

Self-similar scaling of density in complex real-world networks

Neli Blagus*, Lovro Šubelj, Marko Bajec

University of Ljubljana, Faculty of Computer and Information Science, Ljubljana, Slovenia

Abstract

Despite their diverse origin, networks of large real-world systems reveal a number of common properties including small-world phenomena, scale-free degree distributions and modularity. Recently, network self-similarity as a natural outcome of the evolution of real-world systems has also attracted much attention within the physics literature. Here we investigate the scaling of density in complex networks under two classical box-covering renormalizations—network coarse-graining—and also different community-based renormalizations. The analysis on over 50 real-world networks reveals a power-law scaling of network density and size under adequate renormalization technique, yet irrespective of network type and origin. The results thus advance a recent discovery of a universal scaling of density among different real-world networks [Laurienti et al., *Physica A* 390 (20) (2011) 3608-3613.] and imply an existence of a scale-free density also within—among different self-similar scales of—complex real-world networks. The latter further improves the comprehension of self-similar structure in large real-world networks with several possible applications.

Keywords: complex networks, self-similarity, network density, community structure

PACS: 89.75.Hc, 05.45.Df, 89.75.Da, 89.75.Fb

1. Introduction

The study of complex real-world networks and underlying systems has erupted in recent years in various fields of science. Due to their simple and intelligible form, networks enable representation of diverse systems of complex interactions and provide for their common investigation. Thus, several fundamental properties of large real-world networks have been revealed in the past decade. These include small-world phenomena [1], scale-free degree distributions [2, 3], network clustering [1, 4] and robustness [5, 6], degree mixing [7, 8], community and hierarchical structure [9, 10], network motifs [11] and other [12] (for reviews see [13, 14]). More recently, network self-similarity as an inherent property behind the evolution of real-world systems has also attracted much of attention within the physics community [15, 16, 17, 18, 19, 20].

*Corresponding author. Tel.: +386 1 476 81 86.

Email addresses: neli.blagus@fri.uni-lj.si (Neli Blagus), lovro.subelj@fri.uni-lj.si (Lovro Šubelj), marko.bajec@fri.uni-lj.si (Marko Bajec)

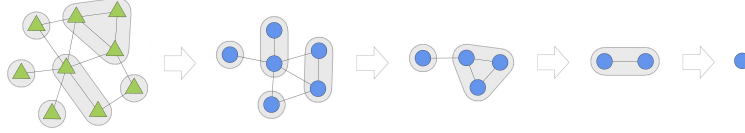


Figure 1: Network renormalization—system coarse-graining—technique [16, 25] applied to a small example network. At each step, the network is covered with boxes that are replaced by super-nodes. The latter are linked when a corresponding link also exists in the (original) network. The process then repeats until only a single node remains or multiple nodes in the case of a disconnected network. (Here the network is randomly tiled with boxes of nodes at a distance smaller than 2.)

Network self-similarity is commonly considered alongside the concept of fractal networks [16, 21]. Fractality is a property of a geometric object that it is exactly or approximately similar to a part of itself [22]. Nevertheless, classical theory of self-similarity requires a power-law scaling between the system size and its parts under some renormalization [23, 24]. The latter is an iterative process where a system is coarse-grained into smaller replicas, thus its essential structural features are preserved [16, 25] (Fig. 1). Hence, fractal or self-similar networks commonly refer only to a self-similar scaling exponent in the aforementioned power-law relation [16, 26, 18, 27]. However, network self-similarity is also investigated in the context of other network properties [15, 17, 18, 19, 28] under various renormalization techniques [16, 25, 29, 30].¹ (Note that fractal scaling laws observed in real-world networks do not necessarily imply a self-similar network [31].)

Guimerà et al. [15] have first observed self-similar community size distributions in a network of human communications. Furthermore, Song et al. [16, 26] have proposed an adequate renormalization technique (Fig. 1) to expose the origin of self-similar fractal scaling in web, collaboration and different biological networks. The latter in fact gives rise to degree disassortativity [26] and resilience to diseases [32], commonly observed for these networks. Still, such scaling cannot coexist with a small-world network topology [33, 34]. Self-similarity has also been considered as a scale-invariance of degree distribution [18, 27] or maximum degree [19, 35] under network renormalization, while Itzkovitz et al. [17] have revealed self-dissimilarity in a motif structure for different biological and technological networks. Authors have also considered network self-similarity in the context of different dynamical processes including percolation [36] and synchronization [37].

Despite the above efforts, there is yet little evidence whether self-similarity exists only in certain networks and which properties are indeed invariant throughout different network scales. We thus here investigate the scaling of density—defined as the number of links to all possible links—with respect to network size under five renormalization techniques borrowed from the field of fractal networks [16, 26] and community detection literature [38, 39]. Analysis on over 50 real-world networks of diverse origin reveals a self-similar power-law scaling of network density and size (under suitable renormalization). The latter advances a recent work of Laurienti et al. [40] who have observed a universal scaling of density among different real-world networks, while Leskovec et al. [41, 42] have also found similar densification laws in evolving networks. The results thus imply an existence of a scale-free density not only among, but also within—among different

¹Throughout the paper we refer to network self-similarity in a general sense.

self-similar scales of—complex networks irrespective of their type and the underlying domain. Hence, under adequate renormalization self-similar real-world networks neither get denser nor sparser with respect to their size, whereas characteristic network topology is also largely retained throughout the renormalization.

The rest of the paper is structured as follows. Section 2 introduces different renormalization techniques and real-world network data adopted in the research. Empirical analysis with formal discussion on real-world and random networks is presented in Section 3, while Section 4 gives final conclusions and discusses future work.

2. Techniques and network data

Self-similarity is primarily studied under the framework of network renormalization [16, 18]. As already discussed, renormalization is an iterative coarse-graining technique, where the original network is covered with boxes, thus each node belongs to exactly one box [16, 25] (Fig. 1). Boxes are then replaced by super-nodes that are linked when a corresponding link also exists in the (original) network. The entire process repeats until no links remain and the number of nodes equals to the number of connected components.

While there exists a number of different box-covering approaches, not all of them are able to reveal self-similar scales in complex networks. Thus, we employ techniques that have already proven useful for exposing self-similarity in various real-world networks [15, 16, 18, 19]. In particular, we adopt methods commonly used in analysis of fractal networks, as well as different community detection algorithms.

Fractal network structure is mainly explored under two general classes of renormalization techniques, namely, node coloring and network burning approaches [16, 29] (for reviews see [20, 27]). In the former, box-covering is mapped to a node coloring problem [43, 44], whereas, in the latter, boxes are grown around a randomly selected seed node. Although there exist several efficient algorithms for node coloring [44, 45], network burning methods offer some distinct advantages [29]. Different authors have proposed a wide range of alternative network coarse-graining techniques including methods based on connectivity patterns [17], skeleton of the network [25], link-covering [46] and other [18, 47, 30, 37].

For the purpose of this research, we adopt two classical network burning approaches. First, box-tiling method, randomly tiles the network with boxes of nodes that are at a distance smaller than l_B [16, 26] (Fig. 1). Second, cluster-growing method, incrementally grows boxes from randomly selected seed nodes within a distance not larger than r_B [29, 47]. Hence, for random configurations, $l_B = 2 \cdot r_B + 1$ [29]. Box-tilling method allows for somewhat easier analytical consideration, whereas cluster-growing approach enables more efficient implementation. For the analysis in Section 3, we set l_B to 3 and r_B to 2 with respect to network small-worlds [1]. Note that the latter extends the definition of an *egonet* [48, 49]—a subnetwork inferred by a central ego node and its neighbors—which can be seen as a local signature of the respective node.

We further adopt several algorithms drawn from community detection literature (for reviews see [38, 39]). Here boxes are identified by communities [9]—groups of nodes densely connected within and only loosely connected between—revealed with selected algorithm, whereas network coarse-graining procedure is else identical as above. Community detection has already been successfully employed to reveal self-similarity in real-world networks [15]. Recent work also implies an existence of community structures on

Table 1: Real-world networks. (n and m correspond to the number of nodes and links, respectively.)

Network	Type	n	m
Zachary's karate club [50]	Social	34	78
Lusseau's dolphins [51]		62	159
Comp. sci. PhD students [52]		1025	1043
<i>Facebook</i> friendships [53]	On-line social	324	2218
<i>Wikipedia</i> who-votes-who [54]		7066	100736
Slovenian comp. science [53]	Collaboration	239	568
Krebs's Internet industry [52]		219	630
Complex networks science [55]		379	914
Paul Erdős collaborations [52]		446	1413
<i>Comput. Geometry</i> archive [56]		3621	9461
<i>General Relativity</i> archive [41]		4158	13422
<i>PGP</i> web-of-trust [57]		10680	24316
<i>Astro Physics</i> archive [42]		17903	196972
US political books [58]	Co-purchase	105	441
<i>amazon.com</i> domain [59]	Web graph	2879	3886
<i>epa.gov</i> domain [52]		4253	8897
Broad-topic queries [60]		5925	15770
US political blogs [58]		1222	16714
<i>Graph Drawing</i> proceedings [52]	Citation	249	635
Stanley Milgram citations [52]		233	994
H. Small & B. Griffith citations [52]		1024	4916
<i>Scientometrics</i> archive [52]		2678	10368
Teuvo Kohonen citations [52]		3704	12673
Joshua Lederberg citations [52]		8212	41430
Ahmed Zewail citations [52]		6640	54173
<i>High E. Particle Phys.</i> archive [61]		27400	352021
Mobile phone records [62]	Communication	345	355
Emails at a university [15]		1133	5451
Emails at <i>Enron</i> [63]		33696	180811
Novel <i>David Copperfield</i> [55]	Information	112	425
<i>Roget's Thesaurus</i> dictionary [64]		994	3640
Java documentation (<i>javax</i>) [65]		1031	4408
<i>ODLIS</i> dictionary [66]		2898	16376
<i>USF</i> association norms [67]		10617	63782
<i>FOLDOC</i> dictionary [68]		13356	91471
<i>WordNet</i> dictionary [52]		75606	119564
Small software project [52]	Software	83	125
<i>JUNG</i> graph framework [69]		398	943
Java language (<i>javax</i>) [69]		1570	7194
Java language (general) [52]		1538	7817
<i>Oregon</i> aut. systems [70]	Internet	22963	48436
<i>Gnutella</i> file sharing [42]		36646	88303
European roads [71]	Technological	1039	1305
Finite automaton [52]		1096	1677
US air lines [52]		332	2126
US power grid [52]		4941	6594
<i>Escherichia Coli</i> regulatory [52]	Biological	328	456
<i>Caenorhabditis Elegans</i> neural [1]		297	2148
Yeast protein interactions [72]		2224	6609
Data modeling [52]	Other	638	1020
<i>Amazon</i> products [73]	Co-purchase	524366	1491774
<i>nd.edu</i> domain [74]	Web graph	325729	1497135
Pennsylvania roads [63]	Technological	1087562	1541514
<i>Wikipedia</i> talk service [54]	Communication	2388953	4656682
<i>Skitter</i> overlay map [41]	Internet	1694616	11094209

various scales of complex real-world networks [75, 76]. Hence, community detection appears to be an adequate alternative to classical box-covering renormalization techniques.

Due to generality, we consider three diverse community detection algorithms. First, we adopt balanced propagation [71] as an example of a highly scalable state-of-the-art algorithm. The approach is based on the label propagation principle of Raghavan et al. [77], while node balancers are introduced to improve the stability of the algorithm (stability parameter is set to $1/4$). Next, we employ a fast hierarchical optimization of modularity Q [78] proposed by Clauset et al. [79] as one of most widely used approaches in the past literature [38]. However, due to many limitations of the measure of modularity Q , high values of Q cannot be regarded as an indication of network community structure [80, 81]. Last, we also consider a spectral algorithm of Newman [55] as a representative of a partitioning approach with origins in classical graph theory [82]. The algorithm reveals communities by extracting the leading eigenvector of network modularity matrix using a power method.

Analysis in Section 3 is conducted on 55 real-world networks that are often analyzed in complex network literature (Table 1), and also on random graphs á la Erdős-Rényi [83] and different generative graph models. The real-world networks range between tens of nodes and tens of millions of links; and include different social—classical, on-line, collaboration etc.; information—web graphs, citation, communication etc.; technological—Internet, software, transportation etc.; biological—protein, genetic and neural; and other networks. Due to the large number of networks considered, detailed description is omitted. Still, networks were carefully chosen thus to represent a relatively diverse set of real-world systems including most types of networks commonly analyzed in the literature. For simplicity, all networks are considered as simple undirected graphs and reduced to largest connected components.

3. Analysis and discussion

In the following we first analyze self-similar scaling of density in real-world networks of moderate size (Section 3.1), while analysis on Erdős-Rényi random graphs and different generative graph models is given in Section 3.2. In Section 3.3 we further consider self-similarity of five larger real-world networks with at least a million links.

3.1. Real-world networks

The algorithms were first applied to 50 real-world networks (Table 1). According to the number of nodes n and density d from original and reduced networks we examine the density scaling with respect to network size. In particular, d is expressed as a power function of n through formula $d = c \cdot n^{-\gamma}$, where γ is a scaling exponent and c is a constant. We measure goodness of fit to the data using coefficient of determination R^2 —how well the network size predicts density—and dependence between both variables corresponding to Spearman’s correlation coefficient ρ —the extent to which network density decreases as network size increases. Moreover, we also evaluate the number of self-similar scales S defining how many renormalized networks are revealed under different techniques.

Mean estimates for each method appear in Table 2. Coefficients R^2 demonstrate that the power-law relationship between the size and density appears to be a good fit to the data under box-covering methods and balanced propagation based renormalization.

Table 2: Estimates of the fit for power-law scaling of network density and size in 50 real-world networks revealed under different renormalization techniques. Values are estimates of the mean over 10 renormalizations of each network and correspond to correlation coefficient ρ , coefficient of determination R^2 , expressed network density d and the number of revealed self-similar scales S . (For each technique, ρ and R^2 are obtained separately for original and renormalized networks, and for renormalized varieties only—first and second row, respectively. Bold values of R^2 indicate relatively high goodness of fit to a power-law, whereas values in italics show poor performance of the respective renormalization technique.)

Technique	ρ	R^2	d	S
Randomized box-tiling	−0.975	0.944	$1.7 \cdot n^{-0.807}$	5.3
	−0.973	0.936		
Randomized cluster-growing	−0.977	0.948	$1.6 \cdot n^{-0.818}$	4.6
	−0.977	0.944		
Balanced propagation	−0.985	0.962	$1.9 \cdot n^{-0.836}$	4.3
	−0.980	0.963		
Modularity optimization	−0.966	0.956	$3.0 \cdot n^{-0.882}$	3.9
	−0.889	<i>0.820</i>		
Spectral analysis	−0.951	0.922	$4.1 \cdot n^{-0.893}$	4.5
	−0.878	<i>0.718</i>		
Original networks	−0.924	0.870	$3.8 \cdot n^{-0.921}$	

(We can reject the null hypothesis—no actual relationship between variables—at one percent significance level, thus results are statistically significant.) Irrespective of renormalization technique, R^2 and ρ for original networks are improved considering also their renormalized varieties. Otherwise, box-covering methods perform better than community detection algorithms, whereas balanced propagation exhibits the most homogeneous relationship between size and density. Spectral algorithm and modularity optimization prove the worst, particularly at observing fits for renormalized networks only. In the case of modularity optimization, this could be largely due to its resolution limit [80], lack of global maximum and degeneracy of optimal partitions [81]. On the other hand, spectral analysis is in fact an optimization of eigenvectors of the modularity matrix. Therefore, it is attributed to the above mentioned modularity limitations, whereas it also reveals modules in random networks [84].

The plots on Fig. 2 illustrate size and density relationships with the scaling exponents γ around -0.85 . Original networks exhibit greater scaling factor (see also Section 3.3), which indicates γ is approaching -1 for adequately large n . This corresponds to commonly observed finding that most large-scale real-world networks tend to be sparse—the number of links appears not to be close to $O(n^2)$ but rather of order $O(n)$. Consecutively, we can simplify density definition with the relationship $d \approx n^{-1}$. Thus, power-law relationship between the network size and density is expected for original networks (without considering reduced varieties). However, among renormalized networks the relationships follow even stronger power-laws. This means that networks obtained on different scales of renormalization process also satisfy power-law relationship between size and density, and implies an existence of density scaling also within real-world networks.

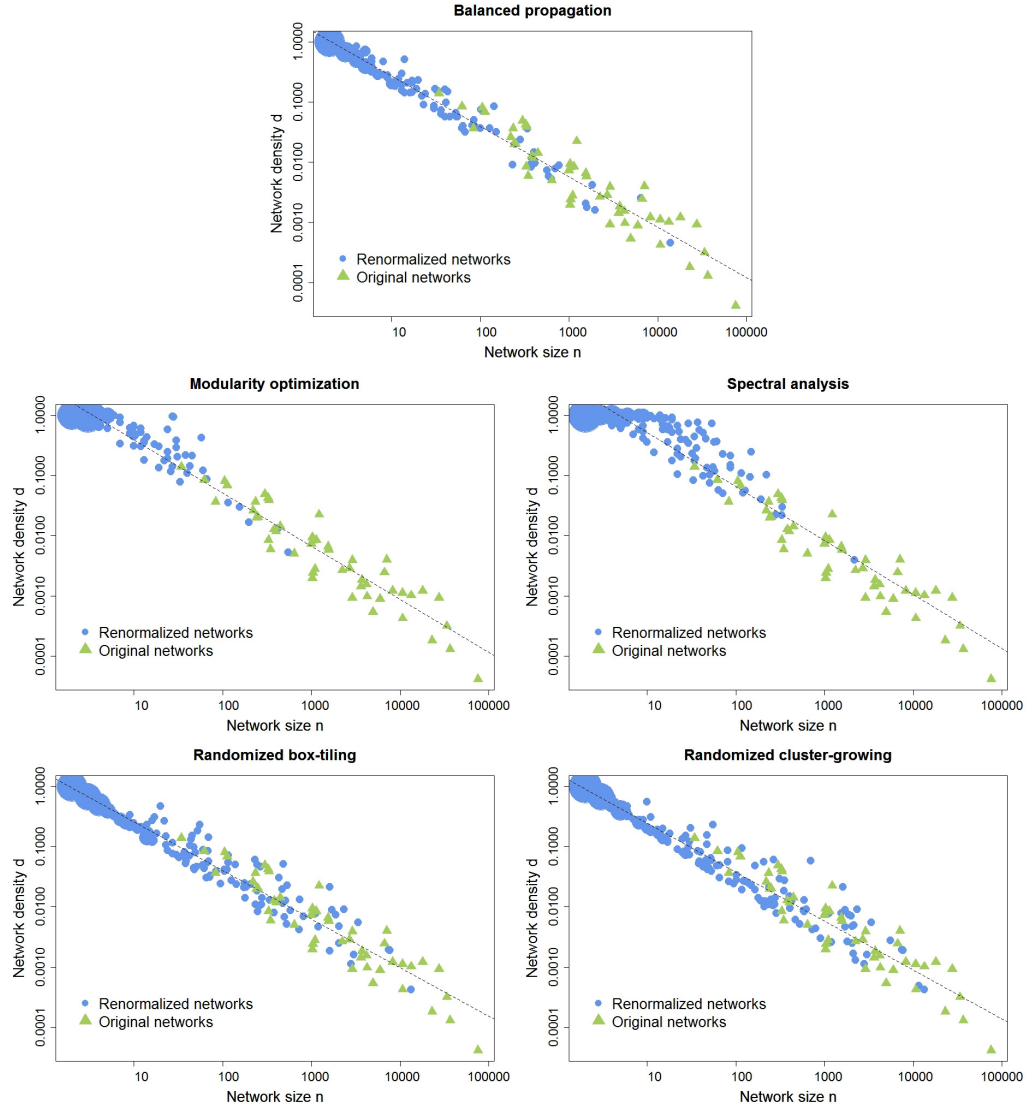


Figure 2: Power-law scaling of network density and size in 50 real-world networks of diverse origin revealed with different renormalization techniques. Plots show scaling of density for a single renormalization of each network under respective technique. (Green triangles correspond to original networks, whereas blue circles represent their renormalized varieties. Symbol sizes are proportional to the number of networks with the same size and density.)

Furthermore, results show similar behavior of exponents γ and constants c for better performing techniques, including box-tiling, cluster-growing, and balanced propagation. This finding implies that box-covering methods find smaller and sparser boxes, similar to communities detected with balanced propagation. Other two algorithms reveal bigger, denser, and also more heterogeneous communities considering density scaling. The values of self-similar scales S are in accordance with these observations. Modularity optimization extracts network with one community in the least number of scales on average (bigger communities). On the other hand, box-tiling obtains a larger number of reduced networks (smaller boxes), which is expected due to the distance l_B setting.

To summarize, the analysis of real-world networks reveals power-law scaling of the network density with respect to network size. Among the employed renormalization techniques, balanced propagation seems to lead to the most optimal reduction of networks according to the density scaling. Results acquired by three best performing techniques indicate an existence of a certain common organizing principle of networks, which dictates linking rules and interactions among nodes. Our findings thus advance a recent discovery of a universal scaling of density among real-world networks [40], since we reveal density scaling also among different self-similar scales of complex real-world networks. In addition, the results are consistent with the densification laws of Leskovec et al. [42, 41]— $m \propto n^\alpha$, where α ranges between 1 and 2 and relates with our exponent γ , which lies between 0 and -1 respectively. Thus, our study expands densification laws to other dimensions of network structure.

Besides density, we also studied the scaling of other network properties with respect to network size. In particular, we analyzed number of links, average and maximum degree, number of articulation points, average path and diameter [1], betweenness and closeness centrality [85] and clustering coefficient [86]. The results reveal significant scaling also between network size and average node or link betweenness—the number of shortest paths going through a node and link respectively. Regarding to a definition of network density and observed power-law relationship between size and density, similar relationship for number of links occurs expectedly. However, due to simplicity, detailed investigation of betweenness centrality scaling is omitted, although a prominent direction for future research.

3.2. Random networks

To further validate our results we apply box-tiling and modularity optimization renormalizations to Erdős-Rényi random graphs with different sizes n and probabilities of linking nodes p . We generate networks with 500, 1000, 2500, 5000, and 10000 nodes and probabilities corresponding to density obtained with modularity optimization based renormalization (Section 3.1), density reported in [40], and probability that should assure sufficient size of the largest network component [83].

Firstly, we test balanced propagation renormalization, since the method performs best on real-world networks. The results prove to be very good, showing fits closely to ideal (R^2 and ρ close to 1 and -1 , respectively). However, detailed investigation shows renormalization for most of the generated networks reveals only a single scale or concludes without reduction, since random networks supposedly have no community structure. For this reason we exclude balanced propagation from the analysis. Thus, we study box-tiling as an illustration of classical box-covering principle and modularity optimization as an

Table 3: Estimates of the fit for power-law scaling of network density and size in Erdős-Rényi random graphs obtained with two renormalization techniques. For each probability of a link between two nodes p , we construct an ensemble of networks of various sizes. Values are estimates of the mean over 10 realizations of each random graph. (See also Table 2.)

p	Technique	ρ	R^2	d	S
$3.0 \cdot n^{-0.882}$	Randomized box-tiling	-0.925	<i>0.820</i>	$2.8 \cdot n^{-0.753}$	4.9
		-0.852	<i>0.787</i>		
	Modularity optimization	-0.957	0.994	$12.3 \cdot n^{-0.963}$	3.0
		-0.583	<i>0.446</i>		
$7.9 \cdot n^{-0.986}$	Randomized box-tiling	-0.939	<i>0.818</i>	$3.7 \cdot n^{-0.797}$	4.5
		-0.882	<i>0.781</i>		
	Modularity optimization	-0.964	0.998	$10.5 \cdot n^{-1.022}$	3.0
		-0.494	<i>0.537</i>		
$2/(n-1)$	Randomized box-tiling	-0.990	0.967	$2.6 \cdot n^{-0.953}$	6.7
		-0.986	0.962		
	Modularity optimization	-0.930	0.916	$6.4 \cdot n^{-1.065}$	4.0
		-0.744	<i>0.817</i>		

example of community based renormalization. Note that, in contrast to the above, the latter reveals non-trivial modules also in random networks (Section 3.1).

The results appear in Table 3. A strong relationship ($R^2 = 1$, $\rho = -1$) arises between size and density of original networks. That occurs due to the settings of probability p . These strong fits cause also high values for original and randomized networks together. The results for randomized varieties of networks show low fits to the data and implies rather diverse density of reduced networks with respect to their size. This is anticipated owing to random network structure. However, the values of R^2 and ρ for randomized networks under $p = 2/(n-1)$ setting are relatively high. Examining plot for box-tiling closely shows diverse density among reduced networks, however, diversity straightens due to the large number of reduction scales. On the other hand, networks reduced under modularity optimization on each scale reveal almost the same density, and thus lead to higher fit. Slightly greater values for renormalized networks under box-tiling seem to occur due to the definition of boxes, which consider only proximity among nodes.

Other variables, including scaling exponent γ , constant c , and revealed self-similar scales S , comprehend greater range than values for real-world networks. This verifies there exists no optimal density characteristic for random networks and denotes that random networks do not exhibit common power-law density scaling.

According to the above, we conclude that results for random networks appear to be weak as anticipated, since random networks should not reveal structures like communities in real-world networks. On the contrary, findings indicate that self-similar density scaling of real-world networks is not obtained by chance, and the scaling exists due to some inner principles which determine network structure.

We have also analyzed several generative graph models, whether they reveal similar

Table 4: Estimates of the fit for power-law scaling of network density and size in five large real-world networks revealed with balanced propagation. Values are estimates of the mean over 10 renormalizations of each network. (See also Table 2.)

Technique	ρ	R^2	d	S
Balanced propagation	-0.990	0.977	$2.9 \cdot n^{-0.926}$	4.9
	-0.980	0.961		
Original networks	-0.900	0.719	$66.2 \cdot n^{-1.175}$	

scaling of density as observed in real-world networks. Expectedly, under balanced propagation renormalization, classical scale-free [3, 87] and small-world graph [88] models show the same behavior as in the case of random graphs (due to the lack of community structure). On the other hand, forest fire model proposed by Leskovec et al. [42, 41] reveals strong power-law scaling of density with estimates similar to those observed in Table 2 (exact results are omitted). Interestingly, the model gives networks that also obey network densification laws, shrinking diameters, community structure and scale-free degree distributions [42], and thus provide a relatively realistic structure of real-world networks. Note that community guided attachment model [42, 41] that also follows densification laws, does not reveal self-similar scaling of density; thus, the latter is indeed not an artifact of the former, but rather extends network density laws to other dimensions.

3.3. Large real-world networks

For a complete analysis, we also analyze the size and density relationship of the largest five real-world networks presented in Table 1. In particular, co-purchase network of different products from Amazon in 2006, complete map of *nd.edu* domain, road network of Pennsylvania, communication network of user discussions on Wikipedia before January 2008, and Internet topology graph from traceroutes in 2005. Due to simplicity, we present study only for the best performing balanced propagation based renormalization, where the maximum number of iterations is limited to 100.

The results are presented in Table 4. Observing only original networks, fits are expectedly low due to small number of networks considered. For the same reason the constant c and exponent γ also differ from the ones in Section 3.1. However, other results show very good fit particularly for original and randomized networks together and reveal a power-law relationship of network size and density (see Fig. 3). (Again, the results are statistically significant at one percent significance level.) As expected due to the size of the networks, the scaling exponent is close to -1 . Number of self-similar scales is higher as in analysis in Section 3.1, since networks are larger and thus reduced in more steps. On the other hand, S does not significantly increase with network size, which implies that renormalization is effective and efficient approach for simplifying large networks.

Fig. 3 illustrates renormalized varieties of three large networks. We consider networks of diverse origin to value how different structure of networks effects the relationship between size and density. For instance, Pennsylvania roads network shows very homogeneous structure, while, on the contrary, other two networks present core-periphery structure typical for social and information networks. However, these diverse network

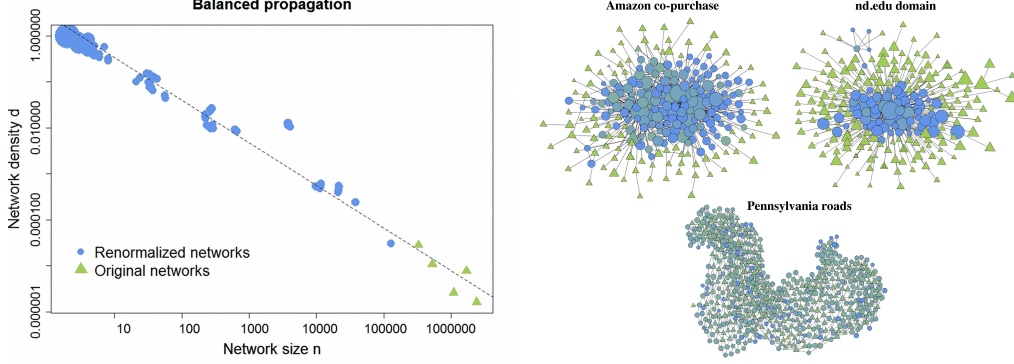


Figure 3: (left) Power-law scaling of network density and size in five real-world networks with millions of links revealed with balanced propagation. Plot shows scaling of density over 10 renormalizations of each network. (Green triangles correspond to original networks, whereas blue circles represent their renormalized varieties. Symbol sizes are proportional to the number of networks with the same size and density.) (right) Density of network structure in renormalized varieties of three large real-world systems of different origin. (Node symbols correspond to degree-corrected clustering coefficient [4] that ranges between 0 and 1—green triangles and blue circles, respectively—while symbol sizes are proportional to the number of nodes in the original network.)

structures do not reflect in the results (Fig. 3, left). Thus, the finding confirms common density scaling in real-world networks irrespective of network type and origin.

Our study improves the comprehension of self-similar structure in real-world networks and implies several possible applications. Firstly, adequate network coarse-graining implies simplification and abstraction of large real-world networks without losing information about original network density. Reduction also enables visualization and improves the comprehension of larger complex networks. Additionally, self-similar density scaling can help at detecting sufficient density according to the size of the sub-graphs in graph sampling applications (e.g., [89]), improve the accuracy of link prediction (for review see [90]) and the quality of synthetic graph generation (e.g., [91]).

4. Conclusions

The paper explores the relationship between size and density of complex real-world networks under different box-covering and community-based renormalization techniques. The analysis was conducted on over 50 real-world networks of various sizes as well as Erdős-Rényi random graphs and different generative graph models. The main contribution of the study is to imply an existence of a scale-free density not only among different real-world networks, but also among their self-similar scales. Common scaling of density thus appears to be a unique property of complex real-world networks irrespective of their type, size and origin. Also, the results reveal balanced propagation based renormalization as the best performing method among the observed algorithms. The study on Erdős-Rényi random graphs, which supposedly exhibit no community structure, validates the above results and confirms that observed scaling of density is distinctive for real-world networks. Hence, our findings expand recent discoveries to other dimensions

of network structure and further improve the comprehension of self-similarity in complex real-world networks. The latter has possible applications in graph sampling, link prediction, synthetic graph generation, network abstraction and visualization.

In our future work we intend to focus on other possible characteristics of density scaling, that could be identified in networks of common type and origin. Furthermore, we will analyze the betweenness centrality scaling with respect to network size in detail. Moreover, the work will also be extended on finding suitable ways for abstracting large real-world networks, while at the same time preserving their fundamental properties.

Acknowledgment

Authors would like to thank Matija Polajnar for providing Slovenian scientists collaboration data and Bojan Klemenc for useful comments. The work has been supported by the Slovene Research Agency *ARRS* within the research program P2-0359.

References

- [1] D. J. Watts, S. H. Strogatz, Collective dynamics of small-world networks, *Nature* 393 (6684) (1998) 440–442.
- [2] M. Faloutsos, P. Faloutsos, C. Faloutsos, On power-law relationships of the Internet topology, *Comput. Commun. Rev.* 29 (4) (1999) 251–262.
- [3] A. L. Barabási, R. Albert, Emergence of scaling in random networks, *Science* 286 (5439) (1999) 509–512.
- [4] S. N. Soffer, A. Vázquez, Network clustering coefficient without degree-correlation biases, *Phys. Rev. E* 71 (5) (2005) 057101.
- [5] R. Albert, H. Jeong, A. Barabási, Error and attack tolerance of complex networks, *Nature* 406 (6794) (2000) 378–382.
- [6] R. Cohen, K. Erez, D. Ben-Avraham, S. Havlin, Resilience of the internet to random breakdowns, *Phys. Rev. Lett.* 85 (21) (2000) 4626–4628.
- [7] M. E. J. Newman, Assortative mixing in networks, *Phys. Rev. Lett.* 89 (20) (2002) 208701.
- [8] M. E. J. Newman, Mixing patterns in networks, *Phys. Rev. E* 67 (2) (2003) 026126.
- [9] M. Girvan, M. E. J. Newman, Community structure in social and biological networks, *P. Natl. Acad. Sci. USA* 99 (12) (2002) 7821–7826.
- [10] E. Ravasz, A. Barabási, Hierarchical organization in complex networks, *Phys. Rev. E* 67 (2) (2003) 026112.
- [11] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, U. Alon, Network motifs: Simple building blocks of complex networks, *Science* 298 (5594) (2001) 824–827.
- [12] Y. Ahn, J. P. Bagrow, S. Lehmann, Link communities reveal multiscale complexity in networks, *Nature* 466 (7307) (2010) 761–764.
- [13] M. E. J. Newman, The physics of networks, *Phys. Today* 61 (11) (2008) 33–38.
- [14] M. E. J. Newman, The structure and function of complex networks, *SIAM Rev.* 45 (2) (2003) 167–256.
- [15] R. Guimerà, L. Danon, A. Díaz-Guilera, F. Giralt, A. Arenas, Self-similar community structure in a network of human interactions, *Phys. Rev. E* 68 (6) (2003) 065103.
- [16] C. Song, S. Havlin, H. A. Makse, Self-similarity of complex networks, *Nature* 433 (7024) (2005) 392–395.
- [17] S. Itzkovitz, R. Levitt, N. Kashtan, R. Milo, M. Itzkovitz, U. Alon, Coarse-graining and self-dissimilarity of complex networks, *Phys. Rev. E* 71 (1) (2005) 016127.
- [18] J. S. Kim, K. I. Goh, B. Kahng, D. Kim, Fractality and self-similarity in scale-free networks, *New J. Phys.* 9 (6) (2007) 177.
- [19] F. Radicchi, J. J. Ramasco, A. Barrat, S. Fortunato, Complex networks renormalization: Flows and fixed points, *Phys. Rev. Lett.* 101 (14) (2008) 148701.
- [20] L. K. Gallos, C. Song, H. A. Makse, A review of fractality and self-similarity in complex networks, *Physica A* 386 (2) (2007) 686–691.

- [21] T. Carletti, S. Righi, Weighted fractal networks, *Physica A* 389 (10) (2010) 2134–2142.
- [22] B. B. Mandelbrot, *The fractal geometry of nature*, W. H. Freeman, 1983.
- [23] A. Bunde, S. Havlin, *Fractals and disordered systems*, Springer-Verlag, Heidelberg, 1996.
- [24] D. Ben-Avraham, S. Havlin, *Diffusion and reactions in fractals and disordered systems*, Cambridge University Press, Cambridge, 2000.
- [25] K. I. Goh, G. Salvi, B. Kahng, D. Kim, Skeleton and fractal scaling in complex networks, *Phys. Rev. Lett.* 96 (1) (2006) 18701.
- [26] C. Song, S. Havlin, H. A. Makse, Origins of fractality in the growth of complex networks, *Nat. Phys.* 2 (4) (2006) 275–281.
- [27] H. D. Rozenfeld, L. K. Gallos, C. Song, H. A. Makse, Fractal and transfractal scale-free networks, e-print arXiv:08082206v1.
- [28] A. L. Barabási, E. Ravasz, T. Vicsek, Deterministic scale-free networks, *Physica A* 299 (3-4) (2001) 559–564.
- [29] C. Song, L. K. Gallos, S. Havlin, H. A. Makse, How to calculate the fractal dimension of a complex network: The box covering algorithm, *J. Stat. Mech.* 2007 (3) (2007) P03006.
- [30] G. Bizhani, V. Sood, M. Paczuski, P. Grassberger, Random sequential renormalization of networks: Application to critical trees, *Phys. Rev. E* 83 (3) (2011) 036110. doi:10.1103/PhysRevE.83.036110.
- [31] S. W. Son, G. Bizhani, C. Christensen, P. Grassberger, M. Paczuski, Irreversible aggregation and network renormalization, *Europhys. Lett.* 95 (5) (2011) 58007.
- [32] Z. Zhang, S. Zhou, T. Zou, G. Chen, Fractal scale-free networks resistant to disease spread, *J. Stat. Mech.* 2008 (9) (2008) P09008.
- [33] G. Csányi, B. Szendrői, The fractal–small-world dichotomy in real-world networks, *Phys. Rev. E* 70 (1-2) (2004) 016122.
- [34] F. Kawasaki, K. Yakubo, Reciprocal relation between the fractal and the small-world properties of complex networks, *Phys. Rev. E* 82 (3) (2010) 036113.
- [35] F. Radicchi, A. Barrat, S. Fortunato, J. J. Ramasco, Renormalization flows in complex networks, *Phys. Rev. E* 79 (2) (2009) 026104.
- [36] M. A. Serrano, D. Krioukov, M. Boguñá, Percolation in self-similar networks, *Phys. Rev. Lett.* 106 (4) (2011) 048701.
- [37] A. Zeng, L. Lu, Coarse graining for synchronization in directed networks, *Phys. Rev. E* 83 (5) (2011) 056123.
- [38] S. Fortunato, Community detection in graphs, *Phys. Rep.* 486 (3-5) (2010) 75–174.
- [39] M. A. Porter, J. P. Onnela, P. J. Mucha, Communities in networks, *Not. Am. Math. Soc.* 56 (9) (2009) 1082–1097.
- [40] P. J. Laurienti, K. E. Joyce, Q. K. Telesford, J. H. Burdette, S. Hayasaka, Universal fractal scaling of self-organized networks, *Physica A* 390 (20) (2011) 3608–3613.
- [41] J. Leskovec, J. Kleinberg, C. Faloutsos, Graphs over time: Densification laws, shrinking diameters and possible explanations, in: *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM, 2005, pp. 177–187.
- [42] J. Leskovec, J. Kleinberg, C. Faloutsos, Graph evolution: Densification and shrinking diameters, *ACM Trans. Knowl. Discov. Data* 1 (1) (2007) 1–40.
- [43] N. Christofides, An algorithm for the chromatic number of a graph, *Computer J.* 14 (1971) 38–39.
- [44] S. Wilf, Backtrack: An $O(1)$ expected time algorithm for the graph coloring problem, *Inf. Proc. Lett.* 18 (1984) 119–121.
- [45] T. H. Cormen, C. E. Leiserson, R. L. Rivest, C. Stein, *Introduction to Algorithms*, MIT Press, 2001.
- [46] W. X. Zhou, Z. Q. Jiang, D. Sornette, Exploring self-similarity of complex cellular networks: The edge-covering method with simulated annealing and log-periodic sampling, *Physica A* 375 (2) (2007) 741–752.
- [47] J. S. Kim, K. I. Goh, B. Kahng, D. Kim, A box-covering algorithm for fractal scaling in scale-free networks, *Chaos* 17 (2) (2007) 026116.
- [48] L. C. Freeman, Centered graphs and the structure of ego networks, *Math. Soc. Sci.* 3 (3) (1982) 291–304.
- [49] M. Everett, S. P. Borgatti, Ego network betweenness, *Soc. Networks* 27 (1) (2005) 31–38.
- [50] W. W. Zachary, An information flow model for conflict and fission in small groups, *J. Anthropol. Res.* 33 (4) (1977) 452–473.
- [51] D. Lusseau, K. Schneider, O. J. Boisseau, P. Haase, E. Slooten, S. M. Dawson, The bottlenose dolphin community of Doubtful Sound features a large proportion of long-lasting associations, *Behav. Ecol. Sociobiol.* 54 (4) (2003) 396–405.

- [52] V. Batagelj, M. A., Pajek datasets, <http://vlado.fmf.uni-lj.si/pub/networks/data/> (2011).
- [53] L. Šubelj, Network data, http://lovro.lpt.fri.uni-lj.si/?navigation=research_support (2011).
- [54] J. Leskovec, D. Huttenlocher, J. Kleinberg, Signed networks in social media, in: Proceedings of the International Conference on Human Factors in Computing Systems, ACM, 2010, pp. 1361–1370.
- [55] M. E. J. Newman, Finding community structure in networks using the eigenvectors of matrices, *Phys. Rev. E* 74 (3) (2006) 036104.
- [56] B. Jones, Computational geometry, <http://compgeom.cs.uiuc.edu/~jeffe/compgeom/> (2002).
- [57] M. Boguñá, R. Pastor-Satorras, A. Díaz-Guilera, A. Arenas, Models of social networks based on social distance attachment, *Phys. Rev. E* 70 (5) (2004) 056122.
- [58] L. A. Adamic, N. Glance, The political blogosphere and the 2004 US election: Divided they blog, in: Proceedings of the KDD Workshop on Link discovery, ACM, 2005, pp. 36–43.
- [59] L. Šubelj, M. Bajec, Ubiquitousness of link-density and link-pattern communities in real-world networks, e-print [arXiv:1104.3162v2](https://arxiv.org/abs/1104.3162).
- [60] J. M. Kleinberg, Authoritative sources in a hyperlinked environment, *J. ACM* 46 (5) (1999) 604–632.
- [61] KDD Cup '03, <http://www.cs.cornell.edu/projects/kddcup/> (2003).
- [62] VAST Challenge '08, <http://www.cs.umd.edu/hcil/VASTchallenge08/> (2008).
- [63] J. Leskovec, K. J. Lang, A. Dasgupta, M. W. Mahoney, Community structure in large networks: Natural cluster sizes and the absence of large well-defined clusters, *Internet Math.* 6 (1) (2009) 29–123.
- [64] D. E. Knuth, *The Stanford GraphBase: A platform for combinatorial computing*, ACM Press, 1993.
- [65] L. Šubelj, M. Bajec, Unfolding communities in large complex networks: Combining defensive and offensive label propagation for core extraction, *Phys. Rev. E* 83 (3) (2011) 036103.
- [66] J. M. Reitz, Online Dictionary of Library and Information Science, <http://vax.wcsu.edu/library/odlis.html> (2002).
- [67] D. L. Nelson, C. L. McEvoy, T. A. Schreiber, University of South Florida free association norms, <http://w3.usf.edu/FreeAssociation/> (2011).
- [68] V. Batagelj, A. Mrvar, M. Zaveršnik, Network analysis of texts, Univ. of Ljubljana, Inst. of Mathematics, Physics and Mechanics, Dep. of Theoretical Computer Science, 2002.
- [69] L. Šubelj, M. Bajec, Community structure of complex software systems: Analysis and applications, *Physica A* 390 (16) (2011) 2968–2975.
- [70] M. E. J. Newman, Network data, <http://www-personal.umich.edu/~mejn/netdata/> (2011).
- [71] L. Šubelj, M. Bajec, Robust network community detection using balanced propagation, *Eur. Phys. J. B* 81 (3) (2011) 353–362.
- [72] D. Bu, Y. Zhao, L. Cai, H. Xue, X. Zhu, H. Lu, J. Zhang, S. Sun, L. Ling, N. Zhang, G. Li, R. Chen, Topological structure analysis of the protein–protein interaction network in budding yeast, *Nucleic Acids Res.* 31 (9) (2003) 2443–2450.
- [73] J. Leskovec, L. A. Adamic, B. A. Huberman, The dynamics of viral marketing, *ACM Trans. Web* 1 (1) (2007) 1–39.
- [74] R. Albert, H. Jeong, A. Barabási, The diameter of the World Wide Web, *Nature* 401 (1999) 130–131.
- [75] G. Tibély, L. Kovanen, M. Karsai, K. Kaski, J. Kertész, J. Saramäki, Communities and beyond: Mesoscopic analysis of a large social network with complementary methods, *Phys. Rev. E* 83 (5) (2011) 056125.
- [76] L. Hébert-Dufresne, A. Allard, V. Marceau, P. Noël, L. J. Dubé, Structural preferential attachment: Network organization beyond the link, e-print [arXiv:1105.5980v2](https://arxiv.org/abs/1105.5980).
- [77] U. N. Raghavan, R. Albert, S. Kumara, Near linear time algorithm to detect community structures in large-scale networks, *Phys. Rev. E* 76 (3) (2007) 036106.
- [78] M. E. J. Newman, M. Girvan, Finding and evaluating community structure in networks, *Phys. Rev. E* 69 (2) (2004) 026113.
- [79] A. Clauset, M. E. J. Newman, C. Moore, Finding community structure in very large networks, *Phys. Rev. E* 70 (6) (2004) 066111.
- [80] S. Fortunato, M. Barthélemy, Resolution limit in community detection, *P. Natl. Acad. Sci. USA* 104 (1) (2007) 36–41.
- [81] B. H. Good, Y. A. De Montjoye, A. Clauset, Performance of modularity maximization in practical contexts, *Phys. Rev. E* 81 (4) (2010) 046106.
- [82] B. Bollobás, *Modern graph theory*, Springer, 1998.
- [83] P. Erdős, A. Rényi, On random graphs I, *Publ. Math. Debrecen* 6 (1959) 290–297.
- [84] A. Lancichinetti, S. Fortunato, Community detection algorithms: A comparative analysis, *Phys. Rev. E* 80 (5) (2009) 056117.

- [85] L. Freeman, Centrality in social networks conceptual clarification, *Soc. networks* 1 (3) (1979) 215–239.
- [86] S. Wasserman, *Social network analysis: Methods and applications*, Cambridge university press, 1994.
- [87] D. Eppstein, J. Wang, A steady state model for graph power laws, in: *Proceedings of the WWW Workshop on Web Dynamics*, 2002, pp. 1–8.
- [88] J. M. Kleinberg, Navigation in a small world, *Nature* 406 (6798) (2000) 845.
- [89] J. Leskovec, C. Faloutsos, Sampling from large graphs, in: *Proceedings of the 12th ACM SIGKDD International conference on Knowledge Discovery and Data Mining*, ACM, 2006, pp. 631–636.
- [90] L. Lü, T. Zhou, Link prediction in complex networks: A survey, *Physica A* 390 (6) (2011) 1150–1170.
- [91] L. S. Heath, N. Parikh, Generating random graphs with tunable clustering coefficient, *Physica A* 390 (23-24) (2011) 4577–4587.