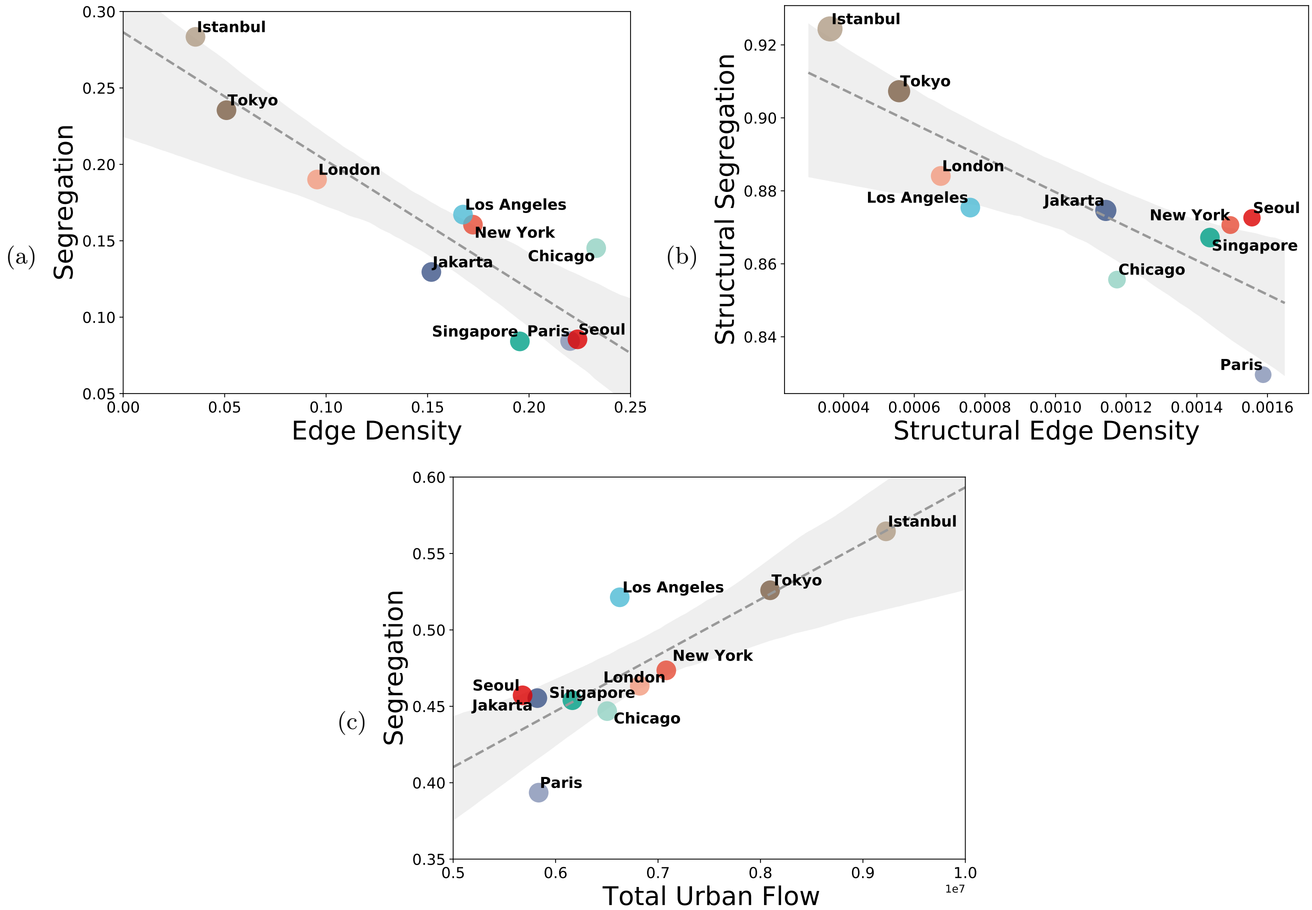
[1] *CoMuNe Lab, Fondazione Bruno Kessler, Via Sommarive 18, 38123 Povo (TN), Italy and*

[2] *Department of Mathematics, University of Trento, Via Sommarive 14, 38123 Povo (TN), Italy*

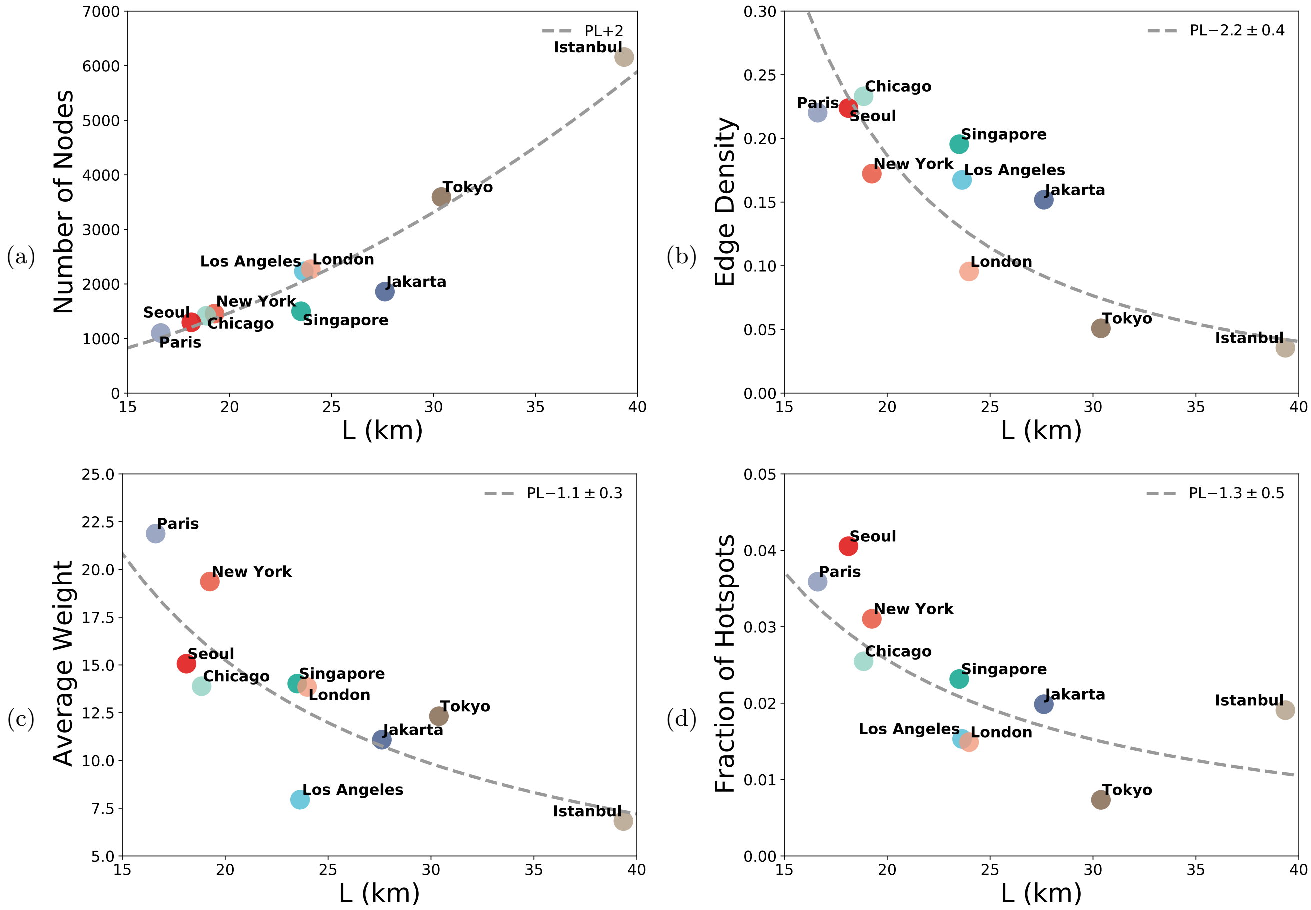
(a) (b)



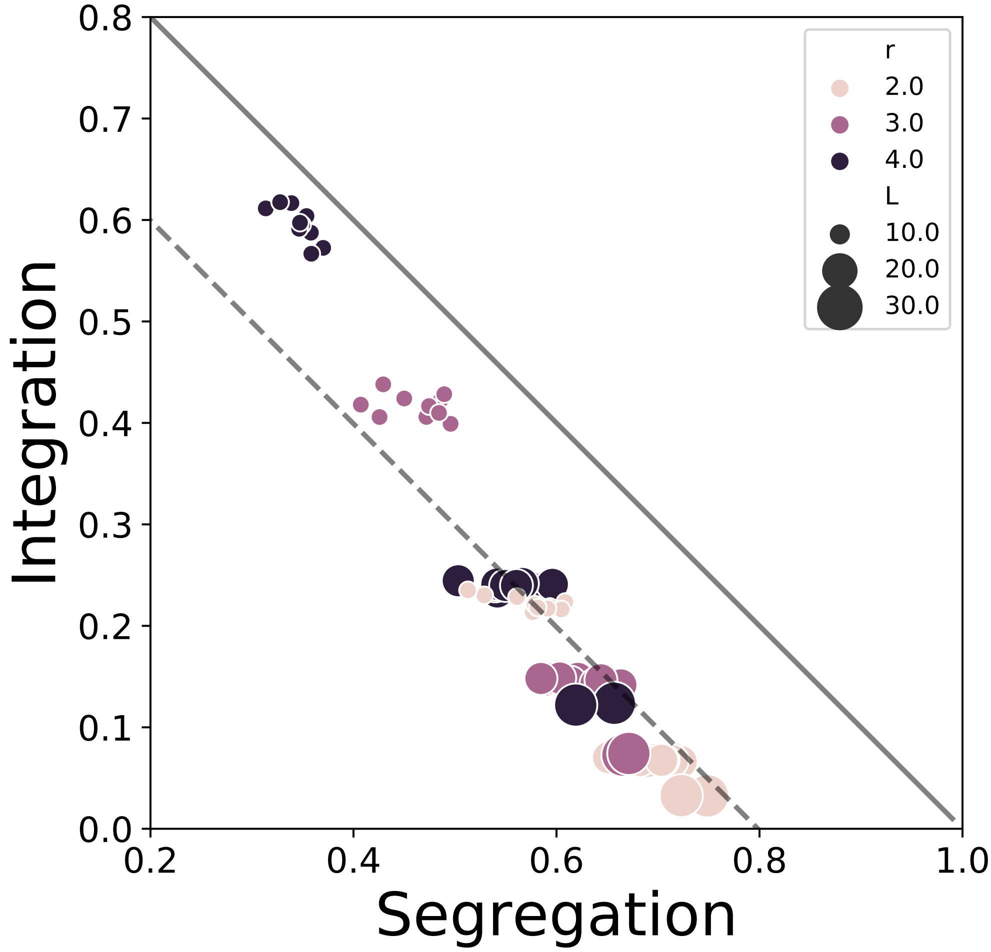
Supplementary Figure 1: **Dependance of integration of topological networks on edge density (a)** The value of integration for the topological functional networks is strictly linked to the edge density. As expected for topological networks, the larger the edge density the larger the integration. **(b)** The integration for the structural networks is also strongly related to edge density.



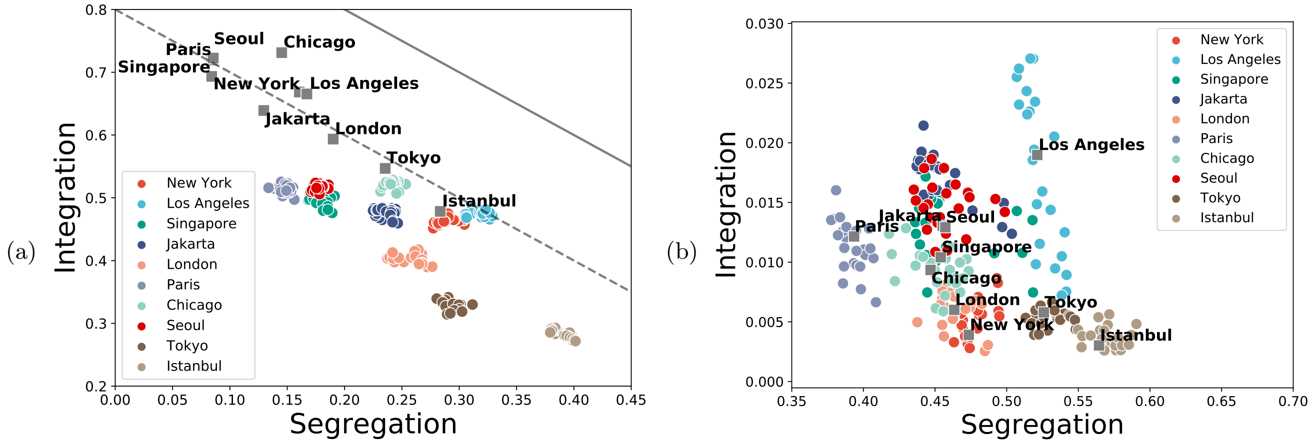
Supplementary Figure 2: **Understanding Structural and Functional Segregation.** Similarly to what illustrated in Fig. 4, here we link the values of segregation for the structural network derived from Open Street Maps and the functional (topological and weighted) networks derived from the Foursquare flows. In this case, the range of values observed for the three networks are not consistent between, suggesting that an improved and correctly normalised definition of segregation is still needed. **(a)** The value of segregation for the topological functional networks is anti-correlated to the edge density, but less tightly than what observed for Integration. **(b)** Similarly, segregation for the structural network grows as edge density increases. **(c)** The value of segregation for the weighted functional networks seems instead to be linked to the total flow recorded in the city, i.e. is the sum of all weights in the network.



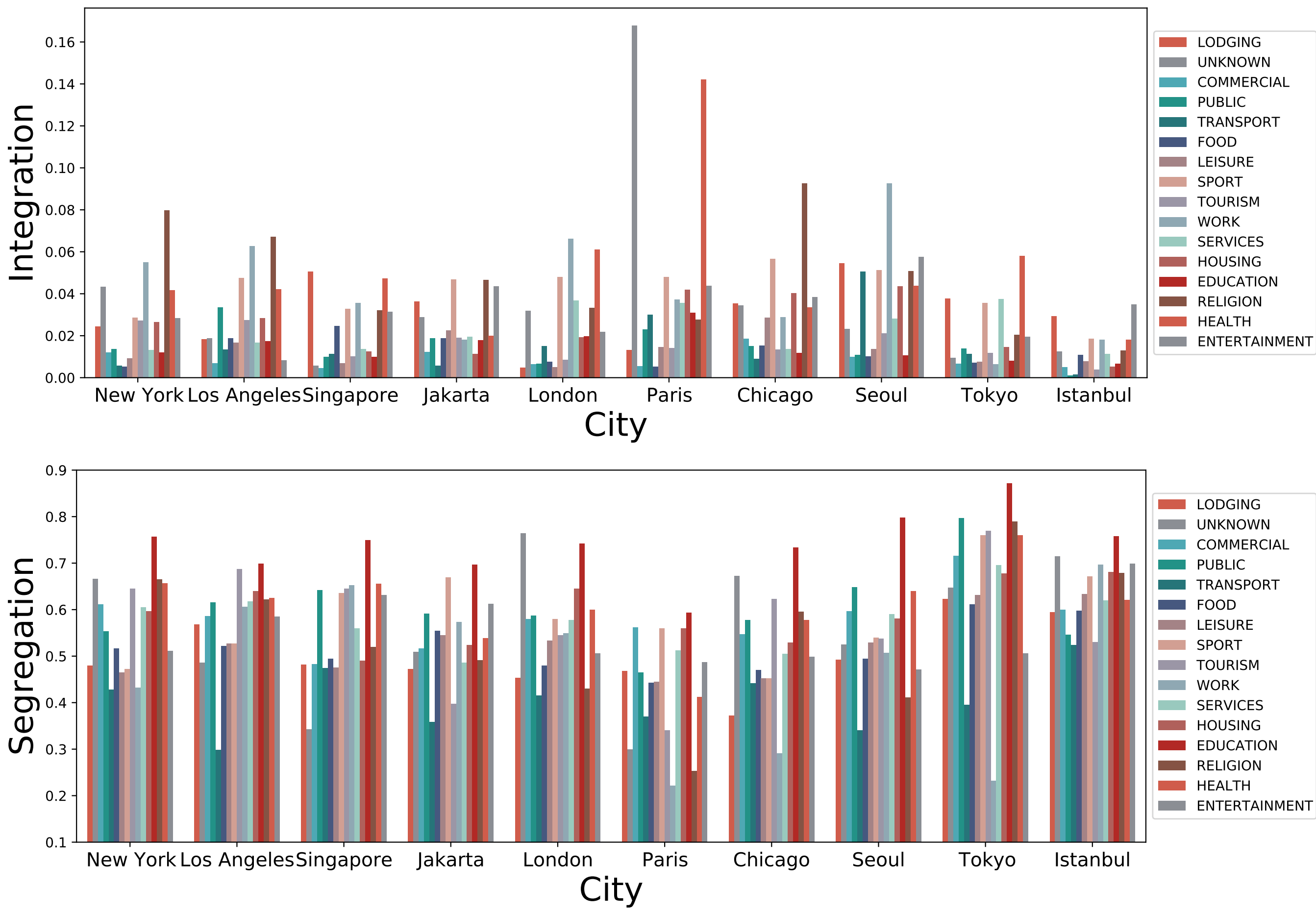
Supplementary Figure 3: **Connecting network properties with urban scaling.** Here we show how the network indicators we extracted from the Foursquare data depend upon city dimensions in terms of L, that is computed as the square root of the total surface area included in the data provided. As we observed in Supplementary Fig.2, functional segregation appears to be proportionate to the total weight of a city. The total weight can be decomposed to the product of three factors: $W_{tot} = N^2 \cdot ed \cdot \langle w \rangle$, where N is the number of nodes, ed the edge density and $\langle w \rangle$ the average weight. In the first three panels we illustrate the scaling behaviour for these three quantities. **(a)** Since we have built the network by coarse graining on a regular grid, it is natural that number of nodes is naturally proportionate to the square of L, i.e. the surface area. **(b)** The edge density decreases for larger cities, which leads to higher values of topological segregation and integration as the city grows. **(c)** Also the average weight of links decreases for larger cities, a factor contributing to a smaller values of segregation as the city grows. **(d)** A final insight on the scaling properties of cities can be extracted by observing that as the size L of the city grow, the fraction of area that is represented by hotspots obtained with the LouBar method (Louail et al 2014) decreases. All dashed lines represent the best fit for a power-law scaling. Given the limited number of points and decades the values have to be considered only as a rough indication which we include in this figure as we hope might inform further studies.



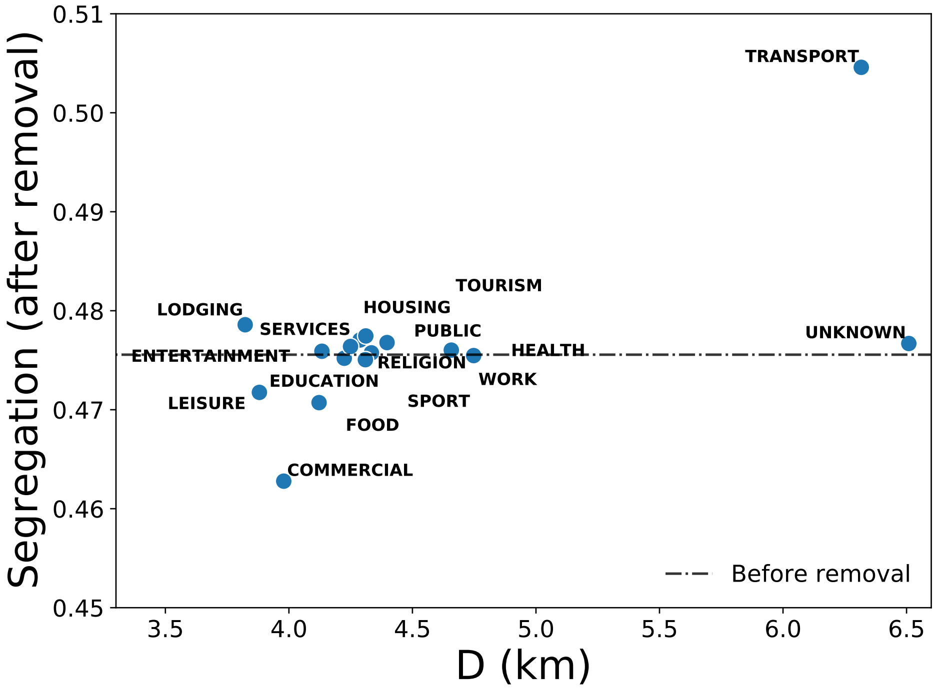
Supplementary Figure 4: **Segregation and Integration of Random Geometric Networks of different sizes.** In this paper, we generate RGNs by i) throwing N nodes in random locations in a square of edge L; ii) connecting all node pairs (i, j) with distance $d(i, j) < r$; iii) rewiring a fraction α of edges. Here, to study the effect of size, we generate networks with identical node density N/L^2 and with no rewiring $\alpha = 0$. For each value of L and r we averaged the values of segregation (modularity Q and integration GCE). The result show that, in this scenario, segregation and integration are strongly anti-correlated. High integration is attained for small networks ($L = 10$) with large r, while the opposite yields high segregation.



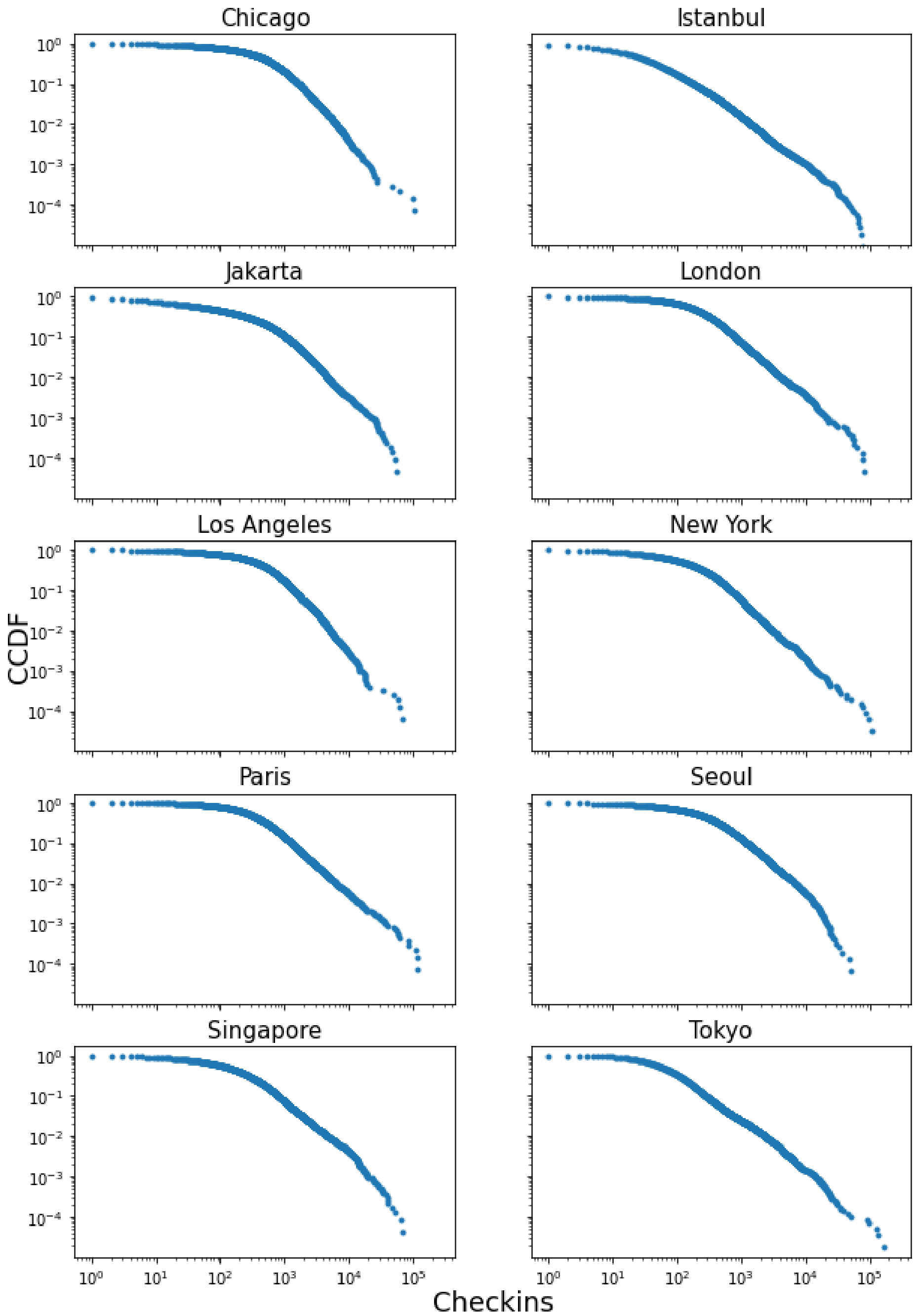
Supplementary Figure 5: **Integration and segregation for topological network and disaggregated by month.** Flows are stratified according to different months (multiple points), while the grey square letter of a city name falls in correspondence of the values for the whole dataset. **(a)** Topological network. Remarkably, the values for topological network extracted by single months exhibit a large deviation from those aggregated over the whole period of analysis. The values for the monthly sub-samples range correspond to those of random geometric networks, suggesting that monthly data would be too under-sampled for making an analysis based only on the topological features of the networks. **(b)** Weighted network. In this case, the richer information captured by nodes allow to compare values for a single month (coloured dots) to those aggregated over the whole period considered (grey squares).



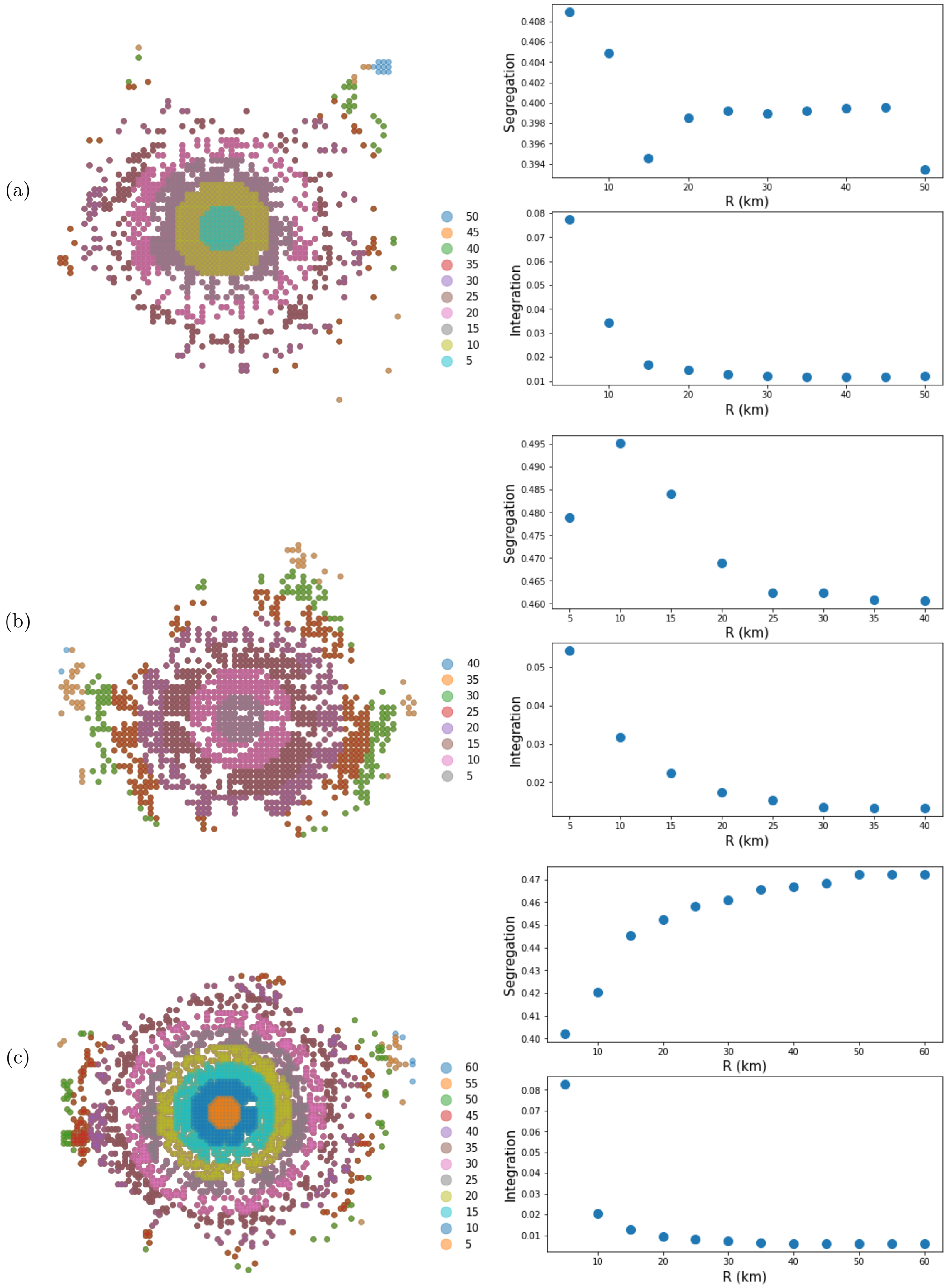
Supplementary Figure 6: **Segregation and Integration for the Single Layer Functional Weighted Networks.**



Supplementary Figure 7: **Average functional integration for different activity categories.** Conversely from what observed in Fig 5c for integration, we observe no clear dependency of the effect of removing a layer with the average distance covered D in movement inside that layer. Again, the transport layer is displaying exceptional behaviour.



Supplementary Figure 8: **The Complementary Cumulative Distribution Function of the number of check-ins per venue.** The fat tail displayed by these curves illustrate the extremely high level of inhomogeneity of the flows captured by different venues, with a small fraction of venues capturing a significant amount of flow.



Supplementary Figure 9: **Test of the measures of Segregation and Integration for the definition of city.** In our dataset, the definition of city boundaries was already provided by Foursquare. This definition of city is possibly non-homogeneous. In this figure, we test the robustness of our metrics to the city definition by radially reducing the boundaries of the network. We took three cities that display a clear central structure **(a)** Paris, **(b)** Seoul and **(c)** London, and divided them into concentric circles of radius ranging as $[5, 10, 15, \ldots, R_{max}]$ around the center of mass of the network. The results of this analysis support that the details of the peripheral boundaries are minor, as in all three cities we observe relevant deviations from the registered values only when the city is reduced in its core for $R < 15$.

* Electronic address: rgallotti@gmail.com
† Electronic address: mdedomenico@fbk.eu

	COM	EDU	ENTERT	FOOD	HEALTH	HOUSING	LEISURE	LODGING	PUBLIC	RELIGION	SERVICES	SPORT	TOURISM	TRANSP	UNKNOWN	WORK
Chicago - Nodes	622	227	176	776	249	158	509	62	370	154	252	224	138	470	176	115
Chicago - IntraLinks	74396	4336	1825	104144	8453	3261	40493	1132	17157	805	4192	4874	1514	31454	840	829
Chicago - InterLinks	158363	54945	33214	181829	69872	40874	114713	19404	98178	26412	55520	51816	29702	120992	29159	22858
Istanbul - Nodes	2483	2184	1336	2796	1042	3272	2714	538	2875	1363	1883	1809	640	1813	1536	2091
Istanbul - IntraLinks	78645	15690	1998	90612	9000	58776	123149	2614	67831	1764	9794	8228	1855	14606	1644	9174
Istanbul - InterLinks	296348	155504	65772	341340	97419	297854	350548	50668	275042	55166	133471	127610	49210	151433	61149	116357
Jakarta - Nodes	839	511	271	1026	315	776	653	209	466	458	615	340	188	574	228	251
Jakarta - IntraLinks	62841	4353	2078	103385	4687	25174	16515	6898	2675	4840	11589	1600	2108	25157	1693	934
Jakarta - InterLinks	164131	64837	39696	197228	56937	128007	102215	58464	50829	65379	96194	36842	36288	118820	38492	30763
London - Nodes	795	213	268	868	153	179	1167	346	665	92	224	376	231	808	268	183
London - IntraLinks	51941	1184	2169	46420	836	567	77615	7354	18515	134	1841	3425	3023	78711	1441	589
London - InterLinks	140579	23752	37800	134904	21187	19280	156034	49889	98442	8687	31076	51692	34995	153718	32502	20336
Los Angeles - Nodes	923	274	244	1112	247	238	634	156	489	148	428	256	173	552	203	156
Los Angeles - IntraLinks	123702	4792	2717	167685	5441	2909	44297	2784	17251	742	8502	4002	2073	25681	2923	870
Los Angeles - InterLinks	266196	69687	53272	300532	71091	56118	180268	44578	129269	28591	99141	64170	41790	162068	57270	31488
New York - Nodes	718	328	215	807	354	317	497	151	532	220	399	338	230	639	268	163
New York - IntraLinks	56306	2937	2006	65811	6211	3526	23904	2405	14118	533	5769	6639	2917	49163	828	1733
New York - InterLinks	124744	41524	27633	131971	57275	43014	77929	23250	77333	18168	53780	50393	34004	111510	24271	21664
Paris - Nodes	391	153	204	540	77	97	355	376	376	69	296	188	160	482	56	113
Paris - IntraLinks	28410	1093	2912	53640	566	320	26223	12231	23093	469	3292	2068	5069	45858	43	425
Paris - InterLinks	82901	20600	34153	96499	13746	12296	72918	64310	77559	10996	42673	32157	36593	96583	5981	14802
Seoul - Nodes	407	229	178	691	120	224	562	125	327	181	167	94	198	386	193	84
Seoul - IntraLinks	28462	2893	3806	75432	1244	4699	45495	2815	6783	1726	1401	656	5323	56111	3603	481
Seoul - InterLinks	104220	40617	42926	143171	24988	51111	123281	33727	60920	32688	30248	19394	50886	118972	41796	16034
Singapore - Nodes	579	441	163	755	168	818	592	118	474	263	481	266	199	755	172	189
Singapore - IntraLinks	38865	5449	1225	99038	904	24290	24209	3417	6037	2329	6555	1604	2124	53865	581	1459
Singapore - InterLinks	121819	60969	27696	164809	20995	131287	112799	33785	62673	39445	68871	33955	32553	134257	19886	29316
Tokyo - Nodes	2203	518	416	2074	199	333	1073	253	920	452	537	440	783	1114	928	225
Tokyo - IntraLinks	95431	2185	5927	92359	320	632	27966	1112	6054	1462	2042	1036	6350	136308	15568	443
Tokyo - InterLinks	204810	36432	47298	200801	12475	21226	121862	21007	68296	23349	39153	28253	60817	194343	86904	18849

Supplementary Table I: **Number of nodes and links (intra- and inter- layer) for each layer of the 10 cities considered in our analysis.**

city	Q	N	Z	GCE	E
Chicago	0.4469	7.0	48.8229	0.0093	103.8097
Istanbul	0.5645	8.0	75.3107	0.003	30.0698
Jakarta	0.4554	8.0	144.0917	0.013	65.0682
London	0.4634	9.0	-2.8755	0.006	65.2905
Los Angeles	0.5213	7.0	290.1588	0.019	67.9652
New York	0.4735	7.0	0.6814	0.0039	85.1583
Paris	0.3935	8.0	-21.7259	0.0122	124.7085
Seoul	0.4572	9.0	88.0736	0.0129	119.7564
Singapore	0.454	10.0	39.3882	0.0104	107.9742
Tokyo	0.5259	11.0	3.2594	0.0058	63.8658

Supplementary Table II: **Table of the values of Segregation (Q) and Integration (GCE) for the weighted functional networks.** We also provide the number of communities N identified, the Z-score of the segregation against a configuration model and the non-normalised efficiency E. We report that segregation is highly significant with the notable exception of New York city.

city	Q	N	Z	GCE	E
Chicago	0.1452	3.0	181.7603	0.7312	0.7322
Istanbul	0.2834	3.0	477.7494	0.4783	0.5121
Jakarta	0.1296	3.0	147.725	0.6391	0.6472
London	0.1901	4.0	212.8296	0.5935	0.5946
Los Angeles	0.1671	3.0	246.2364	0.6655	0.6666
New York	0.1605	3.0	155.8727	0.6685	0.6707
Paris	0.0844	4.0	54.1208	0.7195	0.7196
Seoul	0.0855	3.0	75.2966	0.7226	0.7231
Singapore	0.0841	3.0	83.2197	0.6934	0.6944
Tokyo	0.2354	4.0	297.3499	0.5469	0.5492

Supplementary Table III: **Table of the values of Segregation (Q) and Integration (GCE) for the topological functional networks.** We also provide the number of communities N identified, the Z-score of the segregation against a configuration model and the non-normalised efficiency E. All values of segregation are highly significant. The unnormalised efficiency E is here the same as the GCE by definition.