

Structural invariants in street networks: modeling and practical implications

Alec Kirkley, Hugo Barbosa, Marc Barthelemy, Gourab Ghoshal

▶ **To cite this version:**

Alec Kirkley, Hugo Barbosa, Marc Barthelemy, Gourab Ghoshal. Structural invariants in street networks: modeling and practical implications. 2017. cea-01626272

HAL Id: cea-01626272

https://cea.hal.science/cea-01626272

Preprint submitted on 30 Oct 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

arXiv:1709.05718v1 [physics.soc-ph] 17 Sep 2017

Structural invariants in street networks: modeling and practical implications

Alec Kirkley[1], Hugo Barbosa[1], Marc Barthelemy[2,3], and Gourab Ghoshal[1,4]

[1]Department of Physics & Astronomy, University of Rochester, Rochester, New York
[2]Institut de Physique Théorique, Gif-sur-Yvette, France
[3]Centre d'Analyse et de Mathématique Sociales, Paris Cedex 13, France
[4]Goergen Institute for Data Science, University of Rochester, Rochester, New York

Contents

Main Manuscript **2**

Supporting Information **30**

Structural invariants in street networks: modeling and practical implications

Alec Kirkley,[1] Hugo Barbosa,[1] Marc Barthelemy,[2,3] and Gourab Ghoshal[1,4]

[1]*Department of Physics & Astronomy, University of Rochester, Rochester, New York*

[2]*Institut de Physique Théorique, CEA, Gif-sur-Yvette, France*

[3]*Centre d'Analyse et de Mathématique Sociales, EHESS, Paris Cedex 6, France*

[4]*Georgen Institute for Data Science, University of Rochester, Rochester, New York*

Abstract

We study structural properties of street networks from 97 of the most populous cities worldwide at scales significantly larger than previous studies. We find that the distribution of betweenness centrality (BC), a global structural metric based on network flow, is invariant in all studied street networks, despite the obvious structural differences between them. We also find that the BC distribution is robust to major alterations in the network, including significant changes to its topology and edge weight structure, indicating that the only relevant factors shaping the distribution are the number of nodes in a network, the number of edges, and the constraint of planarity. Through a combination of simulations of random planar graph models and analytical calculations on Cayley trees, this remarkable invariance is demonstrated to be a consequence of a bimodal regime consisting of an underlying tree structure for high betweenness nodes, and a low betweenness regime arising from the presence of loops providing local path alternatives. Furthermore, the high betweenness nodes display a non-trivial spatial dependence, with increasing spatial correlation as a function of the number of roads, leading them to cluster around the barycenter for cities with high density of streets. As the BC is a static predictor of traffic flow, this invariance has important implications for urban planning; indeed, as long as planarity is conserved, bottlenecks will persist and the effect of planned interventions to alleviate congestion will be limited primarily to load redistribution, a feature confirmed by analyzing 200 years of data for central Paris.

Introduction

Recent years have witnessed unprecedented progress in our understanding of urban systems [1–5]. In particular, the study of structural properties of urban infrastructure networks has uncovered unique characteristics of individual cities as well as demonstrated surprising statistical commonalities manifested as scale invariant patterns across different urban contexts [6–9]. Networks of streets and roads are particularly important, allowing residents to navigate the different functional components of a city. Clearly a multitude of factors go into the physical layout of streets, including geographic constraints, design choices, and land parceling among many others. Different street structures result in varying levels of efficiency, accessibility, and usage of transportation infrastructure [10–16]. Consequently structural characteristics of roads have been of great interest in the literature, including the degrees of street junctions, lengths of road segments, cell areas or shapes delineated by streets, and anisotropies among others [17–23].

Street networks fall into the category of planar graphs [24], and their edges constitute a *physical* connection, as opposed to relational connections found in many complex networks [25]. In particular, the geographical embedding (or spatial constraint) leads to strong effects on network topology with limitations on the number of long-range connections and the number of edges incident on a single node (its degree k) [26, 27]. Consequently, degree-based network measures, while well-studied on such systems, lead to rather trivial results; for example the degree distribution is strongly peaked and related metrics such as clustering and assortativity are high [6]. Instead, more interesting information can be gleaned from non-local higher-level metrics such as those based on network centralities, which while strongly correlated with degree in non-spatial networks [28], display highly non-trivial behavior in planar networks [29]. Among the more studied and illuminating metrics of such measures is the betweenness centrality (BC), a path-based measure of the importance of a node in terms of the amount of flow passing through it [30]. More precisely, the BC for node i is defined as

$$g_B(i) = \frac{1}{\mathcal{N}} \sum_{s \neq t \in V} \frac{\sigma_{st}(i)}{\sigma_{st}}, \tag{1}$$

where σ_{st} is the number of shortest paths going from nodes s to t and $\sigma_{st}(i)$ is the number of these paths that go through i [30] . Here $\mathcal{N}$ is a normalization constant, typically of order N^2 where N is the number of nodes, although for reasons that will be apparent later

in the manuscript, we will use here the unnormalized version ($\mathcal{N} = 1$). In principle, one can define a variety of different shortest paths ranging from the number of hops (in the purely topological case) to the shortest distance between two points (geodesic) if the edges are weighted according to Euclidean distances. In the case of urban street networks, given purely structural data, and assuming roughly uniform speed limits and demand, the BC can be used as a proxy for predicted traffic density [31–33]. In such a setting the geodesic paths can be considered as the optimal routes between locations, and thus nodes with high BC should expect to receive more traffic. Consequently, in what is to follow, we will focus on the *weighted* node BC, calculated using the standard Brandes algorithm [34].

A number of studies have been conducted on the BC in various samples of street networks [35–37] finding among other things, a complicated spatial behavior of the high BC nodes [18, 38] as well as its connection to the organization and evolution of cities [39**?** –41]. For non-planar graphs the average BC scales with the degree k in a power law fashion thus $g_B(k) = \sum_{i|k_i=k} \frac{g_B(i)}{N(k)} \propto k^{\eta}$, where k is the degree, $N(k)$ is the number of nodes of degree k, and η is an exponent depending on the graph [42]. In planar graphs, however, the BC behaves in a much more complex manner, as now both topological *and* spatial effects are at play. While for a regular lattice, the BC is a function of the distance from the barycenter, with increasing disorder (as found in street networks) the BC will in general be a complex function of both the distance from the barycenter and of the local topology [38].

Existing analyses of street networks usually suffer from limitations of scale (see [43] for a recent global description), and most comparative studies across cities are typically restricted to one-square-mile samples, while studies on more extensive (but low-resolution) street-maps have been examined for at most tens of cities limited to those in Europe or North America. Furthermore, there have been limited studies of the BC distribution in its entirety, with the majority of analyses instead focusing on the average BC (proportional to the average shortest path [44]) or on its maximum value [45, 46].

To fill this gap, we conduct here a large-scale empirical study of the BC across 97 of the world's largest cities as measured by population (see details about the dataset in the Material and Methods). These cities are sampled from all six inhabited continents and the analysis is conducted at an unprecedented scale of the order of 2×10^3 square-kilometers. We group the different cities in three main categories according to their size (Fig. 1), from small ($N \sim 10^3$ nodes), medium ($N \sim 10^4$) to large road networks ($N \sim 10^5$).

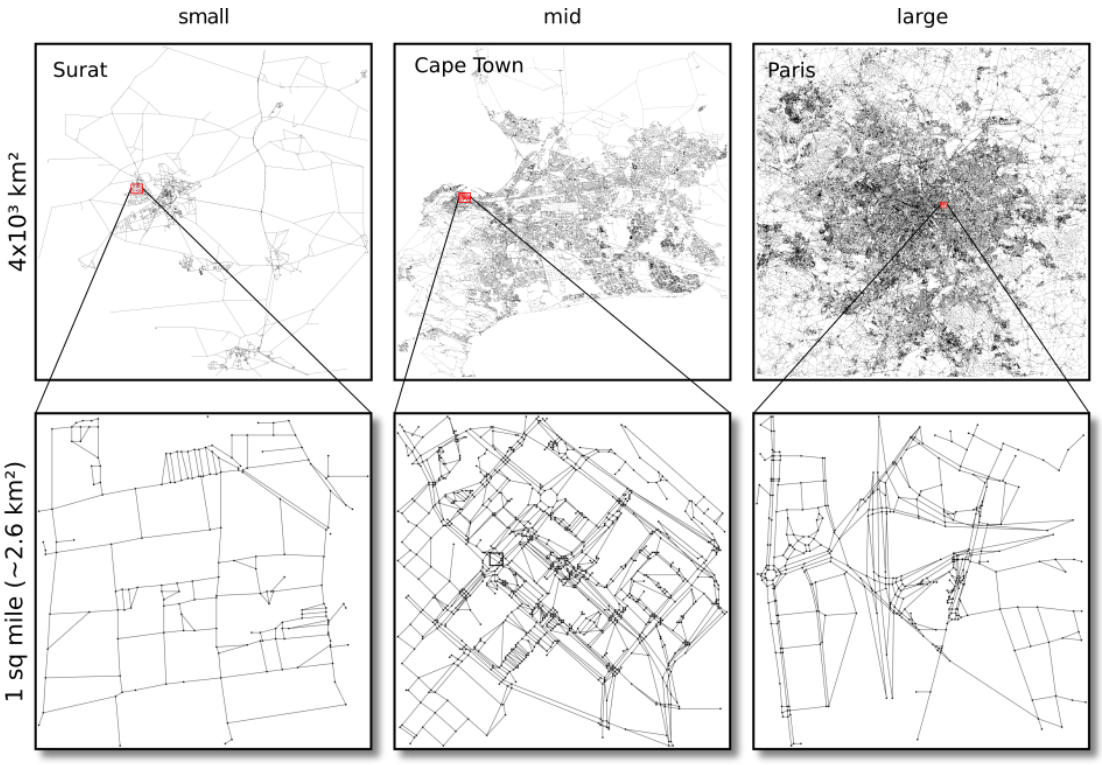


FIG. 1. **Street networks at multiple scales.** Cities split into three categories based on the number of nodes (intersections) in the sampled street networks: small ($N \sim 10^3$), medium ($N \sim 10^4$) and large ($N \sim 10^5$). The upper panel shows the networks at the full sampled range of 10^3 square-kilometers, whereas the lower panel shows selected smaller samples (on the order of one-square-mile).

Results

Betweenness at different scales and rescaling

In Fig. S1A we show the betweenness probability distribution for a selection of the three categories of cities at the resolution of two and a half square kilometers (or one-square-mile), plotted in a log-linear scale. One sees significant variability between cities, within and across categories, with mostly exponential tails (Fig. S2), as also seen for similar samples in [39, 40]. This is somewhat expected given the small sample size, as even controlling for the number of street intersections, the topology of cities are different due to geographic and spatial constraints [47, 48]. Indeed, these variations may show up *within the same city* where multiple samples of a similar resolution within a city display important fluctuations (Fig. S1B). At all scales, we observe a range of behavior in the tails of the BC ranging from

peaked to broad distributions, reflecting local variation in the street network structure and fluctuations in the data. One begins to see a dramatic difference when increasing the scale to two thousand-square-kilometers where the distribution in all cities start to look similar (Fig. S1C,D). We observe that the BC distribution for cities within each category is virtually *identical*, and also that the distribution is *bimodal*, with two regimes separated by a bump roughly at $g_B \sim N$. For larger values of the BC we observe a slow decay signalling a broad distribution. The combination of these features appear to have been overlooked or missed in existing work either due to the low resolution of the sampled street networks, or excess noise due to linear binning on a logarithmic x-axis [49].

These trends are apparent in the BC distribution across *all* 97 cities in our data as seen in Fig. 2A with the two regimes being separated by bumps spread across an interval of $10^3 \leq g_B \leq 10^5$ corresponding to the range of N in our data. Indeed rescaling the betweenness of each node by the number of vertices in the network $g_B \to \tilde{g_B} = g_B/N$, we see the distributions collapse on a single curve with a unique bump separating two clear regimes as seen in Fig.2B, although some variability exists resulting from differences in the number of edges. Remarkably, fitting [50] the distribution of $\tilde{g}_B = g_B/N$ with the function

$$p(\tilde{g}_B) \sim \tilde{g}_B^{-\alpha} e^{-\tilde{g}_B/\beta}, \tag{2}$$

results in a tightly bound range for $\alpha \approx 1$ and a broad size-dependent distribution for β (Fig. S4). Rescaling the tail with respect to β results in a collapse of the curves for all cities as seen in Fig. 2C. (Details for each city in Fig. S3 and Tab. S2). In the following we will provide theoretical explanations for these two facts.

Determinants of the BC distribution

Given the fact that cities in our data are ostensibly quite different in terms of geography or space, as well as their levels of infrastructure and socioeconomic development, the observed remarkable invariance is quite striking. To investigate the factors behind this behavior, we next systematically probe the effect of the main features that may be influencing the BC distribution. Examining Eq. (1), apart from its obvious dependence on the number on nodes N and the number of edges e, the other primary factors are the **(a)** Local topology—the local connectivity patterns of a street intersection as governed by its degree distribution **(b)** Distribution of edge weights that can correspond either to Euclidean distances or some scalar quantity such as speed-limits, and **(c)** Planarity—the effect of space. To do so we

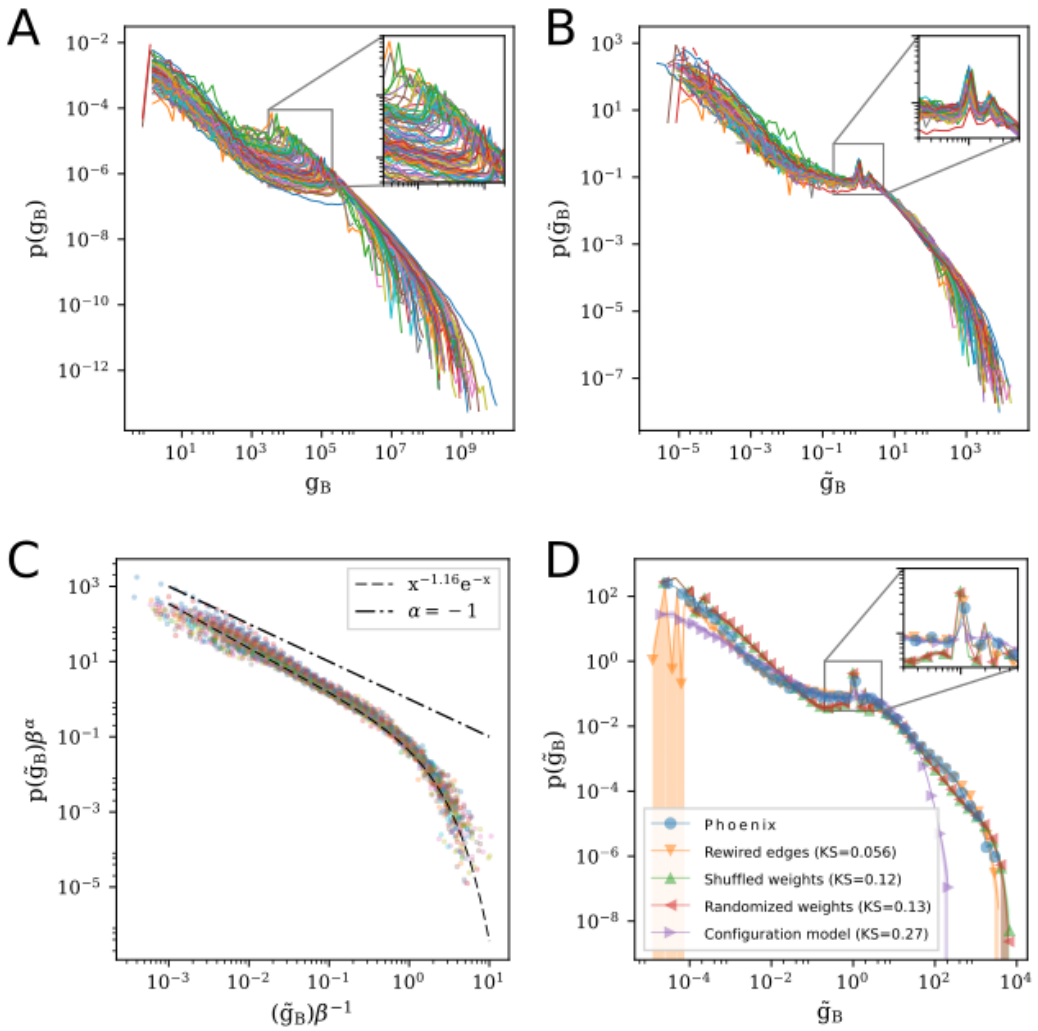


FIG. 2. **Betweenness invariance in urban streets. (A)** The betweenness pdf for all 97 cities at full-resolution. The peak of the distribution for each city is shown as inset. **(B)** The version of the distribution after rescaling by the number of nodes N showing the alignment of the peaks across all cities (also shown as inset). **(C)** The collapse of the tails after rescaling with respect to β. The dashed line shows the analytically computed asymptotic scaling for a Cayley-tree (Eq. (7)). **(D)** The BC distribution of various random graph models described in the text compared to the baseline distribution of Phoenix as a representative example. Shaded area reflects fluctuations around the average over hundred realizations of each model. Apart from the (non-spatial) configuration model, we see minimal changes in the location of the peak(s) (zoomed in inset) or shape of the tail. Also shown are 2-sample KS statistics for phoenix and its corresponding random graph models (details in Figs. S6 and S7).

select the BC distribution of a number of cities as baseline and generate multiple variants of random graphs to compare with the original. In Fig. 2D we show Phoenix (blue circles) as a representative example of a city on which we perform this analysis.

Effect of local topology: In principle, the BC of a given node is rather sensitive to local changes in topology. Consider, for example, the case of a "bridge node" that connects two disjoint clusters via connections to a single node in each cluster. Such a node has a high BC as it necessarily lies between *all* shortest paths between the two clusters. Yet, simply by placing an edge directly between the clusters, one can dramatically decrease the BC of the bridge node. To investigate such effects—which amounts to varying the local neighborhood of a given street intersection—we fix the spatial position of nodes on the 2D plane and generate a Delaunay Triangulation (DT) [51] of the street network. The DT corresponds to the maximum number of edges that can be laid down between a fixed number of nodes distributed within a fixed space, without any edge-crossings. Edges are then randomly eliminated until their number corresponds exactly to our baseline example of Phoenix. A hundred realizations of this procedure was conducted, having the effect of rewiring the local neighborhood of intersections—by changing a node's degree and its neighbors—while still maintaining planarity. In Fig. 2D we plot the average of these realizations (orange triangles), showing differences with the original street network in the lower range of the distribution, yet showing minimal change in both the location of the peak as well as the tail of the distribution. Similar random graphs were generated using a number of other cities as baseline showing the same behavior (Fig. S5).

Effect of edge weights: Next we investigate the effect of Euclidean distances, or the edge-weights on the BC distribution. We fix the number of nodes N and instead of fixing their positions according to the empirical pattern, we now distribute them *uniformly in the 2D plane* with a scale determined by the spatial extent of the city considered (which is Phoenix in our example). Having done this, we generate the DT of the street network and randomly remove edges until we match the number of roads in the data. A hundred different realizations of this procedure has the effect of stretching/compressing the city in multiple directions (either dispersing high density areas or compressing very long road segments) and therefore generating a distribution of distances that are markedly different from the original (Fig. S8). Fig. 2D (red triangles) suggests that while this has a marginally stronger

effect than edge rewiring, the tails of the original and perturbed distributions are quite similar within the bounds of the error-bars. Furthermore, the positions of the peaks remain unchanged. Varying the area (and therefore density of nodes) and conducting the same procedure over multiple cities yielded identical results (Fig. S9), suggesting that the distribution of (spatial) edge-weights has negligible effect on the BC distribution.

While the procedure outlined above does not preserve the local topology (degree of individual nodes) it is possible to change the edge-weights while preserving the degree sequence of nodes, by taking the original street network and randomly sampling from its associated distribution of distances, assigning each edge a number from this distribution (the edge-weights now *do not* correspond to physical distances but can be interpreted instead as a cost function such as speed-limits). In Fig. 2D we show the average of this process over a hundred realizations (green triangles) where each realization corresponds to a reshuffling of the edge weights over the network. We now begin to see some changes in the distribution with a minor shift in the position of the peaks and a moderately heavier tail, although no drastic modifications are apparent. Strikingly, sampling from a whole family of distributions for the edge weights (exponential, power-law, log-normal) produced identical results (Fig. S10). In other words, while there was a small deviation from the original BC distribution as a result of decoupling edge-weights from any spatial dependence, comparing the resulting perturbed distributions yielded little-to-no dependence on the specific distribution of edge-weights.

Relaxing planarity: Finally, we probe the effects of relaxing the condition of planarity. Fixing N, the degree-sequence, and assigning weights sampled from the distance distribution of Phoenix, we use the configuration model [52] (given a degree sequence, a random graph is constructed by uniformly and randomly choosing a matching on the degree stubs emanating from each node) to generate one hundred non-spatial versions of the street network resulting in the markedly different curve in Fig. 2D (purple triangles). The shape of the curve is in line with the known dependence of g_B on the degree for non-spatial networks, with a distribution of degrees peaked around $k = 3$ (Figs. S11 and S12). The markedly different shape of the curve as compared to the actual street network shows that planarity appears to be the dominant factor specifying the BC distribution, with topological effects and edge-weights playing only a negligible role. While this provides an explanation for the observed similarity

across cities despite their significant geospatial variations, it does not by itself provide an explanation for the form of the distribution, its scaling with N, nor its bimodality, and we will provide in the following some theoretical arguments.

Modeling the BC distribution

A clue for the bimodal behavior stems from the fact that it is peaked at N, a feature reminiscent of nodes adjacent to the leaves of a *minimum spanning tree* (MST). Indeed a MST consists of the subset of edges connecting all nodes with the minimum sum of edge-weights [53] and whose betweenness value is of $O(N)$, specifically $N-2$ for degree two nodes adjacent to leaves. Indeed all paths from the leaf to $N-2$ other nodes have to go through this node. In the context of street networks, their analogs are nodes adjacent to dead-ends (or terminal points) provided they are not part of any loops. An examination of the BC distribution of trees therefore, may provide a qualitative explanation for the scaling behavior found in our data.

Cayley-tree approximation: While deriving an exact analytical expression for the BC distribution of generalized MST's is challenging, one can make progress by approximating it as a k-ary tree (where each node has a branching ratio bounded by k). Given that the degree distribution of cities is tightly peaked (Fig. S11), we can make a further approximation by assuming a fixed branching ratio, in which case the k-ary tree reduces to a Cayley tree where all non-leaf nodes have degree k. Assuming all leaf nodes are at the same depth l and adopting the convention $l = L$ for the leaf level and $l = 0$ for the root, a simple calculation reveals that for a node v at level l, the betweenness scales as $g_B(v|k,l) \sim O(Nk^{L-l})$. After a sequence of manipulations (see Materials and Methods), it can be shown that

$$P(g_B) \approx \frac{k^{\log_k\left(\frac{AN}{g_B}\right)}}{N} = Ag_B^{-1}. \tag{3}$$

where A is a constant of proportionality. Therefore the node betweenness of a Cayley tree scales with exponent $\alpha = 1$, consistent with previous calculations of the link betweenness [54]. Qualitatively at least, these arguments provide a possible explanation for both the scaling with N as well as the form of the tail found in the empirical measurements of the BC of city streets (Eq. (2)), implying an underlying tree structure on which the high BC nodes of all cities lie, indicating that the majority of flow in the is concentrated around *a spanning tree* of the street network [55]. A similar feature is also seen for the BC of general weighted (non-planar) random graphs for certain specific families of weight distributions [49].

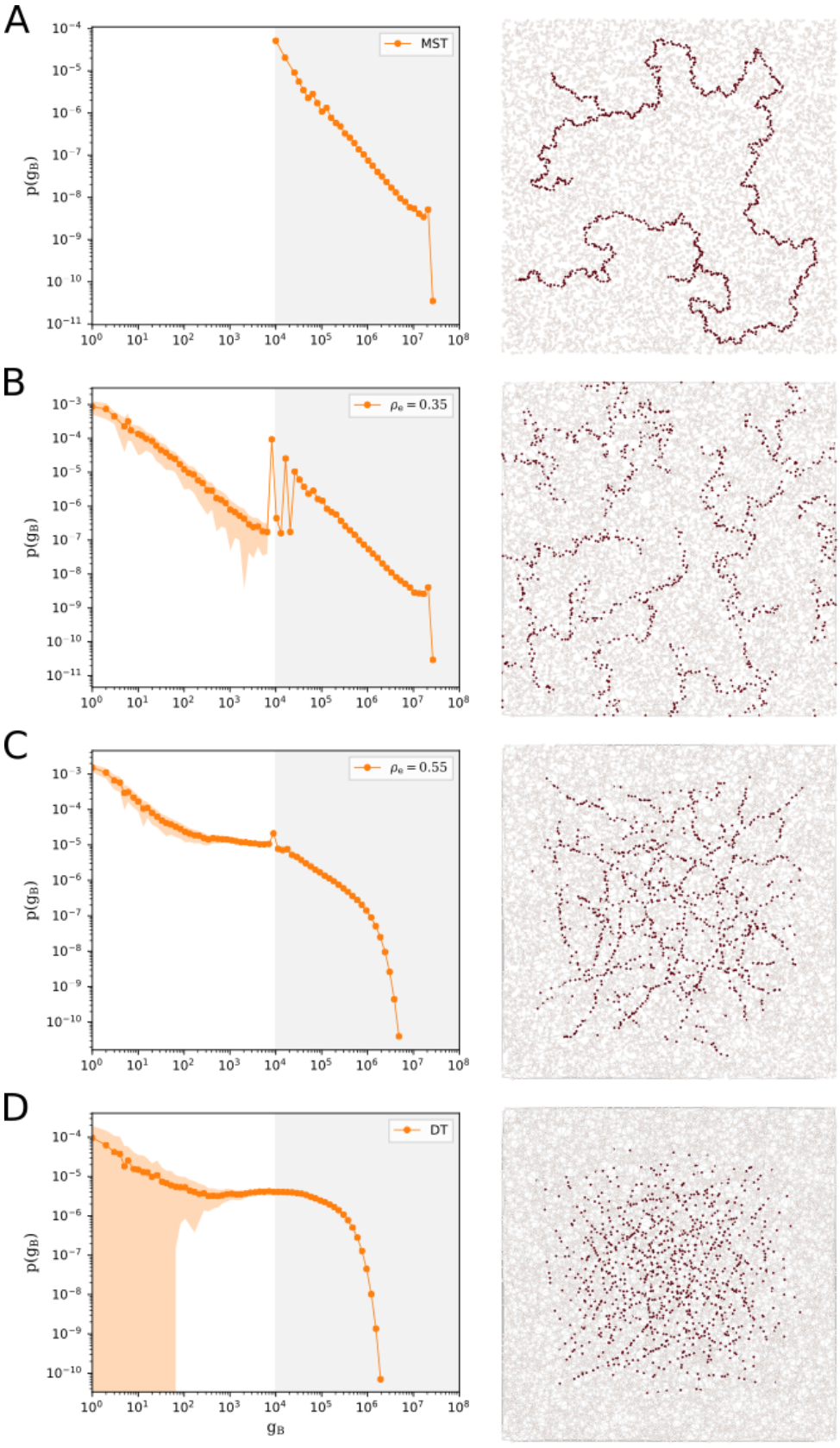


FIG. 3. **Effect of edge-density ρ_e on the betweenness**

FIG. 3. $N = 10^4$ nodes were randomly distributed on the 2D plane and their DT was generated. Edges were removed until the desired edge-density ρ_e was reached. The left panel shows the averaging over a hundred realizations of the resulting BC distribution ranging from the MST constructed over all nodes **(A)** to the DT **(D)** with increasing ρ_e. The orange shaded area corresponds to fluctuations around the average of the realizations, while the silver and white shades separate the "tree-like" region from the "loop-region" respectively. The right panel shows a single instance of the actual generated network corresponding to each ρ_e. Shown in red are the nodes in the 90'th percentile and above in terms of their BC value.

A simple model with variable density: Of course, street networks are not trees and contain loops given by the cyclomatic number $\Gamma = e - N + 1$ (for 1 connected component) where N is the number of nodes, e is the number of edges. In the absence of any loops (such as in the MST) we have that $N = e+1$, and given that N is fixed, the addition of any further edges will necessarily produce loops leading to alternate local paths for navigation. With an increasing number of edges (and therefore more alternate paths), one would expect a large fraction of the (previously) high BC nodes lying on the MST to be bypassed, therefore decreasing their contribution to the number of shortest paths. This will induce the emergence of a low BC regime as well as increasingly sharp cutoffs in the tail, in line with the empirical observations of street networks (Fig. 2). In order to study theoretically the impact of increasing edges on the BC distribution, we study a simple model of random planar graphs. Given that $e \sim O(N)$ and that N itself varies over three orders of magnitude in our dataset, we define a control parameter which we call the *edge density* thus,

$$\rho_e = \frac{e}{e_{DT}}, \tag{4}$$

defined as the fraction of extant edges e compared to the maximal number of possible edges constructed on the set of nodes (given by intersections), as determined by its Delaunay Triangulation e_{DT}, and which varies from $\approx 1/3$ for the MST to 1 for the DT [51]. For a maximally planar graph (i.e one in which no more edges can be added without violating the planarity constraint), we have that $e_{DT} \approx 3N$, so the metric captures the ratio of edges to nodes, or in the context of street networks, the average degree $\langle k \rangle$ of street intersections. Of course in the latter case, the limit $\rho_e \to 1$ is unlikely given some of the geographic constraints inherent in cities.

Having defined this control parameter, we distribute uniformly N nodes in the 2D plane and we first study the MST. In order to be able to vary the density, we generate the Delaunay triangulation on the set of nodes and remove edges until we reach the desired value for ρ_e. The left panel of Fig. 3 shows the BC distribution resulting from a hundred realizations of this procedure for $N = 10^4$ and for increasing values of ρ_e from the MST (top) to the DT (bottom). The BC distribution for the MST seen in Fig. 3A is peaked at N and is bounded by $N^2/2$ which gives here a range of order $[10^4, 10^8]$. In this interval, the BC distribution follows a form close to our calculation for the Cayley tree Eq.(3). As one increases ρ_e and creates loops in the graph, we see the emergence of a bimodal form, with a low BC regime resulting from the bypassing of some of the high BC nodes due to the presence of alternate paths (Fig. 3B). Note that the distribution continues to be peaked at N and the tail maintains its shape. As ρ_e is further increased, the distribution gets progressively more homogenous, yet remains peaked around N even as we approach the limiting case of the DT (Fig. 3D). As a guide to the eye, we shade the "tree-like" region from the "loop-like" region separated by the peak at N.

These results suggest that the observed bimodality seen in the BC distribution for cities stems from the presence of a backbone of high BC nodes, belonging to the MST, decorated with loops. Nodes on these loops contribute to the low BC regime. The transition between the two regimes – low versus high BC nodes – is determined by the minimum non-zero betweenness value for the MST, which is $O(N)$ and the tail may have different peaks, determined by the distribution of branches emanating from the tree. Progressively decorating the tree with loops leads to arbitrarily low betweenness values due to the creation of multiple alternate paths, thus smoothing out the distribution, as the betweenness transitions from an interval $[N, N^2/2]$ for the MST to a continuous distribution over $[1, N^2]$ for the DT.

Spatial distribution of high BC nodes: characterization

Simulations of our random graph model reveals an additional interesting feature. The right-hand panel of Fig. 3 shows a single instance of the actual network generated by our procedure for each corresponding edge-density. Highlighted in red are nodes lying in the 90^{th} percentile and above in terms of their BC. For these nodes, there is a distinct change in spatial pattern with increasing ρ_e. At the level of the MST, they span the network and are tree-like with no apparent spatial correlation; as the network gets more dense, one sees

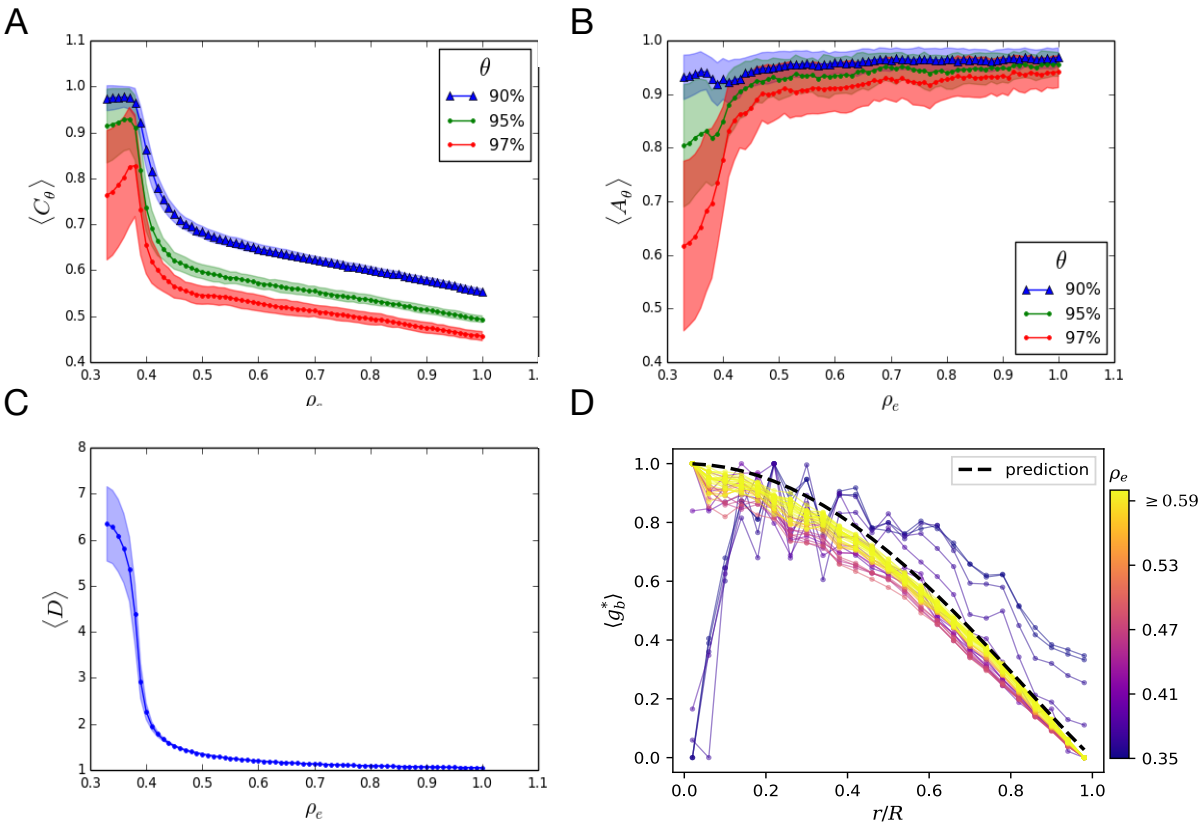


FIG. 4. **Quantifying the spatial effect of edge-density ρ_e on high BC nodes (A)** The metric $\langle C_\theta \rangle$ (Eq. (8)) decreases for denser networks, capturing the tendency of the nodes to be increasingly clustered around their center of mass. **(B)** Correspondingly they also become more isotropic around this center as A_θ (Eq. (10)) approaches 1. **(C)** The network also becomes increasingly geometric as indicated by the decrease in the average detour factor $\langle D \rangle$ (Eq. (11)) measured for the *full network*, which experiences an abrupt transition around $\rho_e \sim 0.4$. The shaded regions represent fluctuations over hundred realizations of the randomization procedure. **(D)** The average BC for nodes at a distance r from the barycenter (rescaled to the interval [0,1]), measured in units of r/R where $R = 50$ is the grid boundary. Curves are colored according to the value of ρ_e. The dashed line corresponds to the analytical calculation for an infinitely dense random geometric graph [56]. The metrics are computed for the networks generated in Fig. 3.

a tendency of these nodes to cluster together and move closer to the barycenter, suggesting a transition between a "topological regime" and a "spatial regime".

To quantify these observed changes, we investigate the behavior of the high BC nodes

at percentile θ through a set of metrics: the clustering C_θ which measures the spread of high BC nodes around their center of mass, the anisotropy factor A_θ which characterizes the spatial anisotropy of this set of nodes, and finally, the detour factor D which measures the average extent to which paths between two locations deviate from their geodesic distance (details in Materials and Methods). In Fig. 4A we plot the quantity $\langle C_\theta \rangle$ for $\theta = 90$, 95, and 97 for different values of the edge-density, finding a clear asymptotic decrease with ρ_e (here $\langle \ldots \rangle$ indicates averaging over realizations). Indeed the decrease is approximately by a factor of two from the MST to the DT, confirming the spatial clustering of the nodes to be a robust effect. In Fig. 4B the plot of $\langle A_\theta \rangle$ in function of ρ_e, for the same set of thresholds as before, indicates a growing isotropic layout and is indicative of a transition from a quasi one-dimensional to a two-dimensional spatial regime. This is confirmed by the corresponding decrease in the detour factor shown in Fig. 4C. Indeed there is a rapid drop around $\rho_e \approx 0.4$ (or equivalently $\langle k \rangle \approx 2$) which is near the density region when the network transitions from a tree-like to a loop-like regime. The appearance of loops in the graph has the additional effect of significantly lowering the detour leading to short paths that are increasingly straight in the geometric sense. Taken together, the behavior of the structural metrics suggests that the spatial position of a node has little relevance to its BC in a sparse network, whereas it assumes increasing importance for dense networks. This is confirmed by plotting the rescaled average BC of nodes as a function of the distance r from the barycenter as shown in Fig. 4D (see Materials and Methods). For low values of ρ_e there appears no distance dependence of the nodes, whereas for $\rho_e > 0.4$, a clear r dependence emerges with the curves converging to the form seen for maximally dense random geometric graphs as calculated in [56]. (Note that while both planar and geometric graphs are embedded in space, the latter allows for edge-crossings and therefore broader degree distributions and larger number of edges for the same N. In light of this difference, the similarity between the two ostensibly different classes of graphs is notable.)

Having observed the spatial behavior of the high BC backbone in random graphs, we next investigate this in the 97 cities. The distribution of ρ_e in Fig. 5A lies in a tight range ($0.4 \leq \rho_e \leq 0.6$) with the majority of cities peaked at $\rho_e \approx 0.5$. The absence of cities with large edge-densities is to be expected, given the nature of street networks, where a node does not exist independently (as in the random graph) but necessarily corresponds to the intersection of streets. Nevertheless, the narrow range is surprising, given the inherent

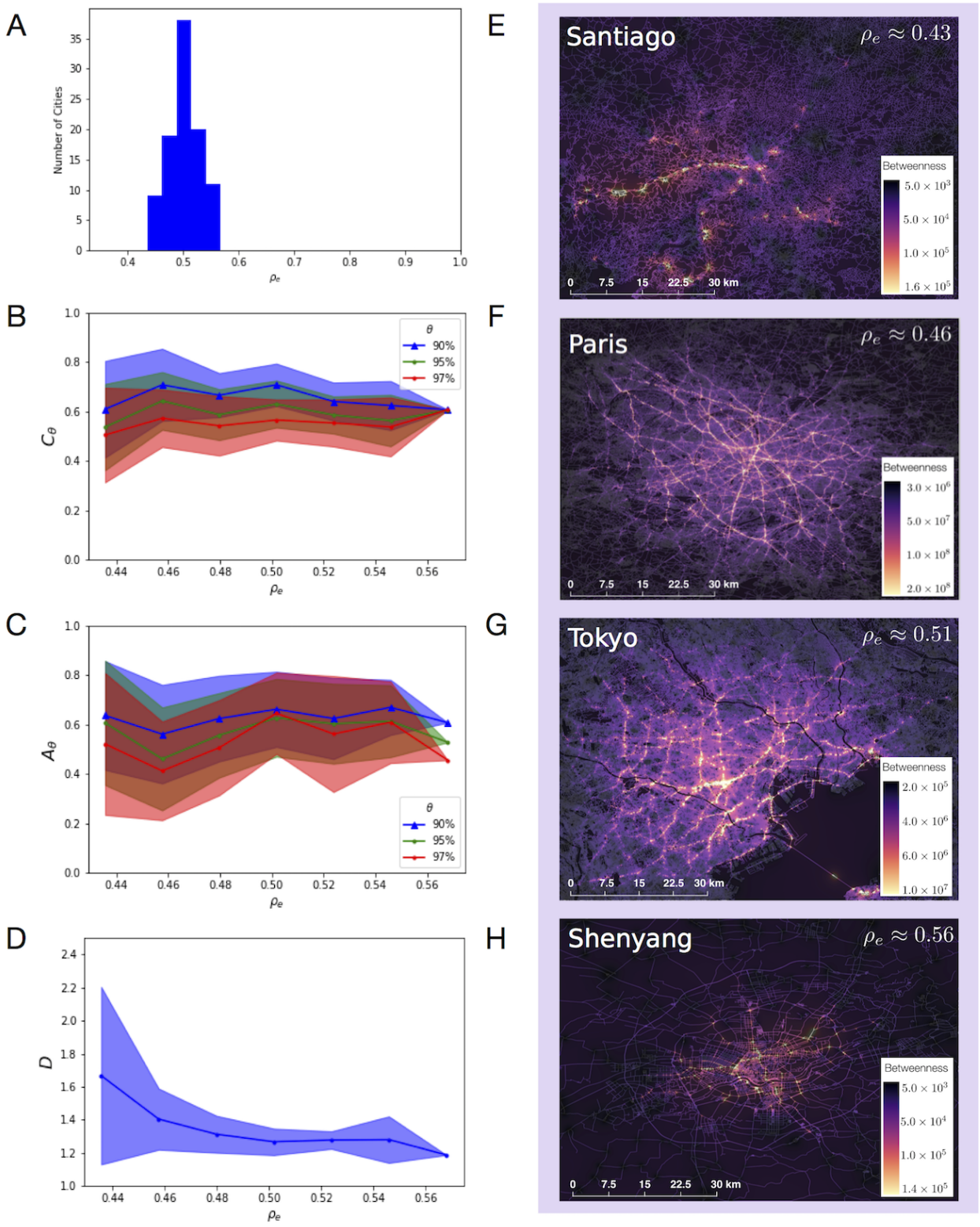


FIG. 5. **Spatial behavior of high BC nodes in real cities**

FIG. 5. **(A)** Distribution of edge-densities for the 97 cities lie in a narrow range $0.4 \leq \rho_e \leq 0.6$ with most cities peaked at $\rho_e \approx 0.5$. **(B)** Variation of the spatial clustering $\langle C_\theta \rangle$ and **(C)** anisotropy ratio $\langle A_\theta \rangle$ with ρ_e for the same range of thresholds used for the random graph models . **(D)** Detour factor for the full street network across cities plotted according to their edge-density. Points are averages over cities within a bin-size of $\rho_e = 0.02$ and the shaded areas represent the fluctuations within the bins. **(E–H)** Spatial layout of intersections in four representative cities in increasing order of ρ_e. The color scale goes from purple to yellow with increased BC. The functional trends of the metrics and the geospatial patterns for the cities are consistent with what is observed for the random graph model described in the text.

geo-spatial and infrastructural differences between the cities. Furthermore, it is notable that they lie in a range that corresponds to the bimodal regime in Fig. 3, providing further explanation for the similarity in their observed BC distributions. On the other hand, this provides a limited window for checking the spatial trends; indeed the curves for $\langle C_\theta \rangle$, $\langle A_\theta \rangle$ and D shown in Figs. 5B,C,D are noisy. Fluctuations arise due to a combination of smaller samples compared to those generated in our random graph simulations, as well as the fact that we are averaging over cities with the same edge-density but different N. Yet, within the extent of fluctuations, the trend is reasonably consistent with that seen in Fig. 4 for the same range of ρ_e. A clearer picture emerges when looking at individual cities; in Fig. 5E-H we show the geospatial layout of the BC distribution for the full street network in four representative cities arranged in increasing order of edge-density. Santiago, being a city with relatively sparse number of streets, shows a tree-like anisotropic pattern for the high BC nodes that are spread mostly along a single axis of the city. Paris and Tokyo, being in the intermediate range, show a complicated lattice-like structure, with a presence of loops among the high BC nodes that span the spatial extent of the cities and appear to be relatively isotropic around the center. Finally, Shenyang, being a city from the upper range of densities, shows a clear (relatively symmetric) clustering of the high BC nodes around the city center.

Beyond static structure: Temporal evolution of BC in cities

The observed changes in the random graph structure seen in Fig. 3, can be thought of as a proxy for the evolution of a city as it experiences change in infrastructure with increased connectivity. While historical data of the evolution of the full street networks in cities is

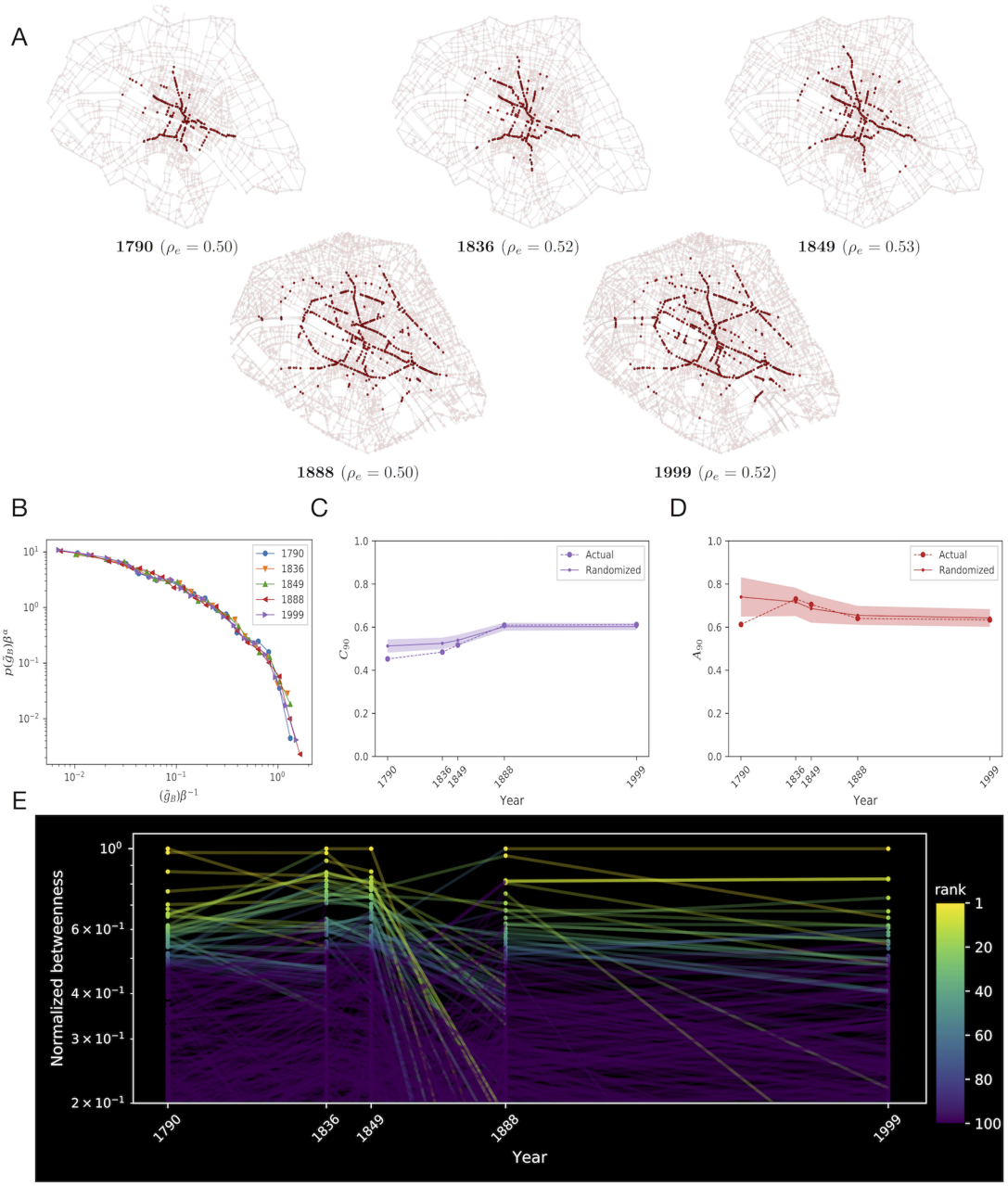


FIG. 6. **Evolution of central Paris from 1790–1999 (A)** Five snapshots of a portion of central Paris spanning two hundred years. Colored in red are the nodes corresponding to the 90'th percentile in terms of their BC. **(B)** The rescaled BC distributions (using the same method as in Fig. 2C) for all five networks showing that they are identical. **(C)** The clustering and **(D)** anisotropy metrics for the nodes in the 90^{th} percentile. Also shown are the corresponding metrics for one hundred realizations of randomized versions of the networks according to the procedure used in Fig. 3. **(E)** Temporal evolution of the BC of individual nodes that are present in all five networks. Points are colored according to rank based on BC.

limited (or at least hard to get), progress can be made by examining smaller subsets. To this effect, we make use of data of five historical snapshots of a portion of central Paris spanning two hundred years (1790–1999), previously gathered to study the effects of central planning by city authorities [41]. The selected portion of Paris is around thirty square kilometers with about 10^3 intersections and road-segments, and represents the essential part of the city around 1790. This particular period $1790 - 1999$ was chosen to examine the effects of the so-called "Hausmann transformation", a major historical example of central planning in a city that happened in the middle of the 19^{th} century in an effort to transform Paris and to improve traffic flow, navigability and hygiene (see [41] and [57] for historical details). In Fig. 6A we show five instances of the street network (1790, 1836, 1849, 1888, 1999), corresponding to the region clipped to 1790. Highlighted in red are the high BC nodes that lie in the 90^{th} percentile. The spatial pattern of the nodes remains virtually identical (with a radial, spoke-like appearance) until 1849, and experience an abrupt change to a ring-like pattern in 1888 which continues to persist for a hundred years. This change in pattern corresponds to the period after the Haussmann transformation, which involved the creation of a number of new roads, broader avenues, new city squares among other things. Yet, it is important to note, that relative to the spatial extent of the region these high BC nodes continue to be located near the city center. Also of note is the relative stability of the edge-density ($\rho_e \approx 0.5$) across the temporal period, reflecting the fact that both nodes and edges are growing at the same rate (Fig. S13).

The rescaled BC distribution, $\tilde{g_B}$, is identical for all 5 snapshots as seen in Fig. 6C despite the structural changes brought about by the Haussman transformation (this is further indication of the marginal effect of local topological variations in the global BC distribution). In Figs. 6C and D, we show the clustering $\langle C_{90} \rangle$and anisotropy metrics $\langle A_{90} \rangle$ for the different eras, which capture the transition from the radial to the ring pattern, but are nevertheless relatively flat in correspondence with what one would expect to see in the random graph for fixed ρ_e. For purposes of comparison, we also plot the averaged metrics for hundred random realizations of each of the five networks, that show a remarkable similarity between the original and randomized cities. To track the evolution of the BC at the local level, we identify those intersections that are present throughout the temporal interval (within a resolution of fifty meters) and compute their betweenness in each instance of the network normalizing by N^2 to provide a consistent comparison, given the historical increase in intersections and

roads. In Fig. 6E we plot the temporal evolution of g_B/N^2 for these intersections, coloring the points according to their corresponding relative rank. While one observes significant fluctuations in the BC at the local level (as expected), the high BC nodes are relatively stable from 1790-1849. After the Haussmann intervention, one observes a dramatic drop in rank of the high BC nodes—corresponding to the "decongesting" spatial transition from a radial to a circular pattern seen in panel A—after which once again the high BC nodes are relatively stable till 1999. Yet it is important to note that the load is simply redistributed to a different part of the network, as can be seen by the transition of the middle-ranked nodes to the top positions in the same periods. Furthermore, as indicated by the spatial layout of these "new" high BC nodes, they continue to be relatively close to the center (few or none are near the periphery), a pattern that is consistent with what one would expect to find for the corresponding random graphs.

Discussion

Taken together our results shed new light on the understanding of structural flow in (spatial) infrastructural networks. The remarkable invariance in the BC distribution across 97 cities seems to be a function of the strong constraint imposed by planarity, leaving only the number of nodes N and the number of edges e as tunable parameters vis-a-vis the BC on the network—a markedly different phenomena than seen for non-planar networks, where betweenness is strongly correlated with degree (or local topology). Analytical calculations on Cayley trees, coupled with simulations of random planar graph models, suggest this to be a consequence of a bimodal regime consisting of a tree-like structure with a tightly peaked branching ratio comprising the high betweenness "backbone" of the network, and a low betweenness regime dominated by the presence of loops. The transition of nodes between regimes is driven by increasing the density of edges in the network, which has the additional effect of introducing a spatial correlation in the high BC nodes—from being dominated by topology in the low density regime to being strongly dependent on spatial location in the high density regime, features also seen in the spatial distribution of the BC in real cities. Given that the number of roads and intersection in our sampled cities vary over three orders of magnitude, the similarity in the BC distribution can be explained as a function of the observed narrow range of ρ_e.

On the other hand, the relative lack of sensitivity of the BC distribution to spatial

layout, including distances and local topological variations, has some interesting implications for urban planning. While the random graph models are closer in spirit to so-called self-organized cities that grow organically, the evolution of Paris suggests that central planning may also have its limitations. The invariance of the BC distribution suggests that traffic cannot be alleviated, but only redirected to different parts of the city. Indeed, the Haussmann transformation succeeded in doing precisely that by improving the navigability of Paris and decongesting the center. However the high BC backbone continued to be closer to the center than the city periphery, a consequence of the spatial distribution being a function of ρ_e. For cities with a higher ratio of roads to intersections, the "decongestion-space" as it were, is expected to be even more limited.

Of course modern metropolises are far more complex than a two-dimensional network of roads, consisting of underpasses, flyovers, bypasses and multimodal transport systems. Indeed, the results of our analysis lend weight to the argument for making such refinements rather than trying to connect different parts of the city via planar roads. Additionally, our analysis does not take into account the *dynamics* of traffic which are a function of how inhabitants actually sample these roads [47]. In fact, a strategy to circumvent the heavy load of cities with high edge-densities may be to manipulate the traffic over an *effective* network by limiting use in certain intersections and roads, thus lowering ρ_e. Such a strategy has long been employed in Singapore, for example, where there is an economic cost applied to the use of roads in the central part of the city, although the motivations for that do not necessarily correspond to the constraints outlined here.

Generally speaking, the study of high BC nodes is an important endeavor as they represent the bottlenecks in the system. In some sense, they represent a generalization of studying the maximum BC node, that governs the behavior of the system in saturation cases where the traffic exceeds the node-capacity. Our analysis however suggests that it is important to take into account the entire high BC set, since the maximum BC node can easily change due to local variations, yet is guaranteed to lie somewhere along the spanning tree that constitutes the backbone of the network. In this respect, further study of the mechanisms governing the spatial distribution of BC is important, especially given the fact that cities lie in a parameter regime with non-negligible spatial correlations in their BC.

Planar graphs are not limited to roads but include other infrastructural networks such as power grids as well as transport networks found in biology and ecology [58]. In particular, leaf

venation networks, arterial networks, and neural cortical networks rely on tree-like structures for optimal function. The lessons from this analysis may well be gainfully employed in these other sectors.

Material and Methods

Construction of street networks

The street networks used in our analysis were constructed from the OpenStreetMaps (OSM) database [59]. For each city we extracted the geospatial data of streets connecting origin-destination pairs within a 30 kilometre radius from the city center (referenced from `latlong.net` [60]), corresponding to a rectangular area of approximately 60×60 square-kilometers with some variability due to road densities, latitude and topographical variations. The 30 kilometer radius was chosen to encapsulate both high density urban regions and more suburban regions with fewer, longer streets. The locations of the street-intersections were found using an Rtree data structure for expedited spatial search [61]. Lattitude and longitude coordinates were projected onto global distances using the Mercator projection, and adjacent intersections lying along the same roads were adjoined by edges with weights equal to the Euclidean distance between the intersections. The resulting street networks are weighted, undirected planar graphs with intersections as nodes, and edges between these nodes approximating the contour of the street network. Aggregate statistics are shown in

	Nodes N	Edges e	Length ℓ (km)	Area A (km^2)	Density ρ
mean	83528.87	130253.05	17461.68	4600.08	18.02
stdev	90335.10	143060.21	15052.83	1926.00	15.43
min	3349.00	5020.00	1793.45	777.07	1.00
25%	18925.00	28518.00	5789.36	3184.32	5.35
50%	62451.00	95797.00	12812.46	4411.81	14.98
75%	118712.00	178773.00	23751.22	5873.67	26.59
max	612418.00	976040.00	82586.30	11562.73	93.47

TABLE I. **Aggregate statistics for the 97 street networks.** Shown are the average, standard deviation, minimum, maximum and various percentile values for the area A, number of intersections (nodes) in the network N, number of roads (edges) e, total length of streets l and the density $\rho = N/A$ of intersections. Details for individual cities shown in Tab. S1.

Tab. I.

BC of Cayley trees

Let us consider a perfect Cayley tree of size N with fixed branching ratio k and all leaf nodes at the same depth. Adopting the convention $l = L$ for the leaf level and $l = 0$ for the root, a node on the l-th level has $k-1$ branches directly below it at the $(l+1)$-th level, each with M_{l+1} children such that the set of branches $\{n_i\}$ stemming from this node will have sizes $\{n_i\} = \{M_{l+1}, ..., M_{l+1}, N - M_l\}$. For fixed k there are $k-1$ copies of the term M_{l+1} which is of the form

$$M_\lambda = \sum_{l'=0}^{L-\lambda} k^{l'} = \frac{1 - k^{L-\lambda+1}}{1-k}. \tag{5}$$

The betweenness value of a vertex v in any tree is given by $g_B(v) = \sum_{i<j} n_i n_j$ where i, j are indices running over the branches coming off of v (excluding v), and n_i, n_j are the number of nodes in each branch [62]. Combining this with Eq. (5) gives us the betweenness of v at level l thus

$$g_B(v|k,l) = \binom{k-1}{2} M_{l+1}^2 + (k-1) M_{l+1} \left(N - M_l\right), \tag{6}$$

from which it is easy to see that for any level l, the betweenness scales as $g_B(v|k,l) \sim O(Nk^{L-l})$. Thus, absorbing k^L into the leading constant and letting $g_B(v|k,l) \approx ANk^{-l}$, we have that since g_B is completely determined by the level l in which it lies in the tree:

$$P(g_B) = \sum_l P(g_B|l)P(l) \approx \sum_l \delta_{g_B, ANk^{-l}} \frac{k^l}{N} = \frac{k^{\log_k\left(\frac{AN}{g_B}\right)}}{N} = A g_B^{-1}. \tag{7}$$

Spatial metrics for high BC nodes

To measure the clustering, we specify a threshold θ—i.e. we isolate nodes with a BC above the θ-th percentile—and then compute their spread about their center of mass, normalizing for comparison across networks of different sizes, thus,

$$C_\theta = \frac{1}{N_\theta \langle \mathbf{X} \rangle} \sum_{i=1}^{N_\theta} ||\mathbf{x}_i - \mathbf{x}_{\text{cm}}||. \tag{8}$$

Here $\mathbf{x}_{\text{cm}} = \sum_{i=1}^{N_\theta} \mathbf{x}_i$, N_θ is the number of high betweenness nodes isolated, $\{\mathbf{x}_i\}$ specify their coordinates, and $\langle \mathbf{X} \rangle$ is the average distance of *all nodes* in the network to the center of

mass of the high BC cluster,

$$\langle \mathbf{X} \rangle = \frac{1}{N} \sum_{i=1}^{N} ||\mathbf{x}_i - \mathbf{x}_{\mathrm{cm}}||. \tag{9}$$

Eq. (8) quantifies the extent of clustering of the high BC nodes relative to the rest of the nodes in the network, with increased clustering resulting in low values of C_θ.

In order to more precisely quantify the transition between the topological and spatial regimes, a clue is provided by the increasingly isotropic layout of the high BC nodes with increasing edge-density. To measure the extent of this observed (an)isotropy, we define the ratio,

$$A_\theta = \frac{\lambda_1}{\lambda_2}, \tag{10}$$

where $\lambda_1 \leq \lambda_2$ are the (positive) eigenvalues of the covariance matrix of the spatial positions of the nodes with BC above threshold θ. The metric is unitless and measures the widths of the spread of points about their principal axes, analogous to the principal moments of inertia. Low values of A_θ correspond to a quasi one-dimensional structure with large anisotropy, whereas the system becomes increasingly isotropic for larger values until it is roughly two-dimensional as $A_\theta \to 1$.

The detour factor measures the average extent to which paths between two locations deviate from their geodesic distance and is given by

$$D = \frac{1}{N(N-1)} \sum_{i \neq j} \frac{d_G(i,j)}{d_E(i,j)}. \tag{11}$$

Here $d_E(i,j)$ is the euclidean distance between nodes i, j and $d_G(i,j)$ is their distance-weighted shortest path in the network G.

Distance dependence of BC

In our simulations, nodes were located on a 100×100 grid with coordinates in $\mathbb{R}^2 \in [-50, 50]$. The center of the grid was chosen as the origin $(0,0)$ and the average betweenness $\langle g_B(r) \rangle$ is computed over all nodes that are located at a distance r from the origin, advancing in units of $r = 1$, until we reach the grid boundary $r = 50$. In order to restrict $\langle g_B(r) \rangle$ to the interval $[0,1]$ we measure the rescaled quantity

$$\langle g_b^\star(r) \rangle = \frac{\langle g_B(r) \rangle - \min\langle g_B(r) \rangle}{\max\langle g_B(r) \rangle) - \min\langle g_B(r) \rangle}, \tag{12}$$

for different values of ρ_e. This was done to compare our results to the corresponding expression in random geometric graphs, which was analytically calculated for (the somewhat artificial) limit of an infinitely dense disk of radius R [56].

---

[1] Bretagnolle, A., Daudé, E. & Pumain, D. From theory to modelling: urban systems as complex systems. *CyberGeo: European Journal of Geography* **335** (2006).

[2] Bettencourt, L. & West, G. A unified theory of urban living. *Nature* **467**, 912–3 (2010).

[3] Pan, W., Ghoshal, G., Krumme, C., Cebrian, M. & Pentland, A. Urban characteristics attributable to density-driven tie formation. *Nature Communications* **4**, 1–7 (2013).

[4] Batty, M. Building a science of cities. *Cities* **29**, S9–S16 (2012).

[5] Barthelemy, M. *The Structure and Dynamics of Cities* (Cambridge University Press, 2016).

[6] Barthelemy, M. Spatial networks. *Physics Reports* **499**, 1–101 (2011).

[7] Goh, S., Choi, M. Y., Lee, K. & Kim, K.-m. How complexity emerges in urban systems: Theory of urban morphology. *Physical Review E* **93**, 052309 (2016).

[8] Bettencourt, L. The Origins of Scaling in Cities. *Science* **340**, 1438–1441 (2013).

[9] Kalapala, V., Sanwalani, V., Clauset, A. & Moore, C. Scale invariance in road networks. *Physical Review E* **73** (2006).

[10] Youn, H., Gastner, M. T. & Jeong, H. Price of anarchy in transportation networks: Efficiency and optimality control. *Physical Review Letters* **101**, 128701 (2008).

[11] Cardillo, A., Scellato, S., Latora, V. & Porta, S. Structural properties of planar graphs of urban street patterns. *Physical Review E* **73**, 066107–8 (2006).

[12] Justen, A., Martínez, F. J. & Cortés, C. E. The use of space-time constraints for the selection of discretionary activity locations. *Journal of Transport Geography* **33**, 146–152 (2013).

[13] Witlox, F. Evaluating the reliability of reported distance data in urban travel behaviour analysis. *Journal of Transport Geography* **15**, 172–183 (2007).

[14] da F. Costa, L., Travençolo, B. A. N., Viana, M. P. & Strano, E. On the efficiency of transportation systems in large cities. *EPL (Europhysics Letters)* **91**, 18003 (2010).

[15] Wang, P., Hunter, T., Bayen, A. M., Schechtner, K. & González, M. C. Understanding Road Usage Patterns in Urban Areas. *Scientific Reports* **2**, 1001 (2012).

[16] Kang, C., Ma, X., Tong, D. & Liu, Y. Intra-urban human mobility patterns: An urban

morphology perspective. *Physica A* **391**, 1702–1717 (2012).

[17] Haggett, P. & Chorley, R. J. *Network analysis in geography* (St. Martins Press, New York, 1969).

[18] Lammer, S., Gehlsen, B. & Helbing, D. Scaling laws in the spatial structure of urban road networks. *Physica A* **369**, 853866 (2006).

[19] Wang, F., Antipova, A. & Porta, S. Street centrality and land use intensity in Baton Rouge, Louisiana. *Journal of Transport Geography* **19**, 285–293 (2011).

[20] Rui, Y., Ban, Y., Wang, J. & Haas, J. Exploring the patterns and evolution of self-organized urban street networks through modeling. *The European Physical Journal B* **86**, 74–8 (2013).

[21] Louf, R. & Barthlemy, M. A typology of street patterns. *Journal of The Royal Society Interface* **11**, 20140924–20140924 (2014).

[22] Strano, E. *et al.* Urban Street Networks, a Comparative Analysis of Ten European Cities. *Environment and Planning B: Planning and Design* **40**, 1071–1086 (2013).

[23] Masucci, A. P., Smith, D., Crooks, A. & Batty, M. Random planar graphs and the London street network. *The European Physical Journal B* **71**, 259–271 (2009).

[24] Clark, J. & Holton, D. A. *A first look at graph theory*, vol. 1 (World Scientific, Teaneck, NJ, 1991).

[25] Newman, M. E. J. *Networks: An Introduction* (Oxford University Press, Oxford, 2010).

[26] Aldous, D. & Ganesan, K. True scale-invariant random spatial networks. *Proceedings of the National Academy of Sciences* **110**, 8782–8785 (2013).

[27] Aldous, D. Routed planar networks. *Electronic Journal of Graph Theory and Applications (EJGTA)* **4**, 42–59 (2016).

[28] Ghoshal, G. & Barabási, A.-L. Ranking stability and super stable nodes in complex networks. *Nature Communications* **2**, 394 (2011).

[29] Barthelemy, M. Crossover from scale-free to spatial networks. *Europhysics Letters* **63**, 915 (2003).

[30] Freeman, L. C. A set of measures of centrality based on betweenness. *Sociometry* 35–41 (1977).

[31] Holme, P. Congestion and centrality in traffic flow on complex networks. *Advances in Complex Systems* **6**, 163–176 (2003).

[32] Ashton, D. J., Jarrett, T. C. & Johnson, N. F. Effect of congestion costs on shortest paths

through complex networks. *Physical Review Letters* **94**, 058701–4 (2005).

[33] Jarrett, T. C., Ashton, D. J., Fricker, M. & Johnson, N. F. Interplay between function and structure in complex networks. *Physical Review E* **74**, 026116–8 (2006).

[34] Brandes, U. A faster algorithm for betweenness centrality. *Journal of mathematical sociology* **25**, 163–177 (2001).

[35] Roswall, M., Trusina, A., Minnhagen, P. & Sneppen, K. Networks and cities: an information perspective. *Physical Review Letters* **94**, 028701 (2005).

[36] Jiang, B. A topological pattern of urban street networks: Universality and peculiarity. *Physica A* **384**, 647–655 (2007).

[37] Chan, S. H. Y., Donner, R. V. & Lämmer, S. Urban road networks — spatial networks with universal geometric features? *The European Physical Journal B* **84**, 563–577 (2011).

[38] Lion, B. & Barthelemy, M. Central loops in random planar graphs. *Physical Review E* **95**, 042310 (2017).

[39] Crucitti, P., Latora, V. & Porta, S. Centrality measures in spatial networks of urban streets. *Physical Review E* **73**, 036125–5 (2006).

[40] Porta, S., Crucitti, P. & Latora, V. The network analysis of urban streets: a primal approach. *Environment and Planning B: planning and design* **33**, 705–725 (2006).

[41] Barthelemy, M., Bordin, P., Berestycki, H. & Gribaudi, M. Self-organization versus top-down planning in the evolution of a city. *Scientific Reports* **3**, 2153 (2013).

[42] Barthelemy, M. Betweenness centrality in large complex networks. *European Physical Journal B* **38**, 163–168 (2004).

[43] Strano, E. *et al.* The scaling structure of the global road network. *arXiv preprint arXiv:1706.01401* (2017).

[44] Gago, S., Hurajová, J. & Madaras, T. Notes on the betweenness centrality of a graph. *Mathematica Slovaca* **62**, 1–12 (2012).

[45] Narayan, O. & Saniee, I. Large-scale curvature of networks. *Physical Review E* **84**, 066108 (2011).

[46] Jonckheere, E., Lou, M., Bonahon, F. & Baryshnikov, Y. Euclidean versus hyperbolic congestion in idealized versus experimental networks. *Internet Mathematics* **7**, 1–27 (2011).

[47] Lee, M., Barbosa, H., Youn, H., Ghoshal, G. & Holme, P. Urban socioeconomic patterns revealed through morphology of travel routes. *arXiv:1701.02973* (2017).

[48] Clark, C. Urban population densities. *Journal of the Royal Statistical Society. Series A* **114**, 490–496 (1951).

[49] Wang, H., Hernandez, J. M. & Van Mieghem, P. Betweenness centrality in a weighted network. *Physical Review E* **77**, 046105 (2008).

[50] Clauset, A., Shalizi, C. R. & Newman, M. E. J. Power-law distributions in empirical data. *SIAM review* **51**, 661–703 (2009).

[51] Lee, D.-T. & Schachter, B. J. Two algorithms for constructing a delaunay triangulation. *International Journal of Computer & Information Sciences* **9**, 219–242 (1980).

[52] Newman, M. E. J., Watts, D. J. & Strogatz, S. Random graphs with arbitrary degree distributions and their applications. *Physical Review E* **64**, 026118 (2001).

[53] Graham, R. L. & Hell, P. On the history of the minimum spanning tree problem. *Annals of the History of Computing* **7**, 43–57 (1985).

[54] Szabó, G., Alava, M. & Kertész, J. Shortest paths and load scaling in scale-free trees. *Physical Review E* **66**, 026101 (2002).

[55] Wu, Z., Braunstein, L. A., Havlin, S. & Stanley, H. E. Transport in weighted networks: partition into superhighways and roads. *Physical review letters* **96**, 148702 (2006).

[56] Giles, A. P., Georgiou, O. & Dettmann, C. P. Betweenness centrality in dense random geometric networks. In *2015 IEEE International Conference on Communications (ICC)*, 6450–6455 (2015).

[57] Jordan, D. *Transforming Paris: The Life and Labors of Baron Haussmann* (University of Chicago Press, Chicago, USA, 1995).

[58] Mileyko, Y., Edelsbrunner, H., Price, C. A. & Weitz, J. S. Hierarchical ordering of reticular networks. *PLoS One* **7**, e36715 (2012).

[59] OpenStreetMap Contributors. Openstreetmap (2015). URL `http://planet.openstreetmap.org`. [Online; accessed 02-Febraury-2016].

[60] latlong.net (2016). URL `http://www.latlong.net/`. [Online; accessed 20-December-2016].

[61] Guttman, A. R-trees: a dynamic index structure for spatial searching. In *SIGMOD '84 Proceedings of the 1984 ACM SIGMOD international conference on Management of data*, vol. 14, 47–57 (ACM, New York, 1984).

[62] Sunil Kumar, R., Balakrishnan, K. & Jathavedan, M. Betweenness centrality in some classes of graphs. *International Journal of Combinatorics* **2014** (2014).

ACKNOWLEDGMENTS

This work was partially supported by the US Army Research Office under Agreement Number W911NF-17-1-0127. MB thanks the city of Paris (Paris 2030) for funding and the geohistoricaldata group for discussions and data.

Supporting Information

Structural invariants in street networks: modeling and practical implications

Alec Kirkley, Hugo Barbosa, Marc Barthelemy, Gourab Ghoshal

Table of Contents

S1 Data **31**

S2 Curve fitting and verification **36**

List of Figures

S1 Betweenness pdf's at multiple scales . . . 35
S2 Distribution of g_B for 1sq mile samples and fits to the tail . . . 36
S3 Tails of the BC distributions with their corresponding truncated-power-law fits. . . . 37
S4 Distribution of truncated-power-law parameters . . . 42
S5 The BC distribution of various random graph models . . . 43
S6 2-Sample KS Statistics for City Random Graph Models . . . 45
S7 2-sample KS Test p-values for City Random Graph Models . . . 46
S8 Edge Length Distributions . . . 47
S9 Rescaling the Convex Hull . . . 48
S10 Various Randomized Edge Weight Distributions . . . 49
S11 Aggregate Degree Distribution in Cities . . . 50
S12 Degree Distributions for Individual Cities . . . 51
S13 Evolution of the number of roads in central Paris . . . 52

List of Tables

S1 Street network statistics . . . 32
S2 Parameter values for the truncated power law distributions fitted to the data. . . . 38

S1 Data

Raw shapefile data for the street geometries and locations was collected from the OpenStreetMaps (OSM) database via the `BBBike.org` service. For each city we extracted the street shapefile data of streets connecting origin-destination pairs within a 30 kilometre radius from the city center. The 30 kilometre radius was chosen to encapsulate both high density urban regions and more suburban regions with fewer, longer streets. After aggregating all the raw shapefile data, we populated the Rtree data structure with the linestring (a collection of lattitude/longitude coordinates approximating the contour of the street) geometry of each street. Then, for each street, we found the other streets intersecting it using the Rtree indexing, and cut the street into separate segments at each intersection point, adding a node at each of these points. Lattitude and longitude coordinates of all nodes were projected onto global distances using the Mercator projection, and then an edge was added between nodes adjacent along a given street, with a weight equal to the Euclidean distance between the nodes. After searching through all streets, and checking for connectivity, the street networks were complete. The type of each street, classified into various categories by OSM ('Primary', 'Secondary', 'Tertiary', 'Service', etc), was then added as an attribute to each edge, and two versions of the street network were created for each city. For each city, the entire street network was created, and in addition, a "refined" street network was created to approximate the network of high congestion streets, where only edges classified as primary, secondary, tertiary, highways, or service roads were kept, and all others were pruned, then the giant component of the resulting network was kept. All analyses were perfomed on the entire street network, except for the randomized cities analyses (i.e. randomizing the weights, rewiring the edges, etc), which were done on the filtered street networks for computational tractability. Descriptive statistics for the entire street networks of individual cities are shown in Table S1.

Table S1: Statistics of the street networks sorted by number of nodes N (i.e., intersection). The total length ℓ corresponds to the sum of the lengths of all streets within a convex hull of area A.

City	Nodes N	Edges e	Length ℓ (km)	Area A (km^2)	Density ρ
Tokyo	612418	976040	82586.30	6552.04	93.47
Moscow	307472	482217	61391.85	11562.73	26.59
Nagoya	300588	496495	57990.36	5891.78	51.02
Osaka	292855	469333	47827.03	5968.06	49.07
Paris	279072	425108	57285.37	8911.36	31.32
Milan	201029	299564	38929.31	8412.68	23.90
Berlin	198498	306742	49006.85	10027.52	19.80
Washington DC	183687	276391	35296.45	6464.47	28.41
São Paulo	180843	283349	32579.38	5619.28	32.18
New York City	178120	288278	43137.88	6729.33	26.47
Madrid	177403	273342	33647.67	6119.91	28.99
Houston	175524	270779	34352.18	5278.51	33.25
Delhi	174732	267204	30124.32	6127.09	28.52
Los Angeles	166993	268304	38984.48	4866.74	34.31
Alexandria	162753	244978	31964.20	6371.30	25.54
Mexico City	158528	254762	27406.55	4436.46	35.73
Chicago	157740	258044	34761.24	3992.69	39.51
Toronto	156919	248099	27457.24	4307.80	36.43
Phoenix	153846	235348	32294.83	6097.68	25.23
Hyderabad	151131	231787	19793.77	4471.25	33.80
Istanbul	149511	235069	24757.37	4127.72	36.22
Buenos Aires	138245	241717	29794.35	3063.27	45.13
Philadelphia	122916	192174	31816.90	6252.88	19.66
Khartoum	122634	200241	16405.95	2972.44	41.26
Manila	118712	178773	16436.52	3031.90	39.15
Boston	118573	177186	28054.72	6991.56	16.96
Barcelona	110982	172526	23751.22	4073.86	27.24
London	105198	139541	25931.41	9401.36	11.19
Lima	99750	160214	13356.04	1841.42	54.17
Riyadh	98569	151902	20417.77	3800.11	25.94
Atlanta	92148	131969	23421.24	5949.59	15.49
Kuala Lumpur	89879	133348	16412.21	3777.69	23.79
Rome	86374	129451	19242.76	6093.85	14.17
San Francisco	85635	135493	18801.50	4855.68	17.64
Sydney	82870	123436	17758.48	3273.63	25.31
Rio De Janeiro	82808	129256	15711.31	2956.19	28.01
Johannesburg	78377	121053	21198.68	5231.78	14.98
Jakarta	74128	112467	12520.81	2484.23	29.84

Continued on next page

City	Nodes N	Edges e	Length ℓ (km)	Area A (km^2)	Density ρ
Taipei	74105	118000	15001.50	3834.81	19.32
Monterrey	73981	117784	12812.46	3724.65	19.86
Bangalore	73759	112712	13753.09	3966.04	18.60
Bogota	73648	117684	11133.31	3927.02	18.75
Miami	72411	115085	15548.93	2186.57	33.12
Bangkok	71582	102908	16316.54	4170.52	17.16
Cairo	70777	109185	16406.14	5653.43	12.52
Guadalajara	70145	113418	12226.81	5873.67	11.94
Shenzhen	65286	101370	14927.35	3810.12	17.13
Dubai	62559	91822	12126.64	2478.23	25.24
Hong Kong	62451	96059	11831.31	2716.85	22.99
Ankara	61133	95797	13571.78	5673.95	10.77
Tehran	57177	88127	12898.16	4411.81	12.96
Cape Town	52096	78827	10794.67	2460.27	21.17
Shanghai	50049	82637	19566.98	5539.79	9.03
Chennai	49278	74444	8786.05	2181.86	22.59
Baghdad	48271	75255	10837.62	4839.89	9.97
Santiago	43001	64873	18578.85	6824.06	6.30
Yangon	40840	64890	7689.51	3418.27	11.95
Kolkata	38924	57162	7258.99	3663.49	10.62
Ho Chi Minh City	38311	58902	9277.30	3679.66	10.41
Guangzhou	35921	57460	14606.66	5447.64	6.59
Luanda	35468	57329	7934.13	2101.68	16.88
Mumbain	32720	49535	7182.38	2772.80	11.80
Singapore	29756	44640	5317.58	777.07	38.29
Lahore	28008	43723	7509.84	4821.33	5.81
Surabaya	26420	39506	5467.28	2964.76	8.91
Abidjan	24499	37922	4564.90	2713.18	9.03
Melbourne	22287	33817	5789.36	3330.35	6.69
Kinshasa	21711	35563	4108.92	1624.39	13.37
Accra	21333	32060	5346.94	2062.73	10.34
Dar es Salaam	20754	31061	3564.65	2073.43	10.01
Dongguan	19294	31452	10738.15	5769.56	3.34
Lagos	18936	28066	5084.47	2406.29	7.87
Xian	18925	30592	12378.79	7923.66	2.39
Nanjing	17500	28518	9911.60	6192.78	2.83
Nairobi	17040	24463	5529.59	3184.32	5.35
Bandung	16755	24529	3715.87	3939.03	4.25
Wuhan	16568	26508	8629.57	6446.66	2.57
Pune	16173	23905	4800.26	3905.53	4.14
Tianjin	15461	25641	10362.45	7058.71	2.19
Hanoi	14864	22934	5505.63	4592.05	3.24
Hangzhou	14829	24512	8933.17	5644.03	2.63

Continued on next page

City	Nodes N	Edges e	Length ℓ (km)	Area A (km^2)	Density ρ
Kabul	14137	21517	3931.07	2919.67	4.84
Ahmadabad	13615	21046	4465.10	3988.31	3.41
Chengdu	12521	20724	7967.42	5327.72	2.35
Suzhou	12501	21104	9317.10	4952.66	2.52
Dhaka	12209	18423	3427.93	4836.00	2.52
Medan	10424	15660	2964.14	1749.25	5.96
Xiamen	9679	15652	4598.72	3758.04	2.58
Shenyang	9624	15853	7538.69	6475.61	1.49
Chongqing	8275	13232	5133.81	4740.14	1.75
Qingdao	7095	11911	4470.93	3476.07	2.04
Fuzhou	6310	9945	4519.80	4750.16	1.33
Harbin	6074	9990	4346.88	5059.06	1.20
Dalian	5654	9122	2989.32	2521.05	2.24
Kuwait	4593	6501	1826.65	4595.19	1.00
Quangzhou	3774	6189	3559.67	3248.89	1.16
Surat	3349	5020	1793.45	2635.62	1.27

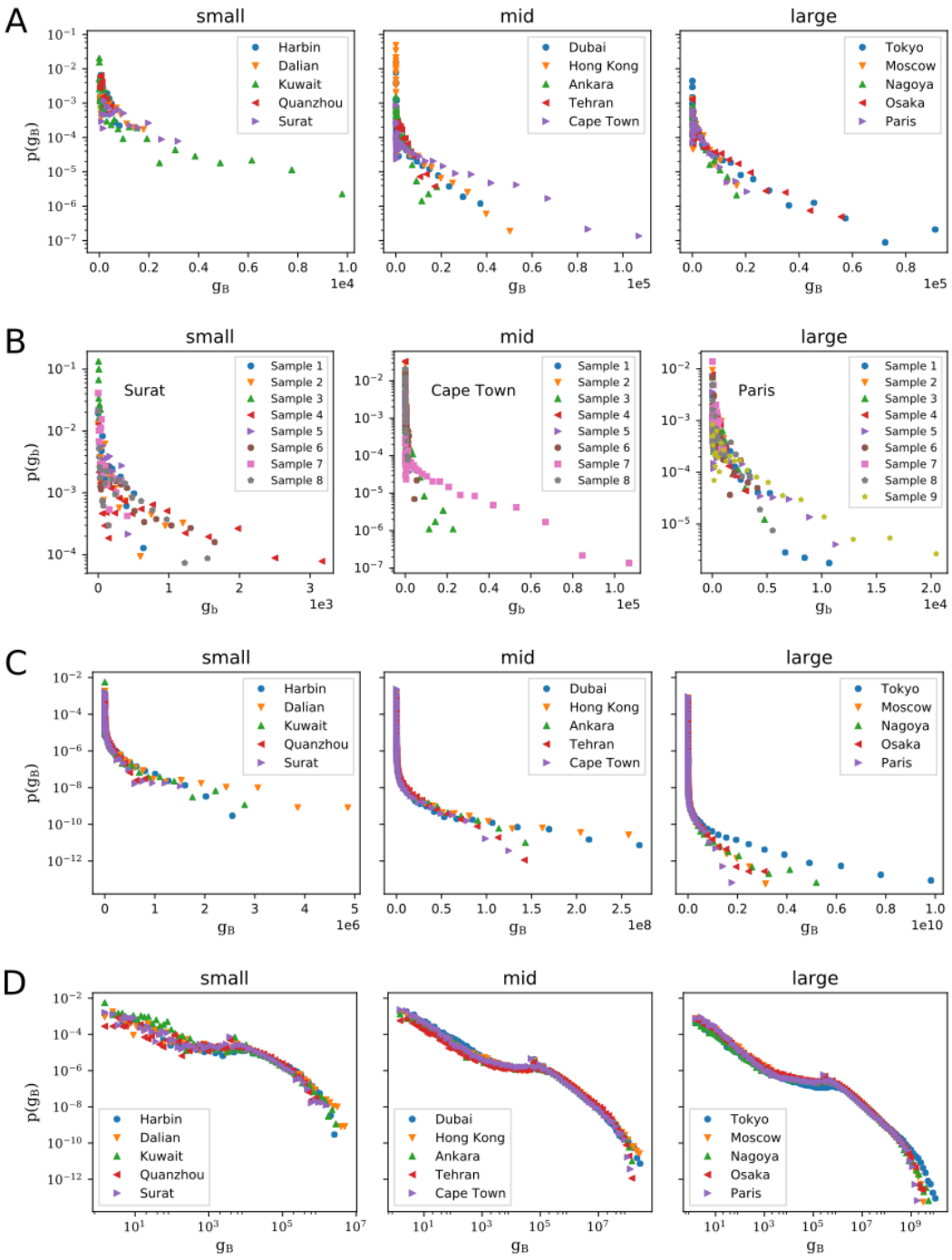


Figure S1: Betweenness pdf's at multiple scales (A) selected one-square-mile samples from each category (log-linear). (B) Multiple one-square-mile samples within a single city picked from each category (log-linear). (C) BC of streets at full resolution $\sim$ 1000 square-miles (log-linear) and finally (D) the same in log-log scale revealing a bimodal distribution.

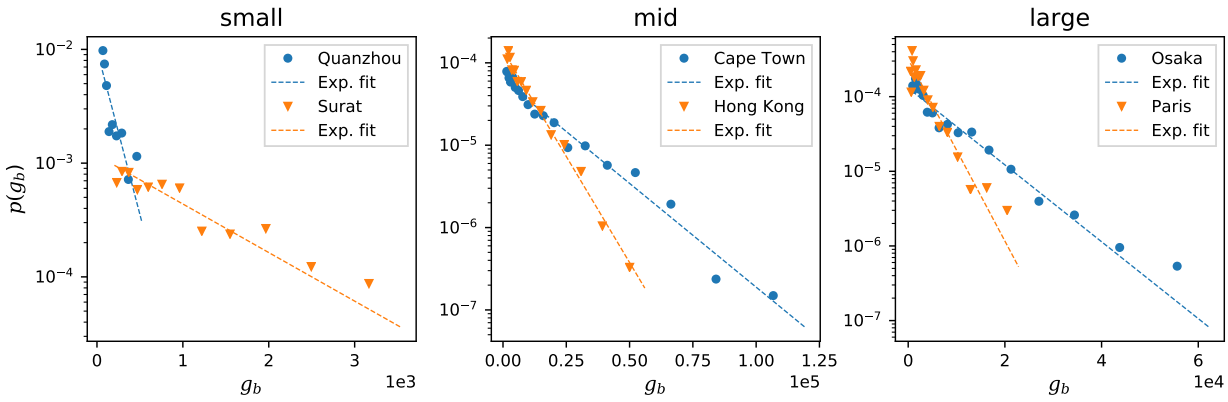


Figure S2: Distribution of g_B for 1sq mile samples for cities of different network sizes and their corresponding exponential fits.

S2 Curve fitting and verification

Curve fits were performed using the maximum likelihood procedures outlined in [50]. As we show in the main document, the tails of the BC distributions are well approximated by a truncated power-law distribution

$$p(\tilde{g}_B) \sim \tilde{g}_B^{-\alpha} e^{-\tilde{g}_B/\beta}.$$

Figure S3 shows the tail of the BC distributions with their corresponding fit lines whereas Table S2 shows the results of the curve fits for each individual city. Additionally, we fit the tails to stretched exponentials of the form

$$p(\tilde{g}_B) \sim \tilde{g}_B^{\gamma-1} e^{-\lambda(\tilde{g}_B)^{\gamma}}.$$

Fits for γ revealed a tightly peaked distribution near $\gamma \sim .3$ for all cities. This indicates that regardless of the exact functional form of the decay, the same power law scaling exponent of ~ -1 persists throughout all cities, with some variation that gets absorbed into the functional form of the tail decay, which is consistent with calculations for the Cayley Tree with fixed branching ratio. Therefore only the results for the truncated power law are reported in the manuscript as the exact functional form of the tails is not germane to the main discussion.

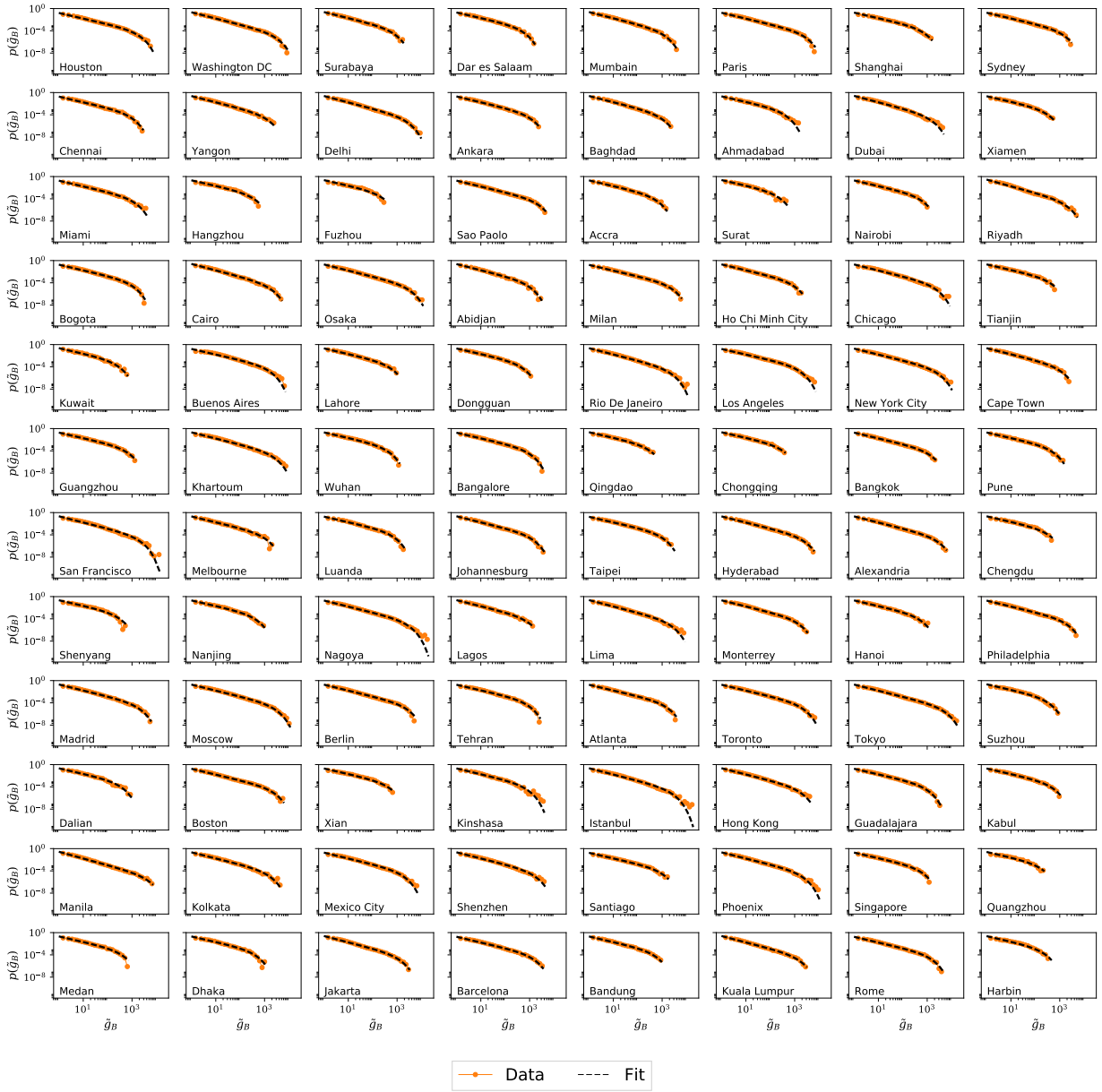


Figure S3: Tails of the BC distributions with their corresponding truncated-power-law fits for all cities.

Table S2: Parameter values for the truncated power law distributions fitted to the data.

City	α	β
Los Angeles	1.000 ± 0.002	833.56
Santiago	1.000 ± 0.003	533.34
Shanghai	1.000 ± 0.003	386.91
Tehran	1.000 ± 0.003	451.24
Taipei	1.000 ± 0.003	544.44
Guangzhou	1.000 ± 0.004	349.23
Luanda	1.000 ± 0.004	372.26
Singapore	1.000 ± 0.005	311.91
Xian	1.000 ± 0.005	246.95
Pune	1.000 ± 0.006	267.52
Wuhan	1.000 ± 0.006	222.75
Hangzhou	1.000 ± 0.006	199.18
Dongguan	1.000 ± 0.006	259.99
Tianjin	1.000 ± 0.006	228.69
Nanjing	1.000 ± 0.006	230.24
Dar es Salaam	1.000 ± 0.006	285.57
Chengdu	1.000 ± 0.007	183.08
Suzhou	1.000 ± 0.007	175.23
Kabul	1.000 ± 0.007	261.67
Dhaka	1.000 ± 0.007	228.10
Ahmadabad	1.000 ± 0.007	214.40
Medan	1.000 ± 0.008	206.62
Shenyang	1.000 ± 0.008	156.62

Continued on next page

City	α	β
Xiamen	1.000 ± 0.009	166.57
Chongqing	1.000 ± 0.009	168.52
Fuzhou	1.000 ± 0.010	133.97
Qingdao	1.000 ± 0.010	159.30
Harbin	1.000 ± 0.011	121.43
Dalian	1.000 ± 0.012	181.44
Quangzhou	1.000 ± 0.014	93.07
Kuwait	1.000 ± 0.015	134.41
Surat	1.000 ± 0.016	116.08
Buenos Aires	1.001 ± 0.002	816.10
Houston	1.008 ± 0.002	1044.60
Guadalajara	1.009 ± 0.003	516.32
Bangalore	1.012 ± 0.003	534.41
Bandung	1.013 ± 0.007	268.97
Nairobi	1.015 ± 0.007	301.92
Hanoi	1.015 ± 0.007	267.13
Chennai	1.018 ± 0.003	501.82
Bogota	1.018 ± 0.003	556.28
Lahore	1.020 ± 0.005	286.75
Sydney	1.021 ± 0.003	736.42
Accra	1.023 ± 0.006	305.32
Rome	1.024 ± 0.002	782.85
Miami	1.025 ± 0.003	679.11
Milan	1.027 ± 0.001	1289.63
Atlanta	1.029 ± 0.002	808.04
Barcelona	1.031 ± 0.002	762.74

Continued on next page

City	α	β
Bangkok	1.037 ± 0.003	677.47
Kinshasa	1.044 ± 0.005	417.81
Paris	1.047 ± 0.001	1418.81
Toronto	1.047 ± 0.002	1104.54
Boston	1.049 ± 0.002	1261.84
Ankara	1.051 ± 0.003	575.70
Ho Chi Minh City	1.052 ± 0.004	502.37
Melbourne	1.056 ± 0.005	558.14
Mexico City	1.057 ± 0.002	929.95
Sao Paolo	1.059 ± 0.002	1224.01
Philadelphia	1.061 ± 0.002	1036.82
Berlin	1.068 ± 0.002	1343.56
New York City	1.073 ± 0.002	1429.50
Moscow	1.077 ± 0.001	1480.18
Madrid	1.078 ± 0.002	1030.24
Surabaya	1.082 ± 0.005	391.68
Chicago	1.083 ± 0.002	1143.91
Monterrey	1.084 ± 0.003	735.52
Nagoya	1.087 ± 0.001	1560.92
Johannesburg	1.087 ± 0.003	698.69
Lagos	1.089 ± 0.006	456.59
Kolkata	1.092 ± 0.004	823.74
Cape Town	1.092 ± 0.004	565.36
Shenzhen	1.096 ± 0.003	868.03
Osaka	1.097 ± 0.001	1700.57
Phoenix	1.097 ± 0.002	1119.66

Continued on next page

City	α	β
Hyderabad	1.097 ± 0.002	1323.22
Khartoum	1.100 ± 0.002	1262.22
Tokyo	1.101 ± 0.001	2698.20
Abidjan	1.101 ± 0.005	504.54
Alexandria	1.102 ± 0.002	1438.91
San Francisco	1.107 ± 0.003	1112.27
Mumbain	1.110 ± 0.005	662.04
Hong Kong	1.111 ± 0.003	921.48
Baghdad	1.113 ± 0.004	519.98
Washington DC	1.115 ± 0.002	1642.06
Yangon	1.117 ± 0.004	700.39
Rio De Janeiro	1.122 ± 0.003	1218.63
Istanbul	1.123 ± 0.002	1410.86
Jakarta	1.132 ± 0.003	672.21
Lima	1.139 ± 0.002	1236.50
Kuala Lumpur	1.140 ± 0.003	830.67
Delhi	1.150 ± 0.002	1393.89
Cairo	1.151 ± 0.003	957.97
Dubai	1.152 ± 0.003	783.79
Riyadh	1.205 ± 0.003	997.07
Manila	1.208 ± 0.002	2497.66

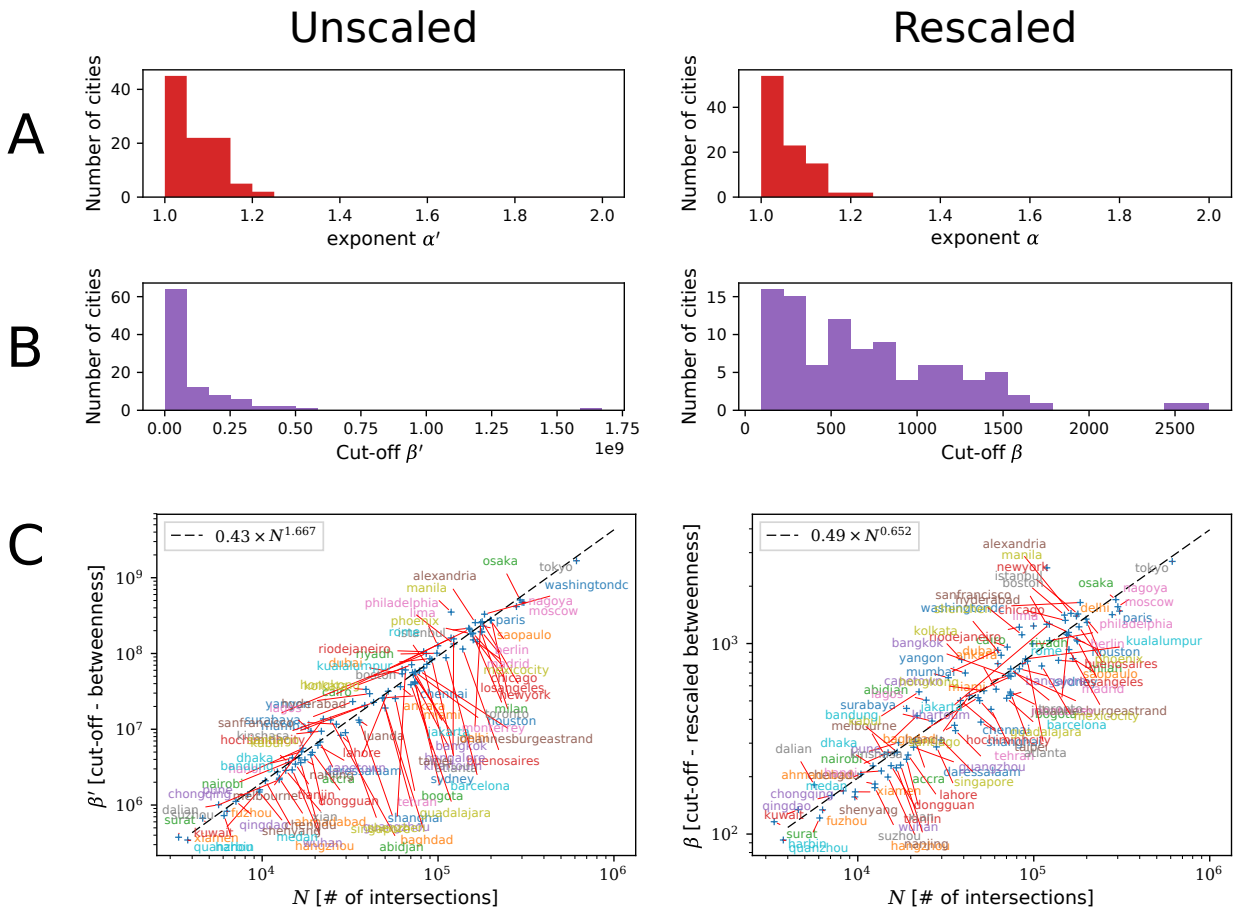


Figure S4: Distribution of the truncated-power-law parameters for both the unscaled (left panels) and rescaled BC (right panels). (A) As expected, the distribution of the α exponent is almost identical in both the rescaled and unscaled BC distributions. (B) The distribution of the exponential cut-offs, β, changes dramatically from the unscaled to the rescaled versions of the BC. (C) These cut-offs also show a marked dependence on system size (number of nodes)

Figure S4 shows the distribution of power law exponents α, α' and decay exponents β, β' obtained for the tails of the betweenness distributions of all cities studied, both for the regular and rescaled betweenness. Also shown are the exponents β, β' as a function of N.

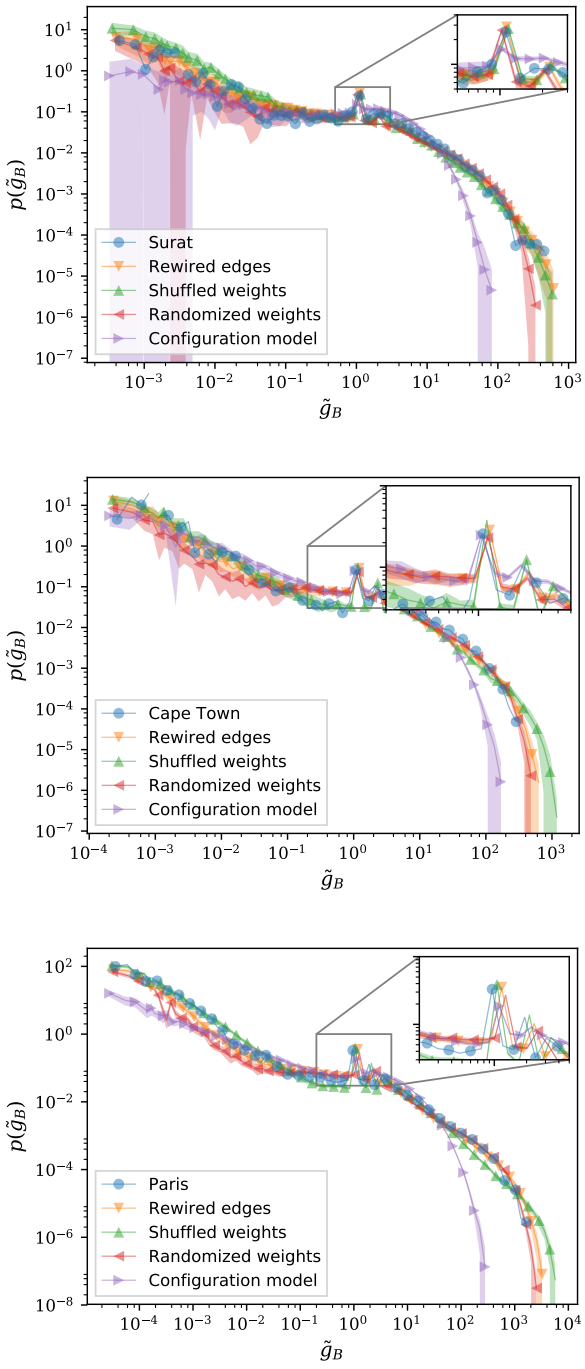


Figure S5: The BC distribution of various random graph models described in the main text compared to the baseline distribution of representative examples of cities of different sizes. Shaded area reflects fluctuations around the average over hundred realizations of each model.

Figure S5 shows the betweenness distributions for the randomized versions of cities at various size scales in the same manner as what was done for Phoenix in Figure 2 in the main manuscript. The similarity in the distributions seen in that figure is replicated in these plots, and the corresponding 2-sample KS statistics for each random graph model (along with the unweighted versions of these simulations) are shown in S6 and S7. To obtain the reported values, the KS statistics were obtained for the comparison of the actual Phoenix street network tail (nodes with betweenness above N) and the tails of each of the 100 realizations of the given random graph model, which were then averaged to get a single value. Although the comparisons are not statistically significant for random graph models of large cities, we do see statistical significance in Surat, a much smaller city of 300 nodes, as well as for random samples from larger cities of sizes up to ~ 500 nodes. The KS statistic for the comparison of the street network with its non-planar configuration model counterpart is more than double that of the next highest KS value, indicating that the constraint of planarity has a much stronger effect on the betweenness distribution than other structural perturbations.

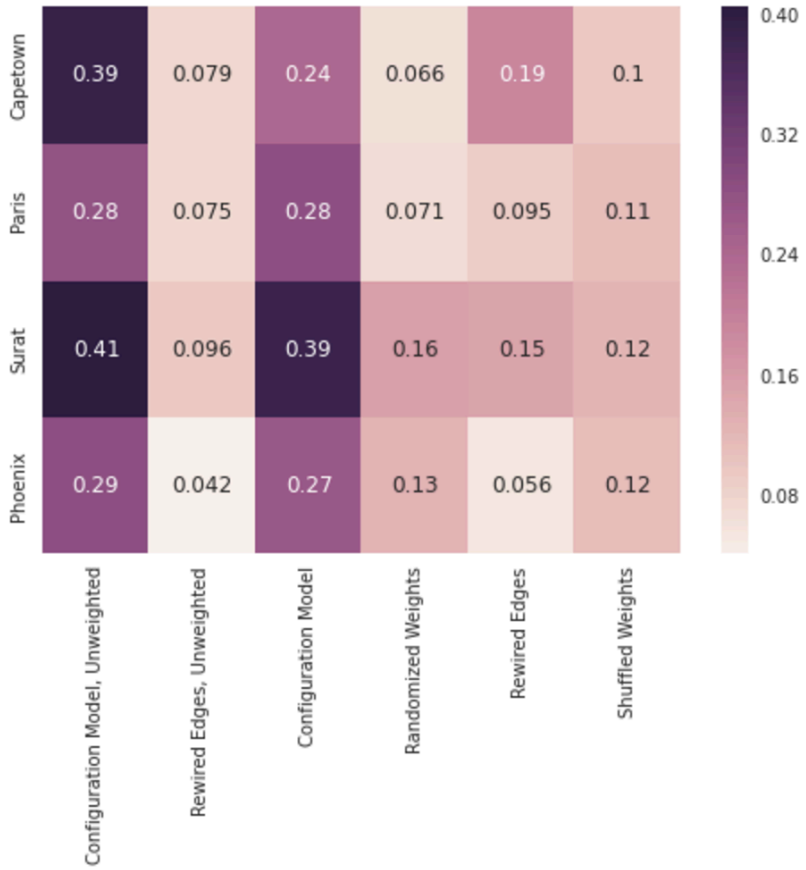
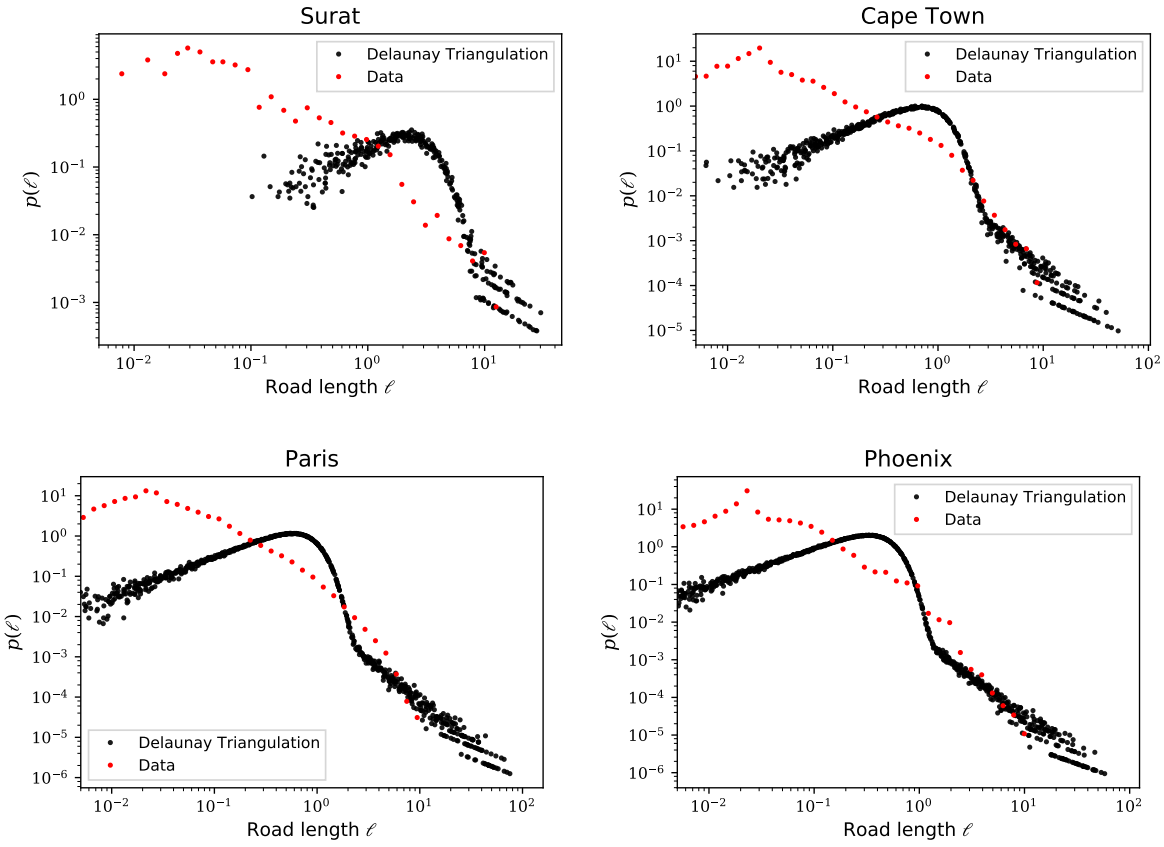


Figure S6: Average KS statistics over all 100 realizations of each random graph model when compared to the tail of the true betweenness distribution for cities at various size scales. In all cases, the non-planar configuration random graphs exhibit the most statistical dissimilarity from the original network in the tails of their betweenness distributions.



Figure S7: Average 2-sample KS p-values over all 100 realizations of each random graph model when compared to the tail of the true betweenness distribution for cities at various size scales.



ca

Figure S8: Distribution of road segment lengths (in km) for three selected cities of different sizes along with the length distribution of their corresponding Delaunay Triangulations.

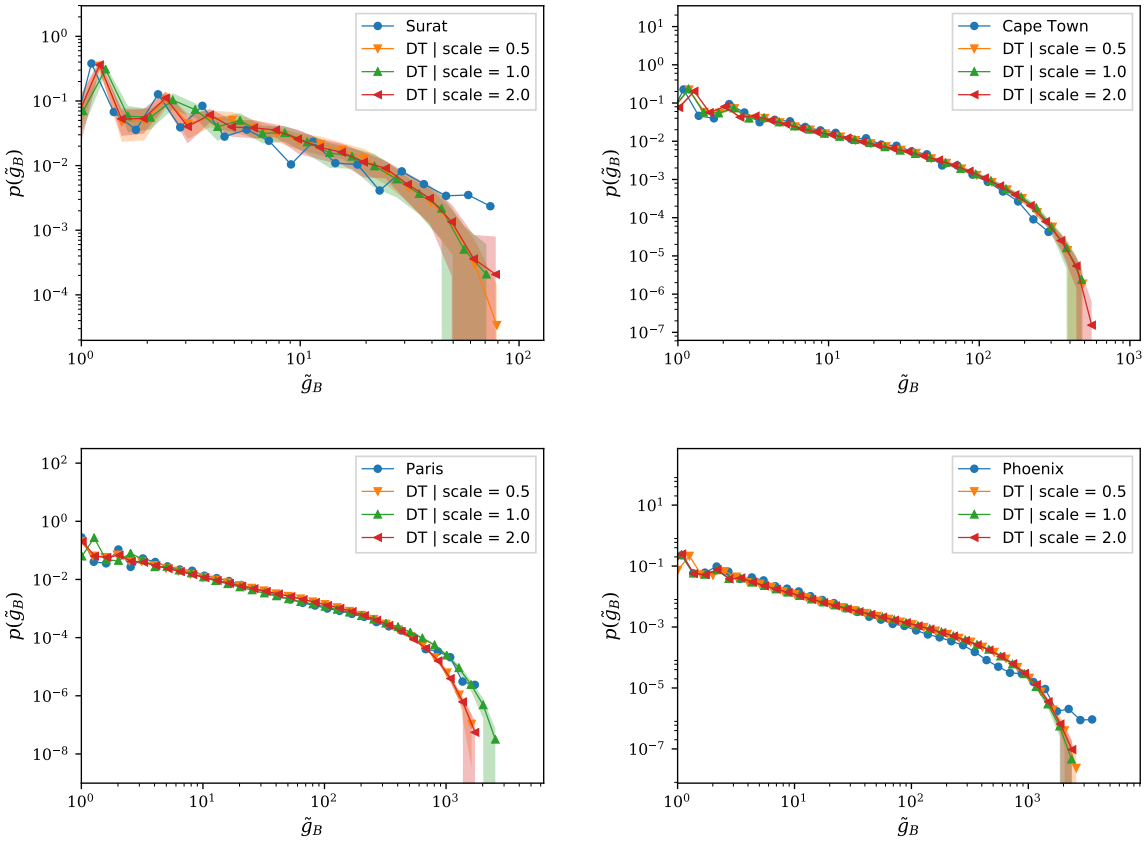


Figure S9: Tails of the BC distributions for selected cities and their corresponding DT for different grid-sizes, having the effect of changing the area and therefore the density of nodes N/A. Shown are the results for half and twice the original areas.

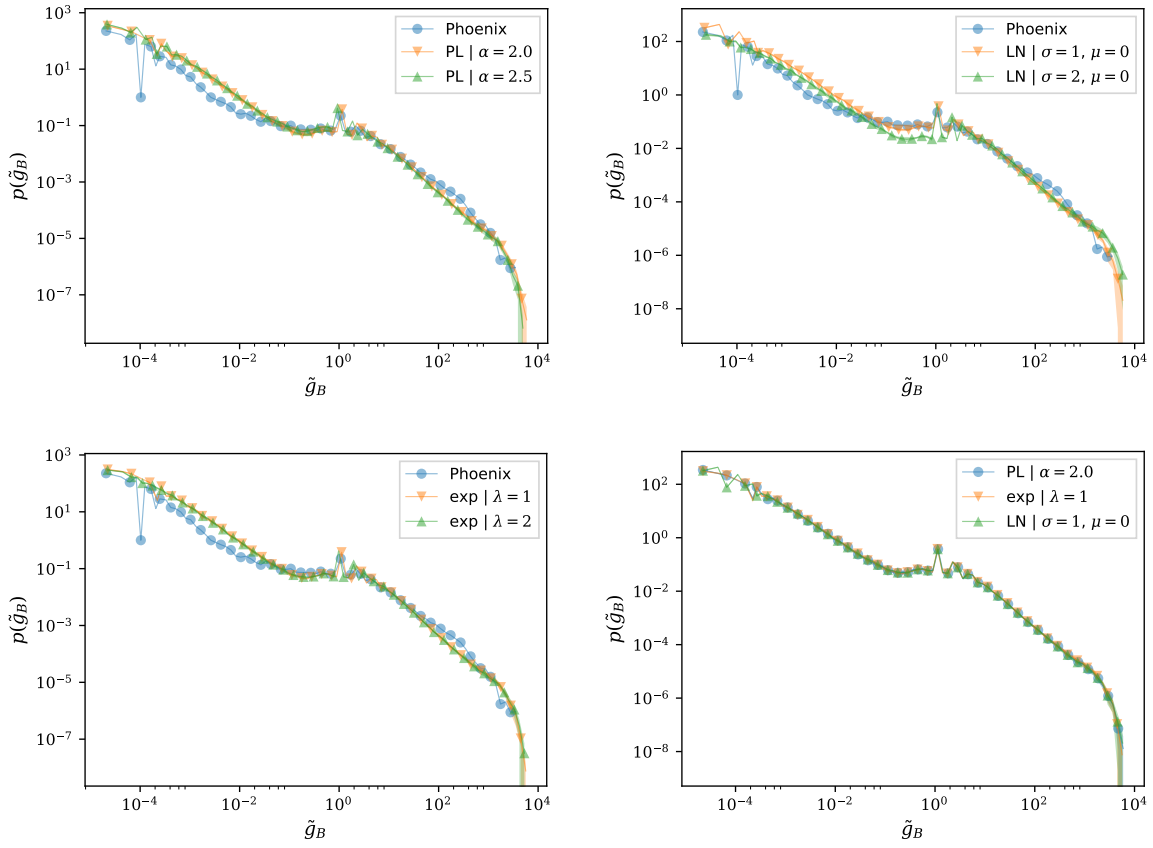


Figure S10: Distribution of $\tilde{g}_B$ for the Phoenix street network, with edge weights generated randomly from multiple families of distributions; power law (PL), exponential (exp) and log normal (LN). The bottom right plot shows three different weight distributions leading to an identical BC profile.

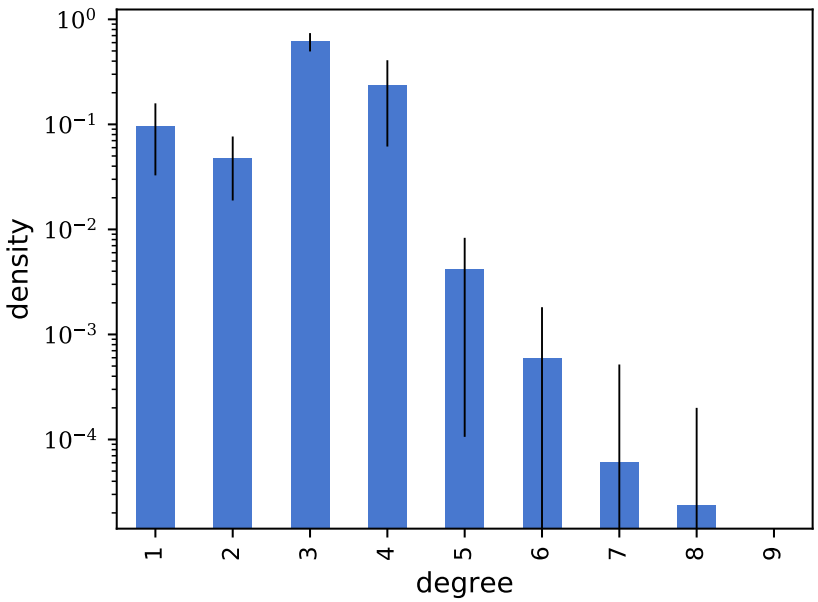


Figure S11: Degree distribution of streets networks averaged over different cities. The shaded-blue area corresponds to the standard deviation.

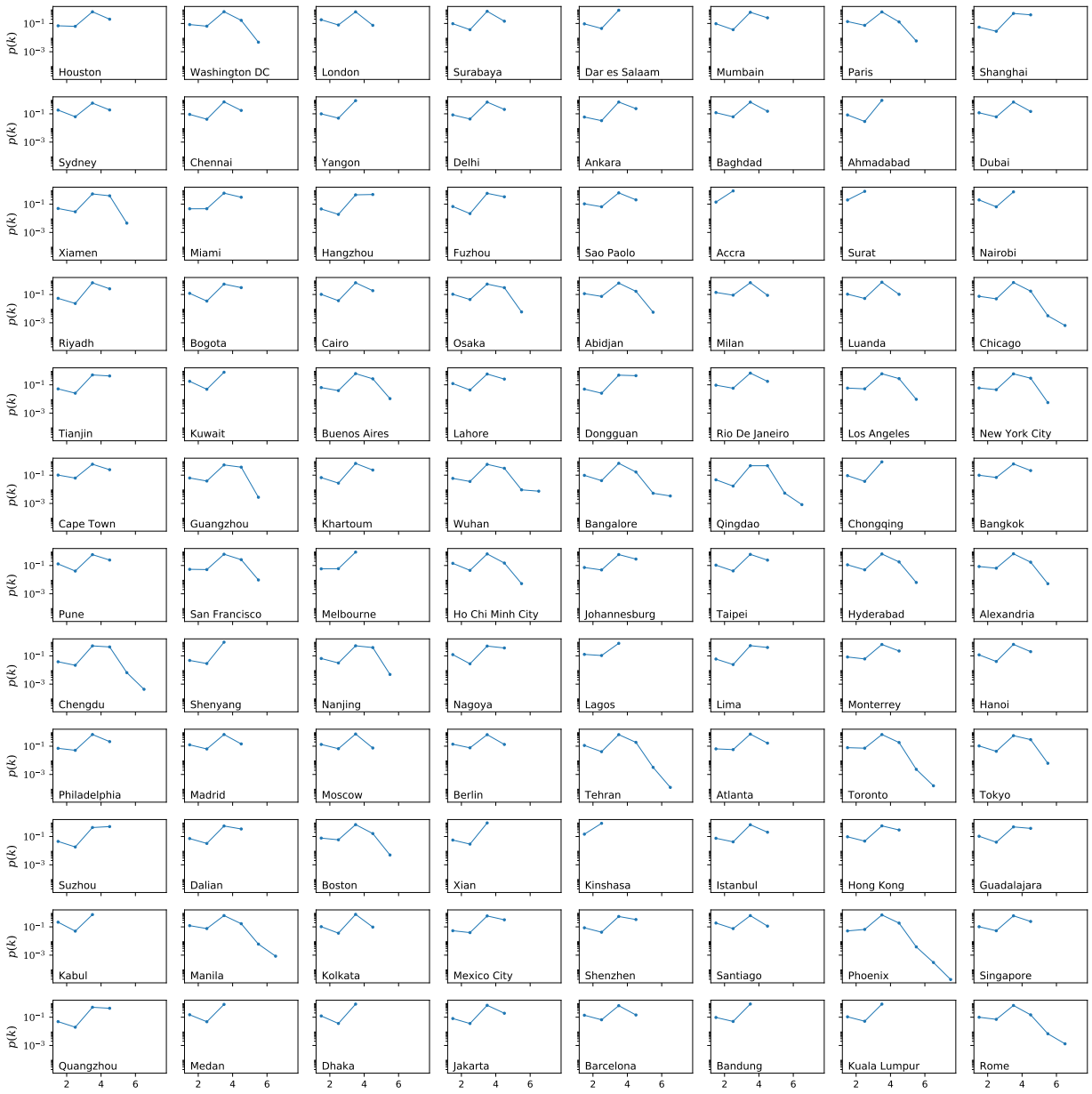


Figure S12: Degree distribution for each individual city.

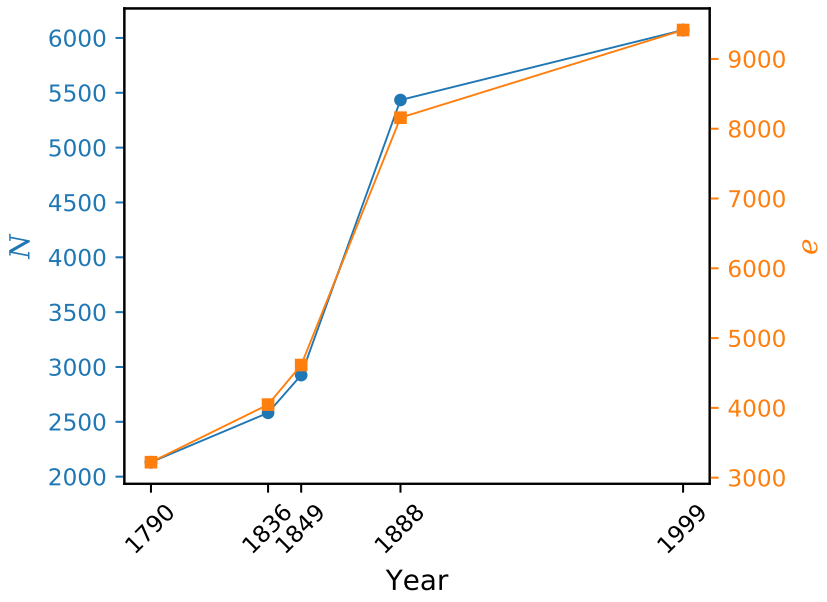


Figure S13: Evolution of the 1789 portion of the Paris street network over a period of approximately 200 years in terms of the number of nodes N and edges e. The edge-density is roughly constant, given that nodes and edges grow at the same rate.