Universal fractal scaling of self-organized networks

Paul J. Laurienti,[1] Karen E. Joyce,[2] Qawi K. Telesford,[2] Jonathan H. Burdette,[1] and Satoru Hayasaka[3, 1, *]

[1] *Department of Radiology, Wake Forest University Health Sciences, Winston–Salem, North Carolina, 27157, USA*
[2] *Department of Biomedical Engineering, Wake Forest University Health Sciences, Winston–Salem, North Carolina, 27157, USA*
[3] *Department of Biostatistical Sciences, Wake Forest University Health Sciences, Winston–Salem, North Carolina, 27157, USA*

(Dated: November 10, 2010)

There is an abundance of literature on complex networks describing a variety of relationships among units in social, biological, and technological systems. Such networks, consisting of interconnected nodes, are often self-organized, naturally emerging without any overarching designs on topological structure yet enabling efficient interactions among nodes. Here we show that the number of nodes and the density of connections in such self-organized networks exhibit a power law relationship. We examined the size and connection density of 47 self-organizing networks of various biological, social, and technological origins, and found that the size-density relationship follows a fractal relationship spanning over 6 orders of magnitude. This finding indicates that there is an optimal connection density in self-organized networks following fractal scaling regardless of their sizes.

PACS numbers: 89.75.Da, 89.75.Hc, 89.75.Fb

There has been considerable interest in the organization of complex networks since the descriptions of small-world [1] and scale-free [2] networks at the end of the 1990s. Of particular interest are naturally occurring complex networks based on self-organizing principles [2]. In particular, self-organized processes have been shown to exhibit some scale-free and fractal behaviors [2, 3]. Barabási and colleagues demonstrated that scale-free degree distributions in many self-organized networks [2, 4–6], which has sparked a great debate [7–9] on the actual existence of scale-free behavior in naturally occurring networks. Although the degree distributions of many networks were initially considered to follow power law distributions [9–13], severe truncation has often been observed [14]. Nevertheless, it is intriguing that self-organized networks can exhibit scale-free degree distributions, and this has led scientists to the search for universality within self-organized systems.

The literature on network organization encompasses a broad range of disciplines and disparate types of networks. The literature boasts networks that range from email communications to protein interactions to word frequencies in texts. The number of nodes and the density of connections in these networks span multiple orders of magnitude, complicating comparisons of metrics extracted from various studies. One common characteristic, however, is that the majority of them are self-organized–from social to technological to biological networks, the interactions between the nodes were not predetermined by a top-down blueprint design.

The work reported here describes a universal relationship between network size (the number of nodes, N) and connection density (the ratio of the number of existing edges to the number of all possible connections, d) across various types of systems. Network parameters from 47 unique networks were collected from the literature or publicly available databases. TABLE I lists N and d, as well as the average node degree K and the total number of edges m from these networks. Although d and/or K have been reported in some of these networks, these metrics are recalculated based on N and m for consistency. Namely, we use the formulae $d = 2m/N(N-1)$ and $K = 2m/N$. Although some of the networks are directed networks, we use the formulae for undirected networks in order to focus on the density of connections regardless of their directions. Note that, from these formulae, the relationship between K and d can be expressed as $d = K/(N-1)$. If N is sufficiently large, $d \simeq K/N$.

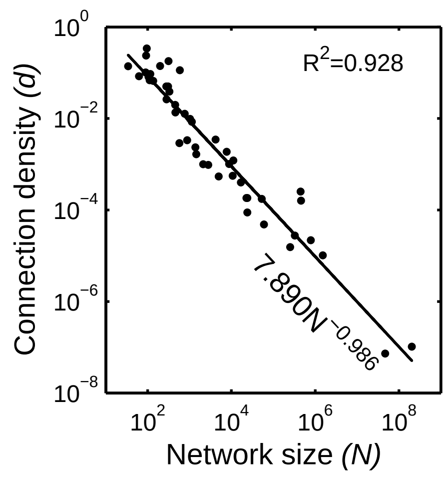

FIG. 1: Log-log plot of the relationship between the number of nodes in a network (network size, N) and the density of connections (d). Each point represents a different network based on the previous literature. The fit shows a power-law relationship that spans more than 6 orders of magnitude with an exponent of -0.986 consistent with a scale-free fractal behavior.

*Electronic address: shayasak@wfubmc.edu

TABLE I: Networks considered in the size-density relationship analysis. Parameters culled from the literature are network size (N) and the number of edges (m). The network density (d) and the mean degree (K) are calculated based on N and m. Networks marked with ‡ contain directed connections. All other networks contain undirected connections. Note that the functional cortical connectivity network was generated by applying a threshold to a correlation matrix, yielding a network that had the edge density of approximately 10%.

Network classes	Networks	N	m	d	K	References
Biological	C. Elegans metabolic	453	2033	1.986×10^{-2}	9.0	[15]
	C. Elegans neural	277	1918	5.018×10^{-2}	13.8	[16, 17]
	E. Coli reaction	315	8915	0.180	56.6	[16, 18]
	E. Coli substrate	282	1036	2.615×10^{-2}	7.3	[16, 18]
	Freshwater food web ‡	92	997	0.238	21.7	[19, 20]
	Functional cortical connectivity	90	405	0.101	9.0	[16, 21]
	Macaque cortex	95	1522	0.341	32.0	[16, 17]
	Marine food web ‡	135	598	6.611×10^{-2}	8.9	[19, 22]
	Metabolic network	765	3686	1.261×10^{-2}	9.6	[4, 19]
	Neural network ‡	307	2359	5.022×10^{-2}	15.4	[19]
	Yeast protein interactions	2115	2240	1.002×10^{-3}	2.1	[19, 23]
Information	Altavista ‡	2.035×10^{8}	2.130×10^{9}	1.028×10^{-7}	20.9	[19]
	Book purchases	105	441	8.077×10^{-2}	8.4	[16]
	Citation ‡	783339	6.716×10^{6}	2.189×10^{-5}	17.1	[19]
	Rogets thesaurus ‡	1022	5103	9.781×10^{-3}	10.0	[24]
	Word adjacency	112	425	6.837×10^{-2}	7.6	[16, 25]
	Word co-occurrence	460902	1.70×10^{7}	1.601×10^{-4}	73.8	[19]
	WWW nd.edu ‡	325729	1.470×10^{6}	2.770×10^{-5}	9.0	[26]
Social	Biology co-authorship	1.520×10^{6}	1.180×10^{7}	1.021×10^{-5}	15.5	[19, 27]
	Company directors	7673	55392	1.882×10^{-3}	14.4	[19, 28]
	Dolphins	62	159	8.408×10^{-2}	5.1	[16, 29]
	Email URV	1133	5452	8.502×10^{-3}	9.6	[30]
	Email messages ‡	59912	86300	4.809×10^{-5}	2.9	[19, 31]
	Email address book ‡	16881	57029	4.003×10^{-4}	6.8	[19, 32]
	Film actors	449913	2.552×10^{7}	2.521×10^{-4}	113.4	[19]
	Football	115	613	9.352×10^{-2}	10.7	[33]
	German directors	4185	30438	3.477×10^{-3}	14.5	[16, 34]
	Jazz	198	2742	0.141	27.7	[35]
	Karate	34	78	0.139	4.6	[36]
	Math co-authorship	253339	496489	1.547×10^{-5}	3.9	[19, 37]
	Newspaper article co-occurrence	459	1422	1.353×10^{-2}	6.2	[16, 38]
	Physics co-authorship	52909	245300	1.753×10^{-4}	9.3	[19, 27]
	Student relationships	573	477	2.911×10^{-3}	1.7	[19, 39]
	Telephone calls ‡	4.7×10^{7}	8.0×10^{7}	7.243×10^{-8}	3.4	[19]
	UK directors	8850	39741	1.015×10^{-3}	9.0	[16, 34]
	US directors	11057	74414	1.217×10^{-3}	13.5	[16, 34]
Technological	Electronic circuits	24097	53248	1.834×10^{-4}	4.4	[19, 40]
	Internet (1998)	10697	31992	5.592×10^{-4}	6.0	[11, 19]
	Internet (2006)	22963	48436	1.837×10^{-4}	4.2	[41]
	Peer-to-peer network	880	1296	3.351×10^{-3}	2.9	[19, 42]
	Power grid (EU)	2783	3762	9.718×10^{-4}	2.7	[43]
	Power grid (US)	4941	6594	5.403×10^{-4}	2.7	[1, 19]
	Software classes ‡	1377	2213	2.336×10^{-3}	3.2	[19, 44]
	Software packages ‡	1439	1723	1.665×10^{-3}	2.4	[19, 45]
	Train routes	587	19603	0.114	66.8	[19, 46]
	Trans-European gas network	24010	25554	8.866×10^{-5}	2.1	[47]
	US airlines	332	2126	3.869×10^{-2}	12.8	[48]

When the network size N and the connection density d are plotted on a log-log plot (FIG 1), there is an obvious linear relationship between the variables. The fit to the data ($d = 7.890N^{-0.986}$) reveals a power law relationship between the size and density of the networks. The scaling exponent approaches negative one (-1), indicating that the relationship is fractal in nature with $1/f$ properties. Despite the wide variety of networks, there is a pronounced power law relationship between the size and the density covering more than 6 orders of magnitude.

The fit to the data is very strong ($R^2 = 0.928$ on $\log_{10}$ transformed data), and there is no indication of truncation at the very large network sizes. It can be seen from FIG 1 that there are two extraordinarily large networks included in the analysis. These networks demonstrate that there is no truncation of the relationship at the extreme values. Even when these networks are removed, the correlation remains very strong ($R^2 = 0.893$) and the exponent is -0.978. Thus, these two points are not unduly influencing the analysis.

When N is sufficiently large, a consequence of a power-law relationship $d \propto N^{-1}$ is that K does not depend on N. This stems from the relationship $d \simeq K/N$, which can be rewritten as $K \simeq dN = cN^{-1}N = c$ where c is a constant. Since our observation above indicates a power-law relationship between d and N with the exponent approximately -1, the scatter plot of K and N does not indicate any association between them (see FIG 2). In other words, a large network size in terms of N is not necessarily associated with large K. It is interesting to note, in FIG 2, that there seems to be a small number of networks with unusually large K compared to the other networks. This is likely a consequence of the mean degree K having a long-tail distribution, as seen in its cumulative distribution plot in FIG 3. In this type of distributions, outliers such as $K > 50$ are likely to occur while the vast majority of K is reasonably small and similar. These outliers seem to occur over the range of N, indicating that such outliers occur randomly without any systematic dependance on N.

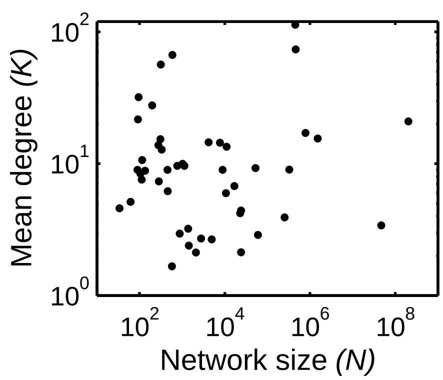

FIG. 2: A scatter plot of the mean degree K of various networks plotted against their network size N. Surprisingly, K does not change systematically over 6 orders of magnitude of N.

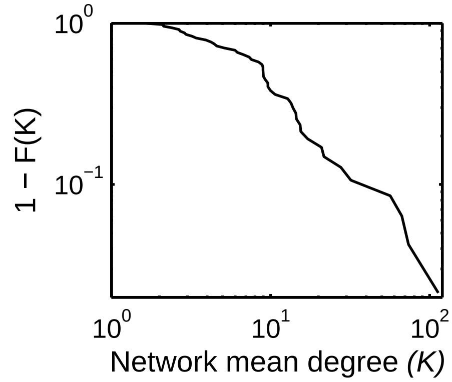

FIG. 3: The complementary cumulative distribution ($1 - F(K)$) of the mean degree K. The distribution exhibits a profile of a long-tail distribution despite the limited number of observations (47 networks).

The findings reported here demonstrate a universal relationship in self-organized networks such that the network size dictates the density. The fractal behavior observed is of particular interest because it indicates that self-organized networks are critically organized. The number of connections within each network is scaled to the size of the network, and this universal behavior likely represents an optimal organization that ensures maximal capacity at a minimal cost. Furthermore, the critical organization would indicate that a density reduction would decrease the communication capabilities of the system. Interestingly, this relationship maintains the mean degree K approximately constant across different network sizes. A similar finding has been reported on the mean degree of the gas and power networks from European countries despite the large disparity in the network size [43, 47]. Our findings further generalizes the constant mean degree K in a variety of network types. It should be noted that the relationship $d \propto N^{-1}$ is not expected from the relationship $d \simeq K/N$ alone, as K could also depend on N rather than being constant. To the best of our knowledge, this work is the first to demonstrate the power-law relationship between d and N, and consequently K being almost constant over 6 orders of magnitude of N.

It is true that one could artificially generate networks that do not exhibit the size-density relationship found above. In fact, the literature contains such artificially generated networks that do not lie near our plotted line. However, such artificially created networks probably do not have real world relevance. We show here the scale-free relationship between network size and connection density in real networks from such diverse origins, supporting the notion of a universal law for network organization.

While replication of these findings from additional networks will be important, there are a number of practical implications of these findings. First, the construction of networks is inherently limited by the sampling procedure used to identify nodes and links. If a self-organized network is found to disobey this relationship, one should seriously consider that there was a bias in the sampling of the network structure. Second, when building artificial networks to be compared to naturally occurring systems, the size-density relationship should roughly follow the $1/f$ relationship. For example, in studies of functional brain networks, cross-correlation matrices of nodal time series are often thresholded to identify links between nodes [49]. The optimal threshold to be applied is not known, and the typical solution is to utilize multiple thresholds [50] producing networks with various densities. Based on the findings presented here, an optimal threshold can be easily determined, resulting in a network following the $1/f$ size-density relationship. Finally, engineered networks for practical applications may realize an optimal cost-benefit trade-off by ensuring that the density of connections is appropriate for the network size.

We show an important, apparently universal feature of self-organized networks: fractal scaling of size and density of connections. This fractal scaling is independent of network types, as the analysis spanned a wide gamut of networks, including biological, information, social, and technological. Thus, it appears that there is an underlying principle to organizing these self-emergent networks, a principle that probably ensures optimal network functioning.

This work was supported in part by the National Institute of Neurological Disorders and Stroke (NS070917, NS056272, and NS039426-09S1) and the Translational Science Institute of Wake Forest University (TSI-K12).

---

[1] D. J. Watts and S. H. Strogatz, Nature **393**, 440 (1998).
[2] A. L. Barabasi and R. Albert, Science **286**, 509 (1999).
[3] P. Bak, C. Tang, and K. Wiesenfeld, Phys Rev Lett **59**, 381 (1987).
[4] H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, and A. L. Barabasi, Nature **407**, 651 (2000).
[5] R. Albert, H. Jeong, and A. L. Barabasi, Nature **406**, 378 (2000).
[6] A. L. Barabasi and E. Bonabeau, Sci Am **288**, 60 (2003).
[7] L. A. Amaral, A. Scala, M. Barthelemy, and H. E. Stanley, Proc Natl Acad Sci U S A **97**, 11149 (2000).
[8] E. F. Keller, Bioessays **27**, 1060 (2005).
[9] T. Gisiger, Biol Rev Camb Philos Soc **76**, 161 (2001).
[10] A. Clauset, C. Shalizi, and M. E. J. Newman, arXiv:0706.1062v1 (2008).
[11] M. Faloutsos, P. Faloutsos, and C. Faloutsos, in *SIGCOMM 1999* (Cambridge, MA, 1999).
[12] M. Mitzenmacher, Internet Mathematics **2**, 525 (2006).
[13] M. E. J. Newman, Contemporary Physics **46**, 323 (2005).
[14] S. Mossa, M. Barthélémy, H. E. Stanley, and L. A. Nunes Amaral, Phys. Rev. Lett. **88**, 138701 (2002).
[15] J. Duch and A. Arenas, Phys Rev E **72**, 027104 (2005).
[16] M. D. Humphries and K. Gurney, PLoS ONE **3**, e0002051 (2008).
[17] M. Kaiser and C. C. Hilgetag, PLoS Comput Biol **2**, e95 (2006).
[18] A. Wagner and D. A. Fell, Proc Biol Sci **268**, 1803 (2001).
[19] M. E. J. Newman, SIAM Review **45**, 167 (2003).
[20] N. D. Martinez, Ecol Monogr **61**, 367 (1991).
[21] S. Achard, R. Salvador, B. Whitcher, J. Suckling, and E. Bullmore, J Neurosci **26**, 63 (2006).
[22] M. Huxham, S. Beaney, and D. Raffaelli, Oikos **76**, 284 (1996).
[23] H. Jeong, S. P. Mason, A. L. Barabasi, and Z. N. Oltvai, Nature **411**, 41 (2001).
[24] D. E. Knuth, *The Stanford GraphBase: A Platform for Combinatorial Computing* (Addison-Wesley, New York, 1993).
[25] M. E. J. Newman, Phys Rev E **74**, 036104 (2006).
[26] R. Albert, H. Jeong, and A. L. Barabasi, Nature **401**, 130 (1999).
[27] M. E. J. Newman, Proc Natl Acad Sci USA **98**, 404 (2001).
[28] M. E. J. Newman, S. H. Strogatz, and D. J. Watts, Phys Rev E **64**, 026118 (2001).
[29] D. Lusseau, Proc Biol Sci **270**, S186 (2003).
[30] R. Guimera, L. Danon, A. Diaz-Guilera, F. Giralt, and A. Arenas, Phys Rev E **68**, 065103(R) (2003).
[31] H. Ebel, L. I. Mielsch, and S. Bornholdt, Phys Rev E **66**, 035103 (2002).
[32] M. E. J. Newman, S. Forrest, and J. Balthrop, Phys Rev E **66**, 035101 (2002).
[33] M. Girvan and M. E. J. Newman, Proc Natl Acad Sci USA **99**, 7821 (2002).
[34] M. J. Conyon and M. R. Muldoon, J Bus Finan Account **33**, 1321 (2006).
[35] P. M. Gleiser and L. Danon, arXiv:cond-mat p. 0307434v2 (2003).
[36] W. Zachary, Journal of Anthropological Research **33**, 452 (1977).
[37] R. d. Castro and J. W. Grossman, Math Intelligencer **21**, 51 (1999).
[38] A. Ozgur and H. Bingol, Computer and Information Sciences - ISCIS **LNCS3280**, 688 (2004).
[39] P. S. Bearman, J. Moody, and K. Stovel, Am J Sociol **110**, 44 (2004).
[40] R. F. i. Cancho, C. Janssen, and R. V. Solé, Phys. Rev. E **64**, 046119 (2001).
[41] M. E. J. Newman, *Unpublished network data available from: http://www-personal.umich.edu/∼mejn/netdata/* (2006).
[42] L. A. Adamic, R. M. Lukose, A. R. Puniyani, and B. A. Huberman, Phys Rev E **64**, 046135 (2001).
[43] R. V. Solé, M. Rosas-Casals, B. Corominas-Murtra, and S. Valverde, Phys. Rev. E **77**, 026102 (2008).
[44] S. Valverde, R. F. Cancho, and R. V. Sole, Europhys Lett **60**, 512 (2002).
[45] M. E. J. Newman, Phys Rev E **67**, 026126 (2003).
[46] P. Sen, S. Dasgupta, A. Chatterjee, P. A. Sreeram, G. Mukherjee, and S. S. Manna, Phys Rev E **67**, 036106 (2003).
[47] R. Carvalho, L. Buzna, F. Bono, E. Gutierrez, W. Just, and D. Arrowsmith, Phys Rev E Stat Nonlin Soft Matter Phys **80**, 016106 (2009).
[48] V. Batagelj and A. Mrvar, *Pajek datasets available from http://vlado.fmf.uni-lj.si/pub/networks/data/* (2006).
[49] E. Bullmore and O. Sporns, Nature Review Neuroscience **10**, 186 (2009).
[50] S. Hayasaka and P. J. Laurienti, Neuroimage **50**, 499 (2010).