

Measuring fractal dimension of metro systems

S Deng¹, W Li^{1,2}, J Gu³, Y Zhu^{1,4,5}, L Zhao¹ and J Han¹

¹ College of Physical Science and Technology, Central China Normal University, Wuhan 430079, China.

² Max-Planck-Institute for Mathematics in the Science, Inselstr. 22, 04103 Leipzig, Germany

³ Mathematics and Science College, Shanghai Normal University, Shanghai 200234, China

⁴ ISMANS, LUNAM Université, 44 av. Bartholdi, 72000 Le Mans, France

⁵ IMMM, UMR CNRS 6283, Université du Maine, 72085 Le Mans, France

E-mail: gitsteven@gmail.com

Abstract. We discuss cluster growing method and box-covering method as well as their connection to fractal geometry. Our measurements show that for small network systems, box-covering method gives a better scaling relation. We then measure both unweighted and weighted metro networks with optimal box-covering method.

1. Introduction

Metro systems are playing an increasingly important role in facilitating urban transportation. In one of our former works by Li *et al* [1], we confronted with the problem of how to measure the fractal dimension of a metro network, we briefly demonstrated how we perform these measurements in the supplementary material of that paper, in this paper we organize our considerations on this issue so as to elucidate some main points that have been studied in the literature on measuring the fractal dimension of a network.

There are two aspects in the structure of a metro network: its geometrical structure and its topological structure. The geometrical structure is a direct map of lines and stations into $\mathbb{R}^2$ or $\mathbb{R}^3$. For such a geometrical object, if there is no a characteristic length scale, fractal geometry is a powerful tool to capture the self-similarity property [2]. An early work studying the geometry fractal property of urban railway system was conducted by Benguigui and Daoud [3], one may be also inclined to do the similar study on metro networks. Geometrical description has its advantages, whereas, it is restricted by practical factors as well as theoretical drawbacks. On the one hand, we found it very difficult to find a complete set of available GPS data; on the other hand, various properties of a metro system can be revealed by constructing its topological network and analysing this network with graph theory. So in the course of studying metro networks, a network accented fractal dimension may always be preferred. The most straightforward approach is to define fractal dimension on the networks.

There are two main definitions of fractal dimension of networks within the current literature: the cluster growing method introduced by Shanker [4], and the box-covering method proposed by Song *et al* [5]. Their studies were based on large size networks, with sizes range from a few thousand to hundreds of thousands. However, metro networks are relatively small, most of which have a size about 100–300 (see figure 2), making scaling relation more difficult to detect. With two definitions of fractal dimension at hand, the metro data enable us to systematically



study which is the better way of revealing scaling relation for small size networks, our work may provide a perspective to further studies on small size networks.

The data of 28 metro networks we studied in this paper were manually collected from corresponding official websites (see Appendix). We first observe the characteristics of these metro networks. In section 3 and section 4 we discuss the cluster growing method and the box-covering method with a coherent connection with geometry fractal. Upon these fundamental considerations together with our observations in the measuring processes, we use box-covering method to measure the fractal dimension of major world-wide metro networks.

2. Properties of typical metro networks

Many real-world networks have properties of small-world and scale-free. Small-world means in many actual networks, one only need a very small number of steps to reach one node from any other node. This implies a slow increase of average distance $\langle l \rangle$ of the network with the system size N [6–8]:

$$\langle l \rangle \sim \ln N \quad \text{or} \quad N \sim e^{\langle l \rangle / l_0}, \quad (1)$$

with l_0 a characteristic length. Scale-free property indicates that the degree distribution of a complex network obeys a power-law distribution with an exponent γ in the range (2, 3) [9–11]:

$$P(k) \sim k^{-\gamma}. \quad (2)$$

The degree distributions of metro networks are all very narrow (see figure 1), even an underling scale-free distribution may exist when the system size becomes large, we can't claim scale-free property for such relatively small systems. Figure 2 gives a scattering pattern between metro system size N and average distance $\langle l \rangle$, this pattern is very diverse from what may expect by Eq. (2). Therefore metro networks are *non-small-world* and *non-scale-free*. This is in agree with our daily experience that we don't have a special express train that connects two distant stations directly nor do we have hubs with large degree in a metro system.

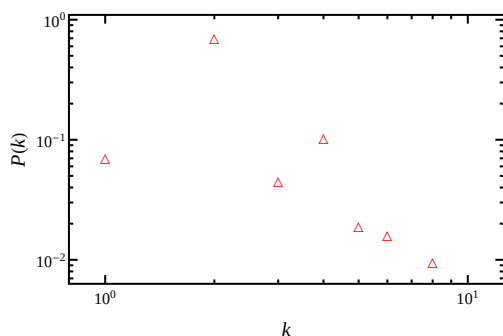


Figure 1. Degree distribution of Paris metro network.

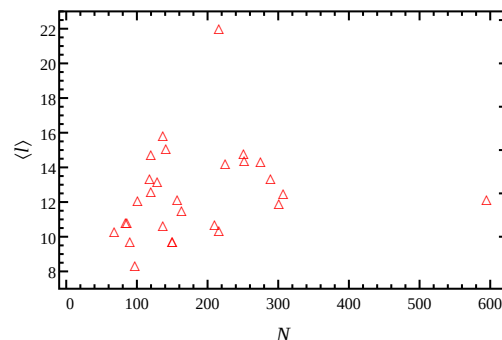


Figure 2. 28 metro networks: system size N versus average distance $\langle l \rangle$.

3. The cluster growing method

For a distance vector $\mathbf{r} = (x_1, x_2, \dots, x_d)$ in the Euclidean space $\mathbb{E}^d$, the Euclidean distance is defined as the l_2 -norm:

$$\|\mathbf{r}\|_2 = \sqrt{x_1^2 + x_2^2 + \dots + x_d^2}. \quad (3)$$

A fractal object $\mathbf{F}$ embedded in this space has a fractal dimension d_f defined as the scaling exponent of the scaling relation

$$V(r) \sim r^{d_f}, \quad (4)$$

where $V(r)$ is the volume of $\mathbf{F}$ enclosed in the d -sphere of radius r . In discrete space, the equivalence of $\mathbb{E}^d$ is the regular lattices $\mathbb{Z}^d$. A homogeneous complex network characterized by a homogeneous degree distribution can be regarded as a fractal object being embedded inside certain $\mathbb{Z}^d$ [4, 5]. For instance, in a one-dimensional ring consists of N nodes, adding edges in turn with probability p to connect each node N_i with a new node N_{i+m} that is m edges away, then the derived network is an untouched one-dimensional ring if $p = 0$ and a two-dimensional $m \times N/m$ lattice if $p = 1$. For $0 < p < 1$ the resulting network has a dimension between 1 and 2 [4]. In a network, the distance between two nodes is taken as the shortest path length. In this sense, this distance is expressed in $\mathbb{Z}^d$ as the l_1 norm

$$\|\mathbf{n}\| = \|n_1\| + \cdots + \|n_d\|, \quad (5)$$

where $\mathbf{n} = (n_1, n_2, \dots, n_d)$ is a distance vector in $\mathbb{Z}^d$. The fractal dimension d_f can be defined in a similar manner as Eq. (4) [4]

$$\langle M(r) \rangle \sim r^{d_f}, \quad (6)$$

where $M_i(r)$ is the number of nodes (mass) contained in a disk of radius r centered at node i . $M_i(r)$ generally varies with the seeded center i , so we need to average over many (say $\mathcal{N}$) seed center nodes. For a network consists of a large number of nodes, we can push $\mathcal{N}$ to be a large number and make $\langle M(r) \rangle$ almost independent of the seed nodes we choose. Whereas for our metro networks, the typical number of stations is less than 300 (see figure 2), randomness seems inevitable. We can always average over all nodes to avoid this randomness, but seed nodes taken from the periphery will bring in a large bias and break the scaling relation (6), as shown in figure 3 and Ref. [12]. As will be demonstrated in section 4, the scaling in the form of (6) doesn't even exist for small-world networks. So we will not utilize cluster growing method to measure fractal dimension of metro networks.

4. The box-covering method

In fractal geometry, another method of measuring fractal dimension is the box-covering method, in which we count the number of boxes N_B that is needed to cover a fractal object $\mathcal{F}$, given the scaling relation between N_B and box size s

$$N_B \sim s^{-d_B}, \quad (7)$$

d_B is then the fractal dimension of $\mathcal{F}$ [2]. On a network the similar tiling is illustrated in figure 4. The box size is defined as chemical distance l_B , which means that the distance of any two nodes within this box is less than l_B [5, 13]. Denoting the minimum number of boxes that is needed to cover the entire network by N_B , the fractal dimension d_B is then given by [2, 5]

$$N_B \sim l_B^{-d_B}, \quad (8)$$

provided that this scaling relation exists, and the network is said to be fractal. In fractal geometry, self-similarity refers to the invariance of a fractal object under scaling transformation. On a network, the self-similarity is the invariance of degree distribution $P(k)$ under iterative coarse graining renormalization procedure as defined in Ref. [5]. Topological fractality and self-similarity both manifest in some networks, such as WWW and actor networks [5, 14]. This is not a surprise, but Song *et al* in [5, 14, 15] found that non-fractal Internet network is actually

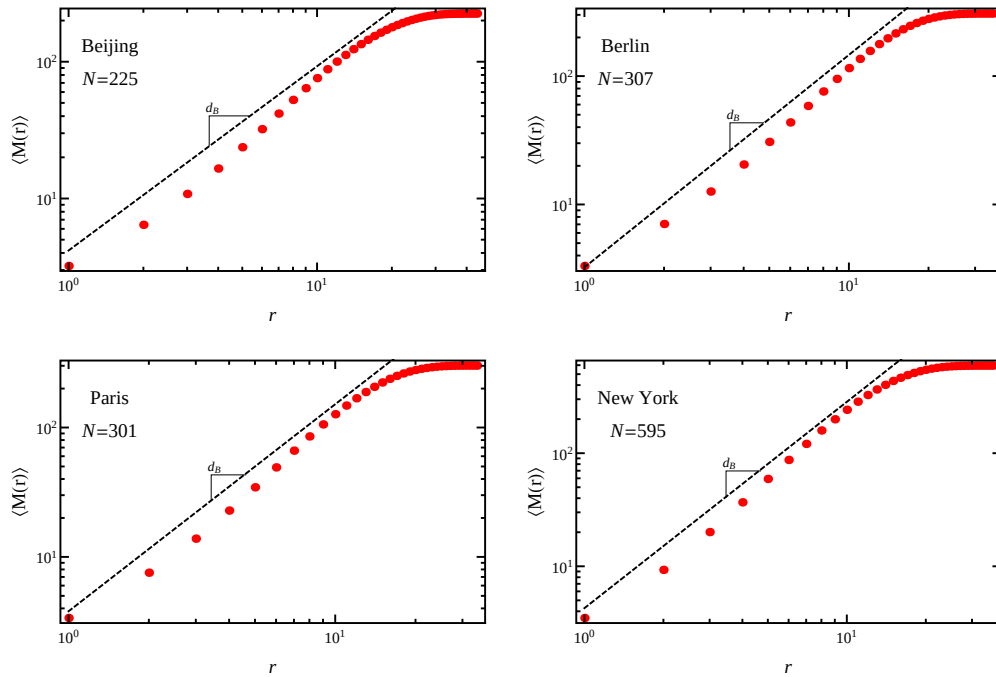


Figure 3. Cluster growing results for metro networks of Beijing, Berlin, Paris and New York: $\langle M(r) \rangle$ versus r , we cut off r at d_{max} , the diameter of the network. The dashed lines are plotted with slopes d_B obtained by box-covering method (see table 1).

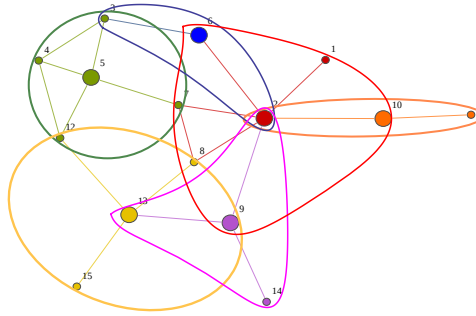


Figure 4. A sample network consists of 15 nodes covered by 6 boxes of size $l_B = 3$.

self-similar. Thus, unlike fractal geometry, fractality and self-similarity are considered to be distinct in complex networks.

The fractality of complex networks is exposed to be related to their underlying tree structure called skeleton [16]. A skeleton is a spanning tree consists of edges with largest betweenness centrality, the remaining edges are called shortcuts. If a network is constructed in a modular way, the edges with the highest betweenness centrality then provide the main communication routes between the modules. In [16] Goh *et al* found that for fractal networks, the number of boxes that is needed to cover the skeleton is almost the same as that of needed by the original fractal network, which implies that the fractality stems from the skeleton tree. In this reasoning,

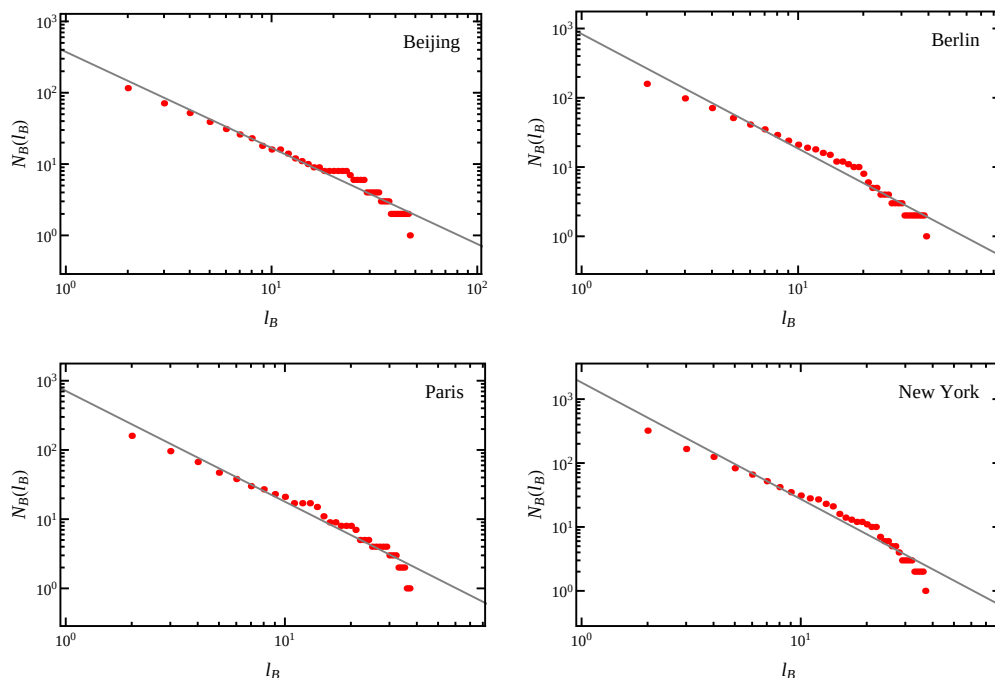


Figure 5. Box-covering results for metro networks of Beijing, Berlin, Paris and New York. The fitted slopes d_B are listed in table 1.

a skeleton tree can be considered as being embedded inside certain lattice of $\mathbb{Z}^d$, as discussed in section 3, such a structure is fractal. We can imagine the original network growing as the skeleton tree branches, as the tree branches, some local shortcuts are added which will preserve the “grand picture” of the tree. The dimension we measure on the original network is thus generally the dimension of its skeleton tree. If the locality of wiring the shortcuts is broken, the “grand picture” of the skeleton is lost, and fractality is erased [14, 16].

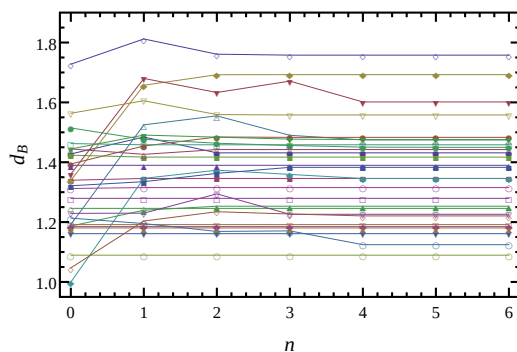
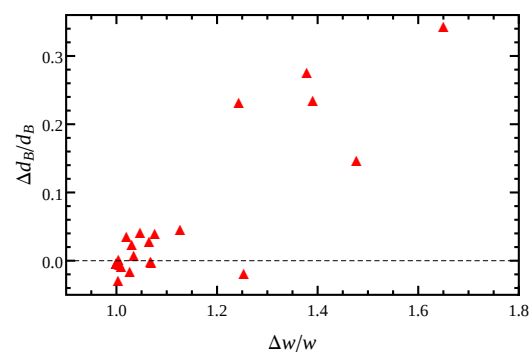
For a network with small-world or scale-free property that is characterized by a broad degree distribution, by using Eq. (4), typically, when choosing a random seed as the disk center, it is very probable to include hubs into the disk. Then $M(r)$ will grow in general as $\sim e^{r/l_0}$, which is not a power-law, hence Eq. (6) and (8) are not equivalent. This is due to the over-covering of hub nodes. On the other hand, box-covering method covers all nodes on the same footing, once a node (include hub) is covered, it cannot be covered again [5]. This distinction between the two methods makes cluster growing method bias in unveiling the essential tree structure when working with inhomogeneous networks. However, for a homogeneous network that can be regarded as being embedded inside a $\mathbb{Z}^d$ lattice, two measuring methods are equivalent, $d_f = d_B$ (cf. figure 3, as can be seen, for metro networks, the first few dots obtained by using cluster growing method are very close to the dashed line obtained by box-covering method).

In section 3, we encountered fitting problem when try to measure d_f . Although metro networks are not scale-free or small-world, their relatively small system size still cause fitting difficulties, we hope box-covering method can tame this problem. The measuring results ought to be deterministic due to the nature of this method, however, choosing the optimal minimum number of boxes to cover the network is NP hard, there are 2^N possibilities in the solution space for a system of size N . A former technique to attack this difficulty is devised by Song *et al* [17], they mapped this problem to graph coloring problem in which the boxes needed are the colors needed to colorize the dual network (see Ref. [17] for details). This greedy coloring

Table 1. Fractal dimension of 28 cities' metro networks

City	d_B	City	d_B	City	d_B
Barcelona	1.469 ± 0.052	London	1.452 ± 0.060	Seoul	1.509 ± 0.025
Beijing	1.347 ± 0.040	Madrid	1.409 ± 0.056	Shanghai	1.342 ± 0.069
Berlin	1.655 ± 0.056	Melbourne	1.231 ± 0.039	Shenzhen	1.185 ± 0.027
Boston	1.250 ± 0.062	Mexico	1.390 ± 0.040	Taipei	1.180 ± 0.056
Busan	1.161 ± 0.031	Milan	1.552 ± 0.062	Tokyo	1.424 ± 0.063
Chicago	1.319 ± 0.071	Montreal	1.090 ± 0.038	Valencia	1.255 ± 0.036
Delhi	1.192 ± 0.045	Moscow	1.462 ± 0.093	Vienna	1.279 ± 0.078
Guangzhou	1.246 ± 0.027	New York	1.824 ± 0.068	Washington	1.243 ± 0.072
Hamburg	1.410 ± 0.068	Osaka	1.445 ± 0.081		
Hong Kong	1.233 ± 0.045	Paris	1.600 ± 0.061		

algorithm is efficient and can give results nearly optimal. Some alternatives include several burning algorithms, where boxes that apparently do not belong to the solution are sequentially excluded [17]. Christian *et al*¹ [13] extended burning algorithm by firstly burning unnecessary boxes and nodes which reduces the solution space considerably, then accompany by splitting the resulting system into subsystems and candidate configurations. In this manner, after several iterations, all candidates are taken into account and the optimal solution can be picked out. But there is a subtlety here, in [13] boxes are calculated with a central node or edge, the box radius r_B and box size l_B is connected by $r_B = (l_B - 1)/2$ for a central node and $r_B = l_B/2$ for a central edge, this may lead to one more box for loops. For example, in a chain of four nodes (1-2-3-4-1), all nodes are separated from other nodes at most by 2 edges ($l_B = 3$), we only need one $l_B = 3$ box to cover the chain; however, we cannot find a box with radius one ($r_B = (l_B - 1)/2 = 1$) which contains all nodes, and we need one more box to cover all nodes. Excepting for this subtlety, we can take this algorithm to be optimal, and we will use it extensively hereafter. Some measuring examples are shown in figure 5. Unlike in figure 3, here we don't need to worry about where to cut off for our measurement. We list the full measuring results in table 1, dimension values are estimated by linear regression, and the errors are estimated as the fitting error.

**Figure 6.** The weighted dimension of 28 metro networks vs. the exponent n .**Figure 7.** The impact of weight on the change of dimension, here $n = 1$.

¹ We thank for their courtesy for sharing their implementation code.

Many real-world networks are weighted, including many metro networks. In metro networks, the weight of an edge is the number of lines it contains. If there are many such weighted edges, not only the weighted edges can sustain a larger conductance, but also the transition between different lines of the metro will be largely improved, one can then get easier to transport to his/her destination. This empirical experience encourages us to regard edge weights as an evaluation of the shrink of the distance. The traditional interpretation of edge weight is the distance of this edge: an unweighted edge has distance 1 and an edge of weight w represent a distance of w . We see this is not quite right in metro networks, a more general connection between the distance and weight is given by [18]

$$l_{ij} = \frac{1}{w_{ij}^n}. \quad (9)$$

In order to “shrink” the distance, we can choose $n > 0$, when $n = 0$ we retain the unweighted case. In figure 6-7, we examine the influence of weights on dimension values. Figure 6 shows that shrinking the distance according to Eq. (9) will lead to a change of dimension value. As we increase n , most dimension values increase first then fall down, all dimension values tend to a constant if n is large enough. In figure 7, taking $n = 1$, the comparison between change of dimension value versus change of weight indicates a increasing trend between these two changes. Unweighted networks should fall on the dashed line, but it is interesting to see a few dots go below the dashed line even though they correspond to weighted network.

5. Summary

We discussed cluster growing method and box-covering method, and their connection with fractal geometry. According to these discussions, the two methods are equivalent for homogeneous systems whereas are distinct for inhomogeneous systems. The non-small-world and non-scale-free properties of metro networks allow us to measure fractal dimension with cluster growing method. But our measurements show that for such small systems, box-covering method are still more powerful in revealing the scaling relation.

Acknowledgments

This work was partially supported by the Programme of Introducing Talents of Discipline to Universities under Grant No. B08033.

Appendix

Data resources

Our data of 28 world major metros, updated till Nov. 7, 2013, can be retrieved from the following web sites:

- Barcelona: www.stm.info/en/info/networks/metro
- Beijing: www.bjsubway.com/
- Berlin: www.bvg.de/index.php/en/index.html
- Boston: www.mbta.com/schedules_and_maps/subway/
- Busan: www.subway.busan.kr/china/main/
- Chicago: www.metrochicago.com/
- Delhi: www.delhimetrorail.com/
- Guangzhou: www.gzmtr.com/
- Hamburg: www.hvv.de/en/
- Hong Kong: www.mtr.com.hk/eng/homepage/cust_index.html
- London: www.tfl.gov.uk/modalpages/2625.aspx
- Madrid: www.metromadrid.es/en/
- Melbourne: www.metrotrains.com.au/

- Mexico: mexicometro.org/
- Milan: www.atm.it/en/Pages/default.aspx
- Montreal: www.stm.info/en/info/networks/metro
- Moscow: engl.mosmetro.ru/
- New York: new.mta.info/
- Osaka: www.kotsu.city.osaka.lg.jp/foreign/english/
- Paris: www.parismetro.com/
- Seoul: www.smrt.co.kr/main/index/index002.jsp
- Shanghai: www.shmetro.com/
- Shenzhen: www.szmc.net/
- Taipei: english.trtc.com.tw/
- Tokyo: www.tokyometro.jp/en/
- Valencia: www.metrovalencia.es/page.php
- Vienna: www.wienerlinien.at/eportal/ep/home.do?tabId=0
- Washington: www.wmata.com/

References

- [1] Li W, Gu J, Liu S, Zhu Y, Deng S, Zhao L, Han J and Cai X 2014 Optimal Transport in Worldwide Metro Networks *preprint* arXiv:1403.7844
- [2] Mandelbrot B B 1983 *The Fractal Geometry of Nature* vol 173 (New York: Macmillan)
- [3] Benguigui L and Daoud M 1991 *Geog. Anal.* **23** 362–368
- [4] Shanker O 2007 *Mod. Phys. Lett. B* **21** 321–326
- [5] Song C, Havlin S and Makse H A 2005 *Nature* **433** 392–395
- [6] Watts D J and Strogatz S H 1998 *Nature* **393** 440–442
- [7] Kleinberg J M 2000 *Nature* **406** 845–845
- [8] Watts D J 2004 *Six Degrees: The Science of a Connected Age* (New York: W.W. Norton & Company)
- [9] Barabási A L and Albert R 1999 *Science* **286** 509–512
- [10] Newman M E 2003 *SIAM Rev.* **45** 167–256
- [11] Dorogovtsev S N and Mendes J F 2013 *Evolution of Networks: From Biological Nets to the Internet and WWW* (Oxford: Oxford University Press)
- [12] Long G and Cai X 2009 *Chin. Phys. Lett.* **26** 088901
- [13] Schneider C M, Kesselring T A, Andrade Jr J S and Herrmann H J 2012 *Phys. Rev. E* **86** 016707
- [14] Song C, Havlin S and Makse H A 2006 *Nat. Phys.* **2** 275–281
- [15] Gallos L K, Song C and Makse H A 2007 *Physica A* **386** 686–691
- [16] Goh K I, Salvi G, Kahng B and Kim D 2006 *Phys. Rev. Lett.* **96** 018701
- [17] Song C, Gallos L K, Havlin S and Makse H A 2007 *J. Stat. Mech: Theory Exp.* **2007** P03006
- [18] Wei D J, Liu Q, Zhang H X, Hu Y, Deng Y and Mahadevan S 2013 *Sci. Rep.* **3**