Chapter 2
Irrelevant and Simple Measures

Many studies on complex networks were about how to characterize them and what are the most relevant measures for understanding their structure. In particular, the degree distribution and the existence of the second moment for an infinite network were shown to be critical when studying dynamical processes on networks. These behaviors are therefore strongly connected to degree fluctuations and the existence of hubs. In the case of spatial networks, the physical constraints are usually large and prevent the appearance of such hubs. These constraints also impact other quantities that are nontrivial for complex networks but that become irrelevant for spatial networks. We review here these measures that are essentially useless for spatial networks and we then discuss older, simple measures that were mostly introduced in the context of quantitative geography.

2.1 Irrelevant Measures

Quantities that depend very much on the spatial constraints turn out in general to be irrelevant for spatial networks. The prime example is the degree distribution which in many complex networks was found to be a broad law and in some cases well fitted by a power law of the form $P(k) \sim k^{-\gamma}$ with $1 < \gamma < 3$ [1]. In this case, degree fluctuations are very large which has a direct impact on many dynamical processes that take place on the network, such as epidemics for example [2]. For spatial networks, however, the degree has to satisfy steric constraints. If we consider the road network, nodes represent the intersections and the degree of a node is the number of streets starting from it and is therefore clearly limited as a result of space. As a consequence, the degree distribution for most spatial networks is not broad but displays a fast decaying tail (such as an exponential for example).

M. Barthelemy, *Morphogenesis of Spatial Networks*,
Lecture Notes in Morphogenesis, https://doi.org/10.1007/978-3-319-20565-6_2

Other irrelevant parameters include the clustering coefficient or the assortativity. Indeed as shown below the clustering coefficient is always large: if a node is connected to two other nodes in a spatial network, they are usually located in its neighborhood which in turn increases the probability that they are connected to each other, leading to a large clustering coefficient. In the following, we will discuss in more detail these different measures.

2.1.1 Degree

We recall here that a graph with N nodes and E edges can be described by its $N \times N$ adjacency matrix A which is defined as

$$A_{ij} = \begin{cases} = 1 \quad \text{if} \quad i \quad \text{and} \quad j \quad \text{are connected} \\ = 0 \quad \text{otherwise} \end{cases} \tag{2.1}$$

If the graph is undirected, then the matrix A is symmetric. The degree of a node is by definition the number of its neighbors and is given by

$$k_i = \sum_j A_{ij} \tag{2.2}$$

The first simple indicator of a graph is the average degree

$$\langle k \rangle = \frac{1}{N} \sum_i k_i = \frac{2E}{N} \tag{2.3}$$

where here and in the following the brackets $\langle \cdot \rangle$ denote the average over the nodes of the network. In particular, the scaling of $\langle k \rangle$ with N indicates if the network is sparse (which is the case when $\langle k \rangle \to const.$ for $N \to \infty$).

In [31, 46], measurements for street networks in different cities in the world are reported. Based on the data from these sources, the authors of [47] plotted (Fig. 2.1a) the number of roads E (edges) versus the number of intersections N. The plot is consistent with a linear fit with slope $\approx$1.44 (which is consistent with the value $\langle k \rangle \approx 2.5$ measured in [46]). The quantity $e = E/N = \langle k \rangle /2$ displays values in the range $1.05 < e < 1.69$, in between the values $e = 1$ and $e = 2$ that characterize tree-like structures and $2d$ regular lattices, respectively. Few exact values and bounds are available for the average degree of classical models of planar graphs. In general, it is known that $e \leq 3$, while it has been recently shown [48] that $e > 13/7$ for planar Erdös–Renyi graphs [48].

The distribution of degree $P(k)$ is usually a quantity of interest and can display large heterogeneities such as it is observed in scale-free networks (see for example [49]). We indeed observe that for spatial networks such as airline networks or the

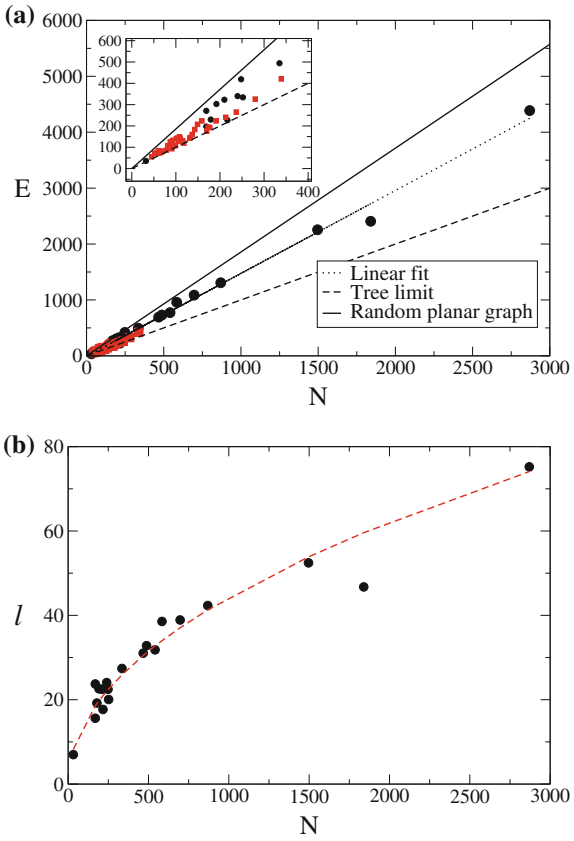

Fig. 2.1 **a** Numbers of roads versus the number of nodes (i.e., intersections and centers) for data from [31] (circles) and from [46] (squares). In the inset, we show a zoom for a small number of nodes. **b** Total length versus the number of nodes. The line is a fit which predicts a growth as $\sqrt{N}$ (data from [31] and figures from [47])

Internet, the degrees are very heterogeneous (see [2]). However, when physical constraints are strong or when the cost associated with the creation of new links is large, a cutoff appears in the degree distribution [8] and in some case the distribution can be very peaked. This is the case for the road network for example, and more generally in the case of planar networks for which the degree distribution $P(k)$ is of little interest. For example, in a study of 20 German cities, Lämmer et al. [29] showed that most nodes have four neighbors (the full degree distribution is shown in Fig. 2.2a) and that the degree rarely exceeds 5 for various world cities [31]. These values are, however, not very indicative: planarity imposes severe constraints on the degree of a node and on its distribution which is generally peaked around its average value.

We note here that in real-world cases such as the road network for example, it is natural to study the usual (or "primal") representation where the nodes are the intersections and the links represent the road segment between the intersection.

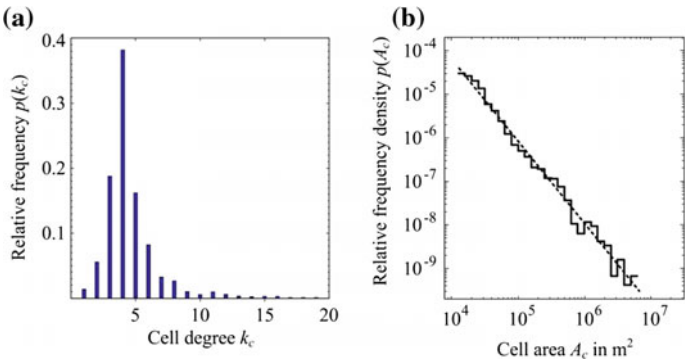

Fig. 2.2 **a** Degree distribution of degrees for the road network of Dresden. **b** The frequency distribution of the cells surface areas A_c obeys a power law with exponent $\alpha \approx 1.9$ (for the road network of Dresden). Figure taken from [29]

However, in another representation, the dual graph can be of interest (see [27]) and for the road network it is constructed in the following way: the nodes are the roads and two nodes are connected if there exists an intersection between the two corresponding roads. One can then measure the degree of a node which represents the number of roads which intersect a given road. Also, the shortest path length in this network represents the number of different roads one has to take to go from one point to another. Even if the road network has a peaked degree distribution, its dual representation can display broad distributions [50]. Indeed, in [50], measurements were made on the dual network for the road network in the US, England, and Denmark and showed large fluctuations with a power-law distribution with exponent $2.0 < \gamma < 2.5$.

2.1.2 Length of Segments

In Fig. 2.1b, we plot the total length ℓ_T of the network versus N for the cities considered in [31]. Data are well fitted by a power function of the form

$$\ell_T = \mu N^{\beta} \tag{2.4}$$

with $\mu \approx 1.51$ and $\beta \approx 0.49$. In order to understand this result, one has to focus on the street segment length distribution $P(\ell_1)$. This quantity has been measured for London in [34] and is shown in Fig. 2.3. This figure shows that the distribution decreases rapidly and the fit proposed by the authors of [34] suggests that

$$P(\ell_1) \sim \ell_1^{-\gamma} \tag{2.5}$$

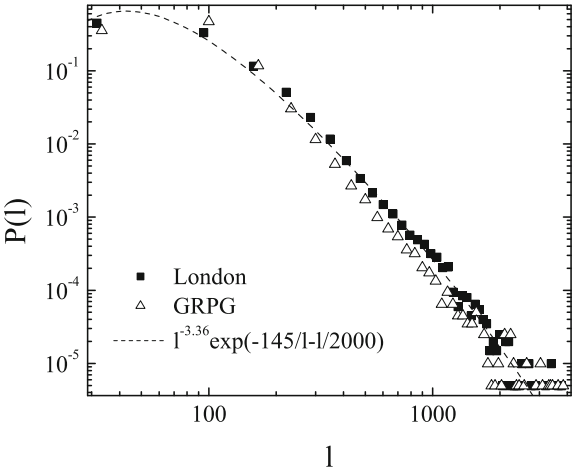

Fig. 2.3 Length distribution $P(\ell_1)$ for the street network of London (and for the model GRPG proposed in [34]). Figure taken from [34]

with $\gamma \simeq 3.4$ which implies that both the average and the dispersion are well defined and finite. If we assume that this result extends to other cities, it means that we have a typical distance ℓ_1 between nodes which is meaningful. This typical distance between connected nodes then naturally scales as

$$\ell_1 \sim \frac{1}{\sqrt{\rho}} \tag{2.6}$$

where $\rho = N/L^2$ is the density of vertices and L is the linear dimension of the ambient space. This implies that the total length scales as

$$\ell_T \sim E\ell_1 \sim \frac{\langle k \rangle}{2} L\sqrt{N} \tag{2.7}$$

This simple argument reproduces well the $\sqrt{N}$ behavior observed in Fig. 2.1b and also the value (given the error bars) of the prefactor $\mu \approx \langle k \rangle L/2$.

2.1.3 Clustering, Assortativity, and Average Shortest Path

Complex networks are essentially characterized by a small set of parameters which are not all relevant for spatial networks. For example, the degree distribution which has been the main subject of interest in complex network studies is usually peaked for planar networks, due to the spatial constraints, and is therefore not very interesting. In the following we will discuss the effect of spatial constraints on other important parameters.

2.1.3.1 Clustering Coefficient

The clustering coefficient of a node i of degree k_i is defined as

$$C(i) = \frac{E_i}{k_i(k_i - 1)/2} \tag{2.8}$$

where E_i is the number of edges among the neighbors of i. This quantity gives some information about local clustering and was the object of many studies in complex networks. For the Erdos–Renyi (ER) random graphs with finite average degree $\langle k \rangle$, the average clustering coefficient is simply given by

$$\langle C \rangle = p \sim \frac{\langle k \rangle}{N} \tag{2.9}$$

where the brackets $\langle \cdot \rangle$ denote the average over the network (p is the probability to connect two nodes). In contrast, for spatial networks, closer nodes have a larger probability to be connected, leading to a large clustering coefficient. The variation of this clustering coefficient in space can thus bring valuable information about the spatial structure of the network under consideration. The clustering coefficient depends on the number of triangles or cycles of length 3 and can also be computed by using the adjacency matrix A. Powers of the adjacency matrix give the number of paths of variable length. For instance, the quantity $\frac{1}{6}\mathrm{Tr}(\mathrm{A}^3)$ is the number C_3 of cycles of length tree and is related to the clustering coefficient. Analogously, we can define and count cycles of various lengths (see for example [51, 52] and references therein) and compare this number to the ones obtained on null models (lattices, triangulations, etc.).

Finally, many studies define the clustering coefficient per degree classes which is given by

$$C(k) = \frac{1}{N(k)} \sum_{i/k_i=k} C(i) \tag{2.10}$$

The behavior of $C(k)$ versus k thus gives an indication on how the clustering is organized when we explore different classes of degrees. However, in order to be useful, this quantity needs to be applied to networks with a large range of degree variations which is usually not the case in spatial networks.

The average clustering coefficient can be calculated for the random geometric graph (see also Chap. 9) and we discuss in the following the argument presented in [53]. If two vertices i and j are connected to a vertex k, it means that they are both in the excluded volume of k. Now, these vertices i and j are connected only if j is in the excluded volume of i. Putting all pieces together, the probability to have two connected neighbors (ij) of a node k is given by the fraction of the excluded volume of i which lies within the excluded volume of k. By averaging over all points i in the excluded volume of k, we then obtain the average clustering coefficient.

We thus have to compute the volume overlap ρ_d of two spheres which for spherical symmetry reasons depends only on the distance between the two spheres. In terms of this function, the clustering coefficient is given by

$$\langle C_d \rangle = \frac{1}{V_e} \int_{V_e} \rho_d(r) dV \tag{2.11}$$

For $d = 1$, we have

$$\rho_1(r) = (2R - r)/2R = 1 - r/2R \tag{2.12}$$

and we obtain

$$\langle C_1 \rangle = 3/4 \tag{2.13}$$

For $d = 2$, we have to determine the area overlapping in Fig. 2.4 which gives

$$\rho_2(r) = (\theta(r) - \sin(\theta(r)))/\pi \tag{2.14}$$

with $\theta(r) = 2 \arccos(r/2R)$ and leads to

$$\langle C_2 \rangle = 1 - 3\sqrt{3}/4\pi \approx 0.58650 \tag{2.15}$$

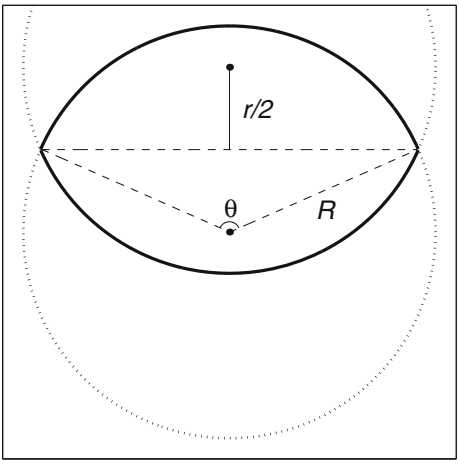

Fig. 2.4 The overlap between the two disks (area comprised within the bold line) gives the quantity $\rho_2(r)$. Figure taken from [53]

Similarly, an expression can be derived in d dimension [53] which for large d reduces to

$$\langle C_d \rangle \sim 3\sqrt{\frac{2}{\pi d}} \left(\frac{3}{4}\right)^{\frac{d+1}{2}} \tag{2.16}$$

The average clustering coefficient thus decreases from the value $3/4$ for $d = 1$ to values of order 10^{-1} for d of order 10 and is independent from the number of nodes which is in sharp contrast with ER graphs for which $\langle C \rangle \sim 1/N$. Random geometric graphs are thus much more clustered than random ER graphs. The main reason—which is in fact valid for most spatial graphs—is that long links are prohibited or rare. This fact implies that if both i and j are connected to k, it means that there are in some spatial neighborhood of k which increases the probability that their inter-distance is small too, leading to a large $\langle C \rangle$.

2.1.3.2 Assortativity

In general, the degrees of the two end nodes of a link are correlated and to describe these degree correlations one needs the two-point correlation function $P(k'|k)$. This quantity represents the probability that any edge starting at a vertex of degree k ends at a vertex of degree k'. Higher order correlation functions can be defined and we refer the interested reader to [54] for example. The function $P(k'|k)$ is, however, not easy to handle and one can define the assortativity [55, 56]

$$k_{nn}(k) = \sum_{k'} P(k'|k)k' \tag{2.17}$$

A similar quantity can be defined for each node as the average degree of the neighbor

$$k_{nn}(i) = \frac{1}{k_i} \sum_{j \in \Gamma(i)} k_j \tag{2.18}$$

where $\Gamma(i)$ denotes the set of neighbors of node i. There are essentially two classes of behaviors for the assortativity. If $k_{nn}(k)$ is an increasing function of k, vertices with large degrees have a larger probability to connect to similar nodes with a large degree. In this case, we speak of an *assortative* network and in the opposite case of a *disassortative* network. It is expected in general that social networks are mostly assortative, while technological networks are disassortative. However, for spatial networks spatial constraint usually implies a flat function $k_{nn}(k)$, since it is usually the distance that governs the existence of a link and not the degree.

2.1.3.3 Average Shortest Path

Usually, there are many paths between two nodes in connected networks and if we keep the shortest one it defines a distance on the network

$$\ell(i, j) = \min_{paths(i \to j)} |path| \tag{2.19}$$

where the length $|path|$ of the path is defined as its number of edges. The *diameter* of the graph can be defined as the maximum value of all $\ell(i, j)$ or can also be estimated by the average of this distance over all pairs of nodes in order to characterize the "size" of the network. For a d-dimensional regular lattice with N nodes, this average shortest path $\langle \ell \rangle$ scales as

$$\langle \ell \rangle \sim N^{1/d} \tag{2.20}$$

In a small-world network (see [7] and Chap. 10) constructed over a $d-$dimensional lattice $\langle \ell \rangle$ has a very different behavior

$$\langle \ell \rangle \sim \log N \tag{2.21}$$

The crossover from a large-world behavior $N^{1/d}$ to a small-world one with $\log N$ can be achieved for a density p of long links (or "shortcuts") [57] such that

$$pN \sim 1 \tag{2.22}$$

The effect of space could thus in principle be detected in the behavior of $\langle \ell \rangle(N)$. It should, however, be noted that if the number of nodes is too small this can be a tricky task. In the case of brain networks, for example, a behavior of a typical three-dimensional network in $N^{1/3}$ could easily be confused with a logarithmic behavior if N is not large enough.

2.1.4 Empirical Illustrations

We discuss here some simple results obtained on transportation networks that illustrate the fact that indeed some measures that are useful for understanding complex networks are actually irrelevant in the case of spatial networks and do not convey interesting information.

2.1.4.1 Power Grids and Water Distribution Networks

Power grids are one of the most important infrastructures in our society. In modern countries, they have evolved for a rather long time (sometimes a century) and are now complex systems with a large variety of elements and actors playing in their functioning. This complexity leads to the relatively unexpected result that their robustness is actually not very well understood and large blackouts such as the huge August 2003 blackout in North America demonstrates the fragility of these systems.

The topological structure of these networks was studied in different papers such as [8, 58, 59]. In particular, in [8, 58], the authors consider the Southern Californian and the North American power grids. In these networks, the nodes represent the power plants, distribution, and transmission substations, and the edges correspond to transmission lines. These networks are typically planar (see for example the Italian case, Fig. 2.5) and we expect a peaked degree distribution, decreasing typically as an exponential of the form $P(k) \sim \exp(-k/\langle k \rangle)$ with $\langle k \rangle$ of order 3 in Europe and 2 in the US. The other studies on US power grids confirm that the degree distribution is exponential (see Fig. 2.6). In [58], Albert, Albert, and Nakarado also studied the load (a quantity similar to the betweenness centrality) and found a broad distribution. The degree being peaked, we can then expect very large fluctuations of load for the same value of the degree, as expected in general for spatial networks. These authors also found a large redundance in this network with, however, 15% of cut edges.

Also, as expected for these networks, the clustering coefficient is rather large and even independent of k as shown in the case of the power grid of Western US (see Fig. 2.7).

Besides the distribution of electricity, our modern societies also rely on various other distribution networks. The resilience of these networks to perturbations is thus an important point in the design and operating of these systems. In [61], Yazdani and Jeffrey study the topological properties of the Colorado Springs Utilities and the Richmond (UK) water distribution networks (shown in Fig. 2.8). Both these networks (of size $N = 1786$ and $N = 872$, respectively) are sparse planar graphs with very peaked degree distributions (the maximum degree is 12).

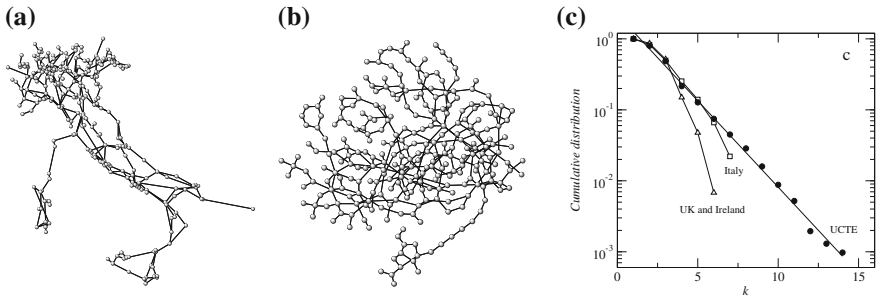

Fig. 2.5 **a** Map of the Italian power grid. **b** Topology of the Italian power grid. **c** Degree distribution for the European network (UCTE), Italy, the UK, and Ireland. In all cases, the degree distribution is peaked and can be fitted by exponential. Figure taken from [59]

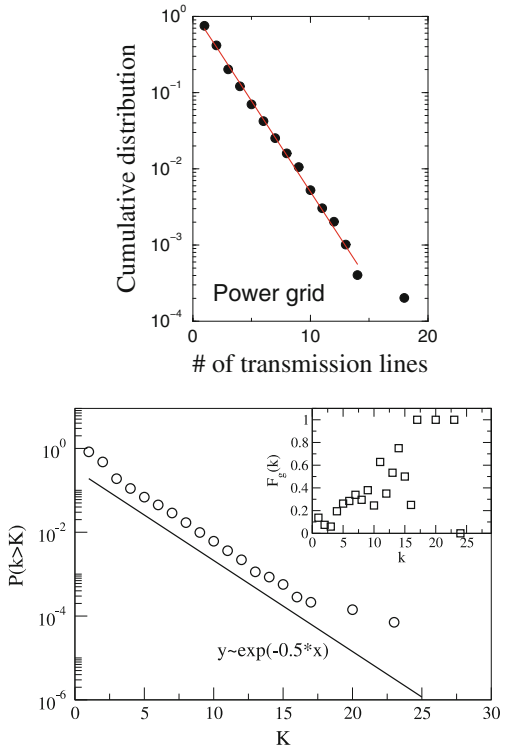

Fig. 2.6 Degree distribution of substations in Southern California (top panel) and for the North American power grid (bottom panel). In both cases, the lines represent an exponential fit. Figure taken from [8, 58], respectively

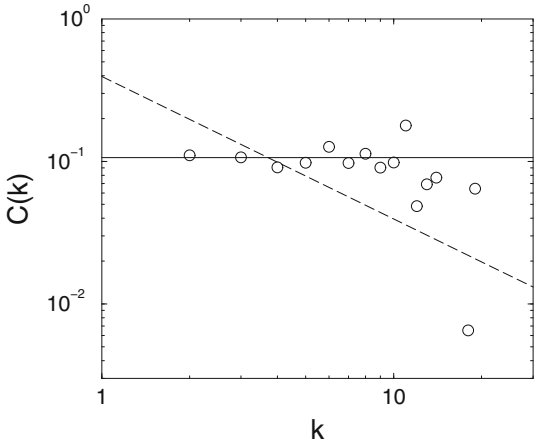

Fig. 2.7 Scaling of the clustering $C(k)$ for the power grid of the Western United States. The dashed line has a slope -1 and the solid line corresponds to the average clustering coefficient. Figure taken from [60]

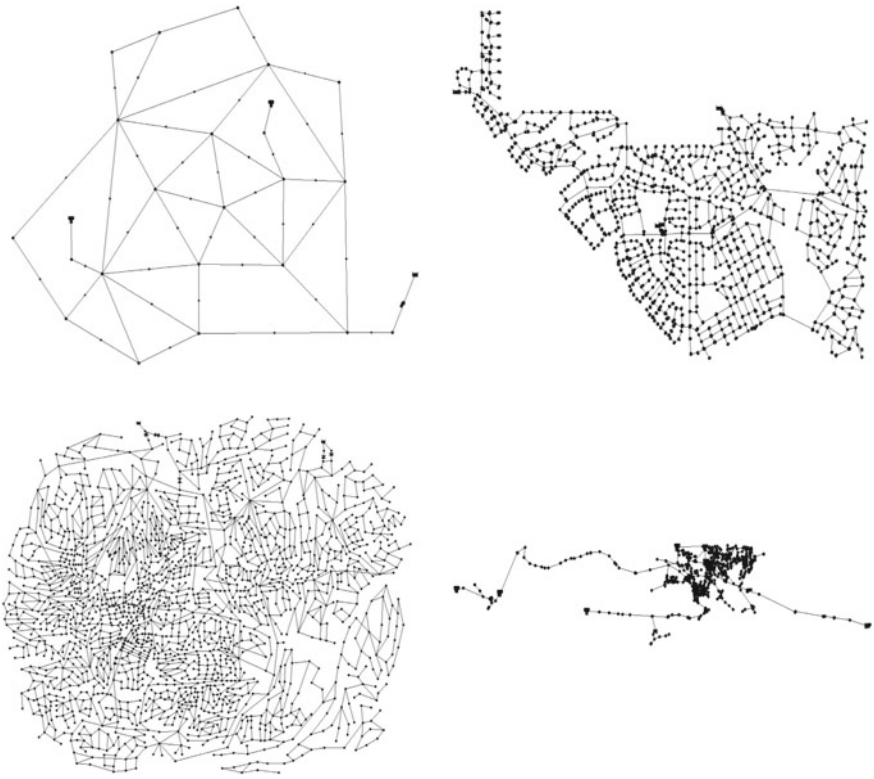

Fig. 2.8 Representation of water distribution networks. Left panels (from top to bottom): Synthetic networks ("Anytown" [62], and "EXNET" [63]). Top-right panel: Colorado Spring Utilities network. Bottom-right panel: Richmond (UK) water distribution network. Figure taken from [61]

2.1.4.2 Subways and Buses

One of the first studies (after the Watts–Strogatz paper) on the topology of a transportation network was proposed by Latora and Marchiori [64] who considered the Boston subway network. It is a relatively small network with $N = 124$ stations. The average shortest path is $\langle \ell \rangle \sim 16$ a value which is large compared to $\ln 124 \approx 5$ and closer to the two-dimensional result $\sqrt{124} \approx 11$.

In [15], Sienkiewicz and Holyst study a larger set made of public transportation networks of buses and tramways for 22 Polish cities and in [65], von Ferber et al. study the public transportation networks for 15 world cities. The number of nodes of these networks varies from $N = 152$ to 2811 in [15] and in the range [1494, 44629] in [65]. Interestingly enough, the authors of [15] observe a strong correlation between the number of stations and the population which is not the case for the world cities studied in [65] where the number of stations seems to be independent from the population (see Sect. 14.3 for a detailed discussion about the connection between

socioeconomical indicators and the properties of networks). For polish cities, the degree has an average in the range [2.48, 3.08] and in a similar range [2.18, 3.73] for [65]. In both cases, the degree distribution is relatively peaked (the range of variation is usually of the order of one decade) consistently with the existence of physical constraints [8].

Due to the relatively small range of variation of N in these various studies [15, 64, 65], the behavior of the average shortest path is not clear and could be fitted by a logarithm or a power law as well. We can, however, note that the average shortest path is usually large (of order 10 in [15] and in the range [6.4, 52.0] in [65]) compared to $\ln N$, suggesting that the behavior of $\langle \ell \rangle$ might not be logarithmic with N but more likely scales as $N^{1/2}$, a behavior typical of a two-dimensional lattice.

The average clustering coefficient $\langle C \rangle$ in [15] varies in the range [0.055, 0.161] and is larger than a value of the order $C_{ER} \sim 1/N \sim 10^{-3} - 10^{-2}$ corresponding to a random ER graph. The ratio $\langle C \rangle / C_{ER}$ is explicitly considered in [65] and is usually much larger than one (in the range [41, 625]). The degree-dependent clustering coefficient $C(k)$ seems to present a power-law dependence, but the fit is obtained over one decade only.

In another study [66], the authors study two urban train networks (Boston and Vienna which are both small $N = 124$ and $N = 76$, respectively) and their results are consistent with the previous ones.

2.1.4.3 Railways

One of the first studies of the structure of railway network [67] concerns a subset of the most important stations and lines of the Indian railway network and has $N = 587$ stations. In the P-space representation (see Chap. 1), there is a link between two stations if there is a train connecting them and in this representation, the average shortest path is of order $\langle \ell \rangle \approx 5$ which indicates that one needs four connections in the worst case to go from one node to another one. In order to obtain variations with the number of nodes, the authors considered different subgraphs with different sizes N. The clustering coefficient varies slowly with N that is always larger than $\approx$0.7 which is much larger than a random graph value of order $1/N$. Finally, in this study [67], it is shown that the degree distribution is behaving as an exponential and that the assortativity $\langle k_{nn} \rangle$(k) is flat showing an absence of correlations between the degree of a node and those of its neighbors.

In [13], Kurant and Thiran studied the railway system of Switzerland and major trains and stations in Europe (and also the public transportation system of Warsaw, Poland). The Swiss railway network contains $N = 1613$ nodes and $E = 1680$ edges (Fig. 2.9). All conclusions drawn here are consistent with the various cases presented in this chapter. In particular, the average degree is $\langle k \rangle \approx 2.1$, the average shortest path is $\approx$47 (consistent with the $\sqrt{N}$ result for a two-dimensional lattice), the clustering coefficient is much larger than its random counterpart, and the degree distribution is peaked (exponentially decreasing).

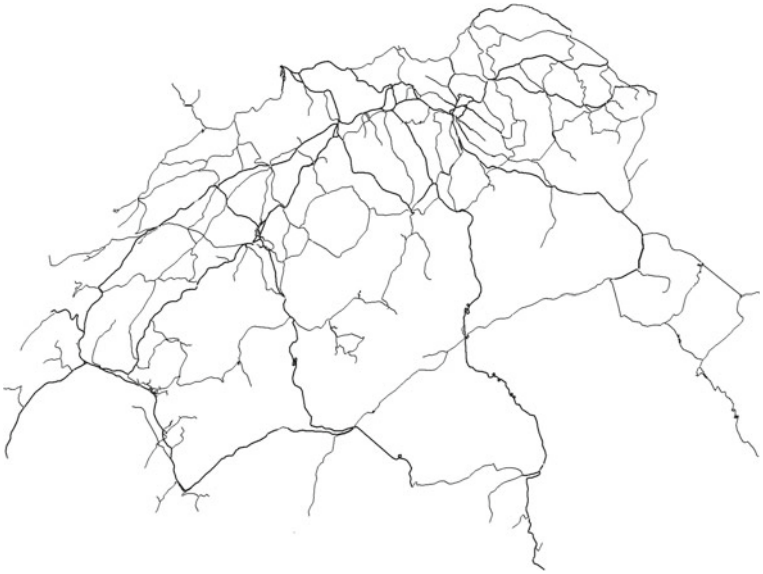

Fig. 2.9 Physical map of the Swiss railway networks. Figure taken from [13]

2.1.4.4 Neural Networks

The human brain with about 10^{10} neurons and about 10^{14} connections is one of the most complex networks that we know. The structure and functions of the brain are the subjects of numerous studies and different recent techniques such as electroencephalography, magnetoencephalography, functional RMI, etc. can be used in order to reconstruct networks for the human brain (see Fig. 2.10 and for a clear and nice introduction see for example [68, 69]).

Brain regions that are spatially close have a larger probability of being connected than remote regions as longer axons are more costly in terms of material and energy [68]. Wiring costs depending on distance are thus certainly an important aspect of brain networks and we can expect spatial networks to be relevant in this rapidly evolving topic. So far, many measures seem to confirm a large value of the clustering coefficient, and a small-world behavior with a small average shortest path length [70, 71]. It also seems that neural networks do not optimize the total wiring length but rather the processing paths, thanks to shortcuts [72]. This small-world structure of neural networks could reflect a balance between local processing and global integration with rapid synchronization, information transfer, and resilience to damage [73].

In contrast, the nature of the degree distribution is still under debate and a recent study on the macaque brain [74] showed that the distribution is better fitted by an exponential rather than by a broad distribution. Besides the degree distribution, most of the observed features were confirmed in latest studies such as [75] where Zalesky

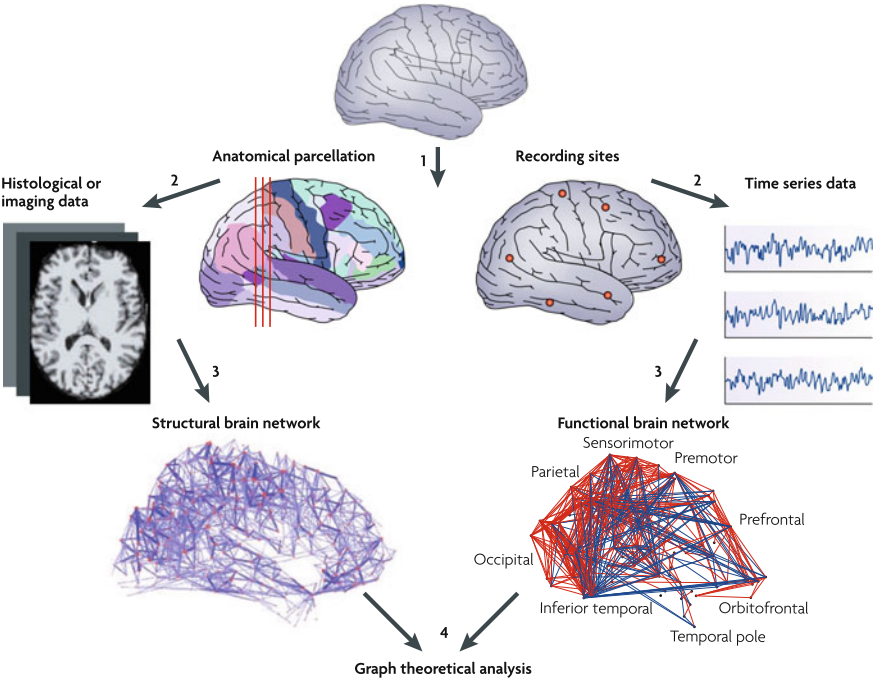

Fig. 2.10 Structural and functional brains can be studied with graph theory by following different methods shown step-by-step in this figure. Figure taken from [68]

et al. propose to construct the network with MRI techniques where the nodes are distinct gray-matter regions and links represent the white-matter fiber bundles. The spatial resolution is of course crucial here and the largest network obtained here is of size $N \approx 4{,}000$. These authors find large clustering coefficients with a ratio to the corresponding random graph value of order 10^2. Results for the average shortest path length $\langle \ell \rangle$ are, however, not so clear due to relatively low values of N. Indeed, for N varying from 1,000 to 4,000, $\langle \ell \rangle$ varies by a factor of order 1.7−1.8 [75]. A small-world logarithmic behavior would predict a ratio

$$r = \frac{\langle \ell \rangle (N = 4000)}{\langle \ell \rangle (N = 1000)} \sim \frac{\log(4000)}{\log(1000)} \approx 1.20 \tag{2.23}$$

while a three-dimensional spatial behavior would give a ratio of order $r \approx 4^{1/3} \approx 1.6$ which is closer to the observed value. Larger sets would, however, be needed in order to be sure about the behavior of this network concerning the average shortest path and to distinguish a $\log N$ from a $N^{1/3}$ behavior expected for a three-dimensional lattice.

Things are, however, more complex than it seems and even if functional connectivity correlates well with anatomical connectivity at an aggregate level, a recent study [76] shows that strong functional connections exist between regions with no direct structural connections, demonstrating that structural and functional properties of neural networks are entangled in a complex way and that future studies are needed in order to understand this extremely complex system.

2.2 Simple Measures

2.2.1 Topological Indices: α and γ Indices

Different indices were defined a long time ago mainly by scientists working in quantitative geography since the 1960s and can be found in [3, 77, 78] (see also the more recent paper by Xie and Levinson [32]). Most of these indices are relatively simple but give valuable information about the structure of the network, in particular if we are interested in planar networks. They were used to characterize the topology of transportation networks: Garrison [79] measured some properties of the Interstate highway system and Kansky [80] proposed up to 14 indices to characterize these networks. The simplest index is called the gamma index and is defined by

$$\gamma = \frac{E}{E_{max}} \tag{2.24}$$

where E is the number of edges and E_{max} is the maximal number of edges (for a given number of nodes N). For nonplanar networks, E_{max} is given by $N(N-1)/2$ for nondirected graphs and for planar graphs we saw in Chap. 1 that $E_{max} = 3N - 6$ leading to

$$\gamma_P = \frac{E}{3N - 6} \tag{2.25}$$

The gamma index is a simple measure of the density of the network but one can define a similar quantity by counting the number of elementary cycles instead of edges. The number of elementary cycles for a network is known as the cyclomatic number (see for example [17]) and is equal to

$$\Gamma = E - N + 1 \tag{2.26}$$

For a planar graph, this number is always less or equal to $2N-5$ which leads naturally to the definition of the alpha index (also coined "meshedness" in [46])

$$\alpha = \frac{E - N + 1}{2N - 5} \tag{2.27}$$

This index lies in the interval [0, 1] and is equal to 0 for a tree and equal to 1 for a maximal planar graph. Using the definition of the average degree $\langle k \rangle = 2E/N$, the quantity α reads in the large N limit as

$$\alpha \simeq \frac{\langle k \rangle - 2}{4} \tag{2.28}$$

which shows that in fact for a large network this index α does not contain much more information than the average degree.

2.2.2 Organic Ratio and Ringness

We note that more recently other interesting indices were proposed in order to characterize specifically road networks [32, 81]. For example, in some cities, the degree distribution is very peaked around 3−4 and the ratio

$$r_N = \frac{N(1) + N(3)}{\sum_{k \neq 2} N(k)} \tag{2.29}$$

can be defined [81] where $N(k)$ is the number of nodes of degree k. If this ratio is small, the number of dead ends and of "unfinished" crossing ($k = 3$) is small compared to the number of regular crossings with $k = 4$ which signals a more organized city. In the opposite case of large r_N (i.e., close to 1), there is a dominance of $k = 1$ and $k = 3$ nodes, which is the sign of a mode "organic" city.

The authors of [81] also define the "compactness" of a city which measures how much a city is "filled" with roads. If we denote by A the area of a city and by ℓ_T the total length of roads, the compactness $\Psi \in [0, 1]$ can be defined in terms of the hull and city areas

$$\Psi = 1 - \frac{4A}{(\ell_T - 2\sqrt{A})^2} \tag{2.30}$$

In the extreme case of one square city of linear size $L = \sqrt{A}$ with only one road encircling it, the total length is $\ell_T = 4\sqrt{A}$ and the compactness is then $\Psi = 0$. At the other extreme, if the city roads constitute a square grid of spacing a, the total length is $\ell_T = 2L^2/a$ and in the limit of $a/L \to 0$ one has a very large compactness $\Psi \approx 1 - a^2/L^2$.

We end this section by mentioning the ringness. Arterial roads (including freeways, major highways) provide a high level of mobility and serve as the backbone of the road system [32]. Different measures (along with many references) are discussed and defined in this paper [32], and in particular, the ringness is defined as

$$\phi_{ring} = \frac{\ell_{ring}}{\ell_{tot}} \tag{2.31}$$

where ℓ_{ring} is the total length of arterials on rings, and the denominator ℓ_{tot} is the total length of all arterials. This quantity ranging from 0 to 1 is thus an indication of the importance of a ring and to what extent arterials are organized as trees.

2.2.3 Cell Areas and Shape

Planar graphs naturally produce a set of nonoverlapping cells (or faces, or blocks) and covering the embedding plane. In the case of the road network, the distribution of the area A of these cells has been measured for the city of Dresden in Germany (Fig. 2.2b) and has the form

$$P(A) \sim A^{-\alpha} \tag{2.32}$$

with $\alpha \simeq 1.9$, which was confirmed by measures on other cities [11]. This broad law is in sharp contrast with the simple picture of an almost regular lattice which would predict a distribution $P(A)$ peaked around ℓ_1^2.

It is interesting to note that if we assume that $A \sim 1/\ell_1^2 \sim 1/\rho$ and that the density ρ is distributed according to a law $f(\rho)$ (with a finite $f(0)$); a simple calculation gives

$$P(A) \sim \frac{1}{A^2} f(1/A) \tag{2.33}$$

which behaves as $P(A) \sim 1/A^2$ for large A. This simple argument thus suggests that the observed value $\approx$2.0 of the exponent is universal and reflects the random variation of the density. More measurements are, however, needed at this point in order to test the validity of this hypothesis.

The authors of [29] also measured the distribution of the form or shape factor defined as the ratio of the area of the cell to the area of the circumscribed circle:

$$\phi = \frac{4A}{\pi D^2} \tag{2.34}$$

(for practical applications, D can be also taken as the longest distance in the cell). They found that most cells have a form factor between 0.3 and 0.6, suggesting a large variety of cell shapes, in contradiction with the assumption of an almost regular lattice. These facts thus call for a model radically different from simple models of regular or perturbed lattices. In Chaps. 3 and 7, we will discuss more thoroughly this quantity ϕ and its distribution.

2.2.4 Route Factor, Detour Index

When the network is embedded in a two-dimensional space, we can define at least two distances between the pairs of nodes. There is of course the natural Euclidean distance $d_E(i, j)$ which can also be seen as the "as crow flies" distance. There is also the total "route" distance $d_R(i, j)$ from i to j by computing the sum of length of segments which belong to the shortest path between i and j. The route factor (also called the detour index or the circuity, or directness [82]) for this pair of nodes (i, j) is then given by (see Fig. 2.11 for an example)

$$Q(i, j) = \frac{d_R(i, j)}{d_E(i, j)} \tag{2.35}$$

This ratio is always larger than one and the closer to one it is, the more efficient the network. From this quantity, we can derive another one for a single node defined by

$$\langle Q(i) \rangle = \frac{1}{N-1} \sum_j Q(i, j) \tag{2.36}$$

which measures the "accessibility" for this specific node i. Indeed the smaller it is and the easier it is to reach the node i (Accessibility is a subject in itself–see for example [83]—and there are many other measures for this concept and we refer the interested reader to the articles [84–86]). This quantity $\langle Q(i) \rangle$ is related to the quantity so-called "straightness centrality" [87] defined as

$$C^S(i) = \frac{1}{N-1} \sum_{j \neq i} \frac{d_E(i, j)}{d_R(i, j)} \tag{2.37}$$

If one is interested in assessing the global efficiency of the network, one can compute the average over all pairs of nodes (also used in [88])

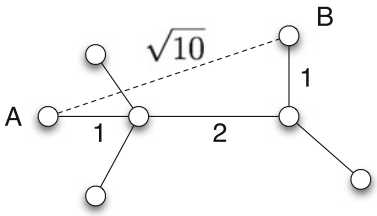

Fig. 2.11 Example of a detour index calculation. The "as crow flies" distance between the nodes A and B is $d_E(A, B) = \sqrt{10}$ while the route distance over the network is $d_R(A, B) = 4$ leading to a detour index equal to $Q(A, B) = 4/\sqrt{10} \simeq 1.265$

$$\langle Q \rangle = \frac{1}{N(N-1)} \sum_{i \neq j} Q(i, j) \tag{2.38}$$

The average $\langle Q \rangle$ or the maximum Q_{max}, and more generally the statistics of $Q(i, j)$, is important and contains a lot of information about the spatial network under consideration (see [89] for a discussion on this quantity for various networks). For example, one can define the interesting quantity [89]

$$\phi(d) = \frac{1}{N_d} \sum_{ij/d_E(i,j)=d} Q(i, j) \tag{2.39}$$

(where N_d is the number of nodes such that $d_E(i, j) = d$) whose shape can help for characterizing combined spatial and topological properties (see also Chap. 7 for empirical examples).

2.2.5 *Cost, Efficiency, and Robustness*

The minimum number of links to connect N nodes is $E = N - 1$ and the corresponding network is a tree. We can also look for the tree which minimizes the total length given by the sum of the length of all links

$$\ell_T = \sum_{e \in E} d_E(e) \tag{2.40}$$

where $d_E(e)$ denotes the length of the link e. This procedure leads to the minimum spanning tree (MST) which has a total length ℓ_T^{MST} (see also Sect. 12.2 about the MST). Obviously, the tree is not a very efficient network (from the point of view of transportation for example) and usually more edges are added to the network, leading to an increase of accessibility but also of ℓ_T. A natural measure of the "cost" of the network is then given by

$$C = \frac{\ell_T}{\ell_T^{MST}} \tag{2.41}$$

Adding links thus increases the cost but improves accessibility or the *transport performance* P of the network which can be measured as the minimum distance between all pairs of nodes, normalized to the same quantity but computed for the minimum spanning tree

$$P = \frac{\langle \ell \rangle}{\langle \ell_{MST} \rangle} \tag{2.42}$$

Another measure of efficiency was also proposed in [90, 91] and is defined as

$$E = \frac{1}{N(N-1)} \sum_{i \neq j} \frac{1}{\ell(i,j)} \tag{2.43}$$

where $\ell(i, j)$ is the shortest path distance from i to j. This quantity is zero when there are no paths between the nodes and is equal to one for the complete graph (for which $\ell(i, j) = 1$). The combination of these different indicators and comparisons with the MST or the maximal planar network can be constructed in order to characterize various aspects of the networks under consideration (see for example [46]).

Finally, adding links improves the resilience of the network to attacks or dysfunctions. A way to quantify this is by using the *fault tolerance (FT)* (see for example [92]) measured as the probability of disconnecting part of the network with the failure of a single link. The benefit/cost ratio could then be estimated by the quantity FT/ℓ_T^{MST} which is a quantitative characterization of the trade-off between cost and efficiency [92].

Buhl et al. [46] measured different indices for 300 maps corresponding mostly to settlements located in Europe, Africa, Central America, and India. They found that many networks depart from the grid structure with an alpha index usually low. For various world cities, Cardillo et al. [31] found that the alpha index varies from 0.084 (Walnut Creek) to 0.348 (New York City) which reflects in fact the variation of the average degree. Indeed for both these extreme cases, using Eq. (2.28) leads to $\alpha_{NYC} \simeq (3.38-2)/4 \simeq 0.345$ and for Walnut Creek $\alpha_{WC} \simeq (2.33-2)/4 \simeq 0.083$. This same study seems to show that triangles are less abundant than squares (except for cities such as Brasilia or Irvine).

Measures of efficiency are relatively well correlated with the alpha index but display broader variations demonstrating that small variations of the alpha index can lead to large variations in the shortest path structure. Cardillo et al. [31] plotted the relative efficiency (see Chap. 1)

$$E_{rel} = \frac{E - E^{MST}}{E^{GT} - E^{MST}} \tag{2.44}$$

versus the relative cost

$$C_{rel} = \frac{C - C^{MST}}{C^{GT} - C^{MST}} \tag{2.45}$$

where GT refers to the greedy triangulation (the maximal planar graph). The cost is here estimated as the total length of segments $C \equiv \ell_T$ and the obtained result is shown in Fig. 2.12 which demonstrates two things. First, it shows—as expected—that efficiency is increasing with the cost with an efficiency saturating at $\sim$0.8. In addition, this increase is slow: typically, doubling the value of C shifts the efficiency from $\sim$0.6 to $\sim$0.8. Second, it shows that most of the cities are located in the high-cost–high-efficiency region. New York City, Savannah, and San Francisco have the

largest value of the efficiency (∼0.8) with a relative cost value around ∼0.35. It seems, however, at this stage difficult to clearly identify different classes of cities and further studies with a larger number of cities are probably needed in order to confirm the typology proposed in [31].

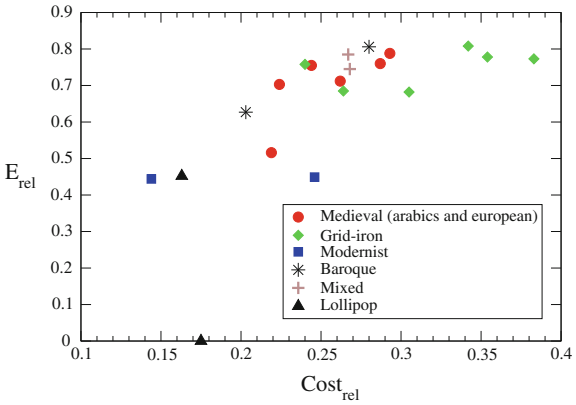

Fig. 2.12 Relative efficiency versus relative cost for 20 different cities in the world. In this plot, the point (0, 0) corresponds to the MST and the point (1, 1) to the greedy triangulation. Figure taken from [31]

Finally, we mention the study [93] on the street and subway networks of Paris and London. The accessibility in these cities is studied in terms of self-avoiding random walks displaying several differences. In particular, Paris seems to have a larger average accessibility than London, probably due to a large number of bridges.

http://www.springer.com/978-3-319-20564-9

Morphogenesis of Spatial Networks

Barthelemy, M.

2018, XIX, 331 p. 187 illus., Hardcover
ISBN: 978-3-319-20564-9