



OPEN

# Scaling behaviours in the growth of networked systems and their geometric origins

SUBJECT AREAS:  
COMPLEX NETWORKS  
STATISTICAL PHYSICS

Jiang Zhang<sup>1</sup>, Xintong Li<sup>1</sup>, Xinran Wang<sup>2</sup>, Wen-Xu Wang<sup>1</sup> & Lingfei Wu<sup>3,4</sup>

Received  
16 December 2014

Accepted  
17 March 2015

Published  
29 April 2015

Correspondence and requests for materials should be addressed to J.Z. (zhangjiang@bnu.edu.cn) or W.-X.W. (wenxuwang@bnu.edu.cn)

<sup>1</sup>School of Systems Science, Beijing Normal University, <sup>2</sup>College of Resources Science and Technology, Beijing Normal University, <sup>3</sup>School of Human Evolution and Social Change, Arizona State University, <sup>4</sup>Center for the Study of Institutional Diversity, Arizona State University.

Two classes of scaling behaviours, namely the super-linear scaling of links or activities, and the sub-linear scaling of area, diversity, or time elapsed with respect to size have been found to prevail in the growth of complex networked systems. Despite some pioneering modelling approaches proposed for specific systems, whether there exists some general mechanisms that account for the origins of such scaling behaviours in different contexts, especially in socioeconomic systems, remains an open question. We address this problem by introducing a geometric network model without free parameter, finding that both super-linear and sub-linear scaling behaviours can be simultaneously reproduced and that the scaling exponents are exclusively determined by the dimension of the Euclidean space in which the network is embedded. We implement some realistic extensions to the basic model to offer more accurate predictions for cities of various scaling behaviours and the Zipf distribution reported in the literature and observed in our empirical studies. All of the empirical results can be precisely recovered by our model with analytical predictions of all major properties. By virtue of these general findings concerning scaling behaviour, our models with simple mechanisms gain new insights into the evolution and development of complex networked systems.

Discovered quite recently, but with roots that go back to decades ago in biology and ecology<sup>1–4</sup>, the scaling behaviour is generally shared by a variety of complex networked systems, such as cities<sup>5–8</sup>, online communities<sup>9–11</sup>, and complex networks<sup>12</sup>. The scaling behaviours in the form

$$X \sim N^\gamma \quad (1)$$

capture the fact that some macroscopic variables  $X$  scale with the system size  $N$ , where  $N$  represents the number of entities in a networked system. Based on empirical observations of the exponent  $\gamma$ , two categories, super-linear and sub-linear scaling behaviours, have been identified.

For instance, in cities, if  $X$  represents the gross domestic product (GDP), the total wage, or the number of crimes, and  $N$  is the population of a city, a super-linear scaling with  $\gamma \approx 1.17^{5-7,13}$  is found.  $\gamma > 1$  indicates that as a city grows, the total number of interactions increases at a faster rate, leading to more wealth and innovation per capita, but also with more crime and pollution as side effects. In online communities,  $X$  represents the total number of activities of users, such as clicks, micro-blogs and tags<sup>9–11</sup>, and  $N$  is the total number of active users.  $\gamma$  lies in the range [1.17–1.48]. The super-linear scaling behaviours have also been observed in complex networks, but in this context, it is termed accelerating growth<sup>12,14</sup> or densification<sup>15</sup>, and is characterized by a higher generating rate for links than for nodes during network growth. This phenomenon is found in scientific collaborations<sup>16</sup>, citation network<sup>15</sup>, Internet autonomous system<sup>15,17</sup>, food webs<sup>18</sup>, and neural networks<sup>19</sup>.

Sub-linear scaling as another type of scaling behaviour with exponent smaller than 1 is also prevalent in complex networked systems. For instance, the area and road volume of a city are found to scale sub-linearly with its population<sup>5</sup>. In online communities, the number of distinct tags scales sub-linearly with the system size<sup>9</sup>. In information retrieval or in linguistics, the general Heap's law captures a sub-linear scaling between the number of distinct words and the total number of words<sup>20,21</sup>. In ecological communities, the diversity of higher taxa scales sub-linearly with the number of species<sup>22</sup>. The sub-linear scaling is also found in river networks<sup>23</sup>, various combinatorial systems<sup>24</sup> and etc.

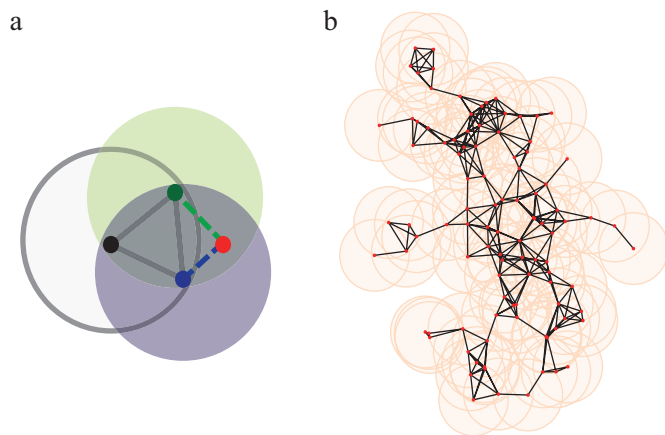
Although some effort has been dedicated to explaining the scaling behaviours based on network approaches<sup>2,3,25</sup>, forest-fire model<sup>15</sup>, random walks<sup>9</sup>, Kronecker graphs<sup>26</sup> and the recently developed city model<sup>27</sup>, the origins of the scaling behaviours have not been fully understood yet. In particular, the general findings in diverse systems prompts us to wonder if there exists some simple but universal underlying mechanism that



accounts for both classes of scaling behaviours. In this paper, we propose a geometrical network model to address this fundamental problem. Partially inspired by the hyperbolic space model for construction of scale-free networks<sup>28</sup> and the hidden geometry of complex networks<sup>29</sup>, we propose a spatial-constrained attachment (SCA) model to uncover the origins of both super-linear and sub-linear scaling behaviours in geometric space of arbitrary dimension. The self-organized phenomena produced by our SCA model are in good agreement with a variety of empirical findings, including the scaling behaviours governing the entire number of links, the time elapsed, and the volume versus the system size measured as the number of nodes in the network. The simple mechanism of SCA allows us to derive analytical results for all critical network properties, such as the accelerating growth, degree distributions of nodes, and the clustering coefficient. We also slightly modify the model by considering some realistic restrictions to better mimic real situations. We apply our model to online communities, a citation network and nighttime light clusters, finding that the empirical observations from these cases, including the scaling behaviours, Zipf size distribution<sup>30</sup>, and aggregation patterns, can be quantitatively reproduced.

## Results

**The Model.** Our model assumes that a geometric graph grows in an  $L^d$  hypercube embedded in  $d$ -dimensional Euclidean space<sup>31–33</sup> according to the SCA mechanism, where  $L$  is the linear size of the hypercube and  $d$  is the spatial dimension. Suppose that initially, a single node is present at the center of the hypercube as the seed of the growing graph. At each time step  $t$ , one new node  $P$  is generated, and its coordinates (location)  $x_P = (x_1, x_2, \dots, x_d)$  are randomly selected with equal probability of being located anywhere throughout the hypercube. However,  $P$  can survive only if  $P$  matches with at least one of the existing node  $Q$  with location  $x_Q$ , such that  $|x_P - x_Q| < r$ , where  $r$  is a given parameter and  $|\cdot|$  is the Euclidean distance. The surviving node  $P$  is then attached to all the existing nodes that belong to the  $r$ -radius ball of  $P$ . In other words, interactions can be built exclusively within a certain area, limited by parameter  $r$ . This process is repeated until a network of the desired size is achieved (see Fig. 1). The geometric network grows at increasing speed in the sense that a newly added node is more likely to attach to existing nodes as the network volume increases. The central area around the seed is of higher density than the other areas because “dense gets denser.”



**Figure 1 | An illustration of the basic model in  $d = 2$ -dimensional space.** (a) The filled disks represent existing nodes, the red disk represents a newly added node that will survive. The dark lines represent existing links, and the dashed lines represent the newly added links. The shaded areas represent the interaction regions of existing nodes. (b) A simulation of the basic model after  $t = 3625$  steps and  $N(t) = 100$  nodes. The shaded area represents  $V(t)$  in two-dimensional space. In all simulations  $r = 1$ .

Although the evolution of complex networked systems in different disciplines is affected by many factors, our model captures some underlying mechanisms that are shared among a variety of systems with spatial constraint and play significant roles in the rising of scaling behaviours. Here, space can be classified into real geographical or physical spaces and, abstract spaces. Cities<sup>27</sup>, Internet autonomous systems<sup>17</sup> and brain<sup>19</sup> are subject to the former and, similarity spaces (scientific collaborations, citation networks, online communities)<sup>15,16,28,34</sup>, semantic spaces (languages, tagging systems)<sup>9</sup> and niche spaces (ecological communities, food webs)<sup>35</sup> are subject to the latter. In our model, the SCA mechanism uncovers the fact that the survival probability of a new node and its contribution to the growth of interactions are significantly determined by its closeness to others that can be measured by its relative location in a space. Indeed, the establishment of many kinds of interactions is strongly affected by the closeness among entities in a certain space, such as scientific collaborations, citations, social ties in cities and online communities, connections of routers, neuronal connections, and co-occurrence of tags. In general, two closer entities are of higher probability to build a connection between them, resulting from either the cost of establishing connections in physical spaces, or similarity induced connections in abstract spaces. On the other hand, the location of a new node may be determined by a number of factors, precluding us from specifying the exact location of every newly generated node. In this regard, it is reasonable to assume that new nodes are randomly located in the statistical point of view. Both closeness induced connections and random birth of new nodes can be captured by our SCA mechanism. Although it is not sufficiently concrete to model the evolution of every aspect of real networked systems, the SCA mechanism in our model is common in many real systems, accounting for the broad implications of our model in offering deeper understanding of scaling behaviours.

**Analytic Results.** We derive the analytical results of our model in thermodynamic limit ( $L, t \rightarrow \infty$ ). As  $t$  approaches infinity, the spatial shape of the network can be approximated by a symmetrical  $d$ -dimensional ball. The radius  $R(t)$  (which is defined as the maximum distance between any existing nodes and the seed (center)) of the ball at time  $t$  is linearly dependent on  $t$ :

$$R(t) \sim t. \quad (2)$$

It is valid because that the average time between two updates to the nodes at the perimeter of the  $d$ -dimensional ball is approximately  $\sim 1/R(t)$  and there are approximately  $\sim R(t)$  positions at the perimeter to be updated; thus the average speed at which the radius increases is a constant ( $\sim R(t) \cdot 1/R(t)$ , see Supplementary Information Section 1). Then the total volume  $V(t)$  that is filled by all  $d$ -dimensional balls with radius  $r$  can be approximated by  $V(t) \sim t^d$ . Suppose that the density of nodes at any location with spherical coordinate  $(\rho, \Theta)$  is  $\mu(\rho, \Theta, t)$ , where  $\rho \leq R(t)$  is the spherical radius and  $\Theta = (\theta_1, \theta_2, \dots, \theta_{d-1})$  is the vector of spherical angles. Then after time step  $\tau_\rho \sim \rho$ , the density  $\mu(\rho, \Theta, t)$  within the infinitesimal space  $d\sigma$  starts to become non-zero and increases with the constant rate  $d\sigma/L$  until  $t$ . Thus, the total number of nodes in the infinitesimal volume  $d\sigma$  is

$$\int_\tau^t (d\sigma/L^d) dt \sim (t - \tau\rho) d\sigma/L^d, \quad (3)$$

yielding

$$\mu(\rho, \Theta, t) \sim \frac{R(t) - \rho}{L^d}. \quad (4)$$

The density of nodes at radius  $\rho$  is

$$\mu(\rho, t) \sim \frac{(R(t) - \rho)\rho^{d-1}}{L^d}. \quad (5)$$

Consequently, the total number of nodes within the radius  $\rho$  is approximated by  $\rho^d$  if  $\rho < R(t)$ , in agreement with the fractal



dimension  $d$  of the network in the thermodynamic limit<sup>36,37</sup>. Finally, the total number of nodes in the whole network can be calculated by

$$N(t) = \int_0^{R(t)} \mu(\rho, t) d\rho \sim R(t)^{d+1} \sim t^{d+1}. \quad (6)$$

Note that each node at location  $(\rho, \Theta)$  at time  $t$  is connected to  $\mu(\rho, \Theta, t)$  neighbors on average, and there are  $\mu(\rho, \Theta, t) d\sigma$  nodes in total in the infinitesimal space  $d\sigma$ , meaning that the total number of links in  $d\sigma$  is  $v(\rho, \Theta, t) \sim \mu(\rho, \Theta, t)^2$ . Thus, the total number of links in the entire network is

$$E(t) = \iint v(\rho, \Theta, t) d\sigma \sim t^{d+2}. \quad (7)$$

We can reformulate all the variables as functions of the number of nodes at  $t$  to obtain the scaling behaviours. The first scaling, between the total number of edges and nodes, can be obtained by eliminating  $t$  in both Eqs. (6) and (7):

$$E(t) \sim N(t)^{\frac{d+2}{d+1}}. \quad (8)$$

This formula is consistent with Eq. (1) if we consider the total number of interactions to correspond to the number of links and the population is proportional to the number of nodes in a system. We see that the exponent  $\gamma = (d+2)/(d+1)$  depends exclusively on the dimension of the space in which the spatial network is embedded. Another interesting scaling behaviour, between the volume and the size of the network, can be derived by Eq. (6):

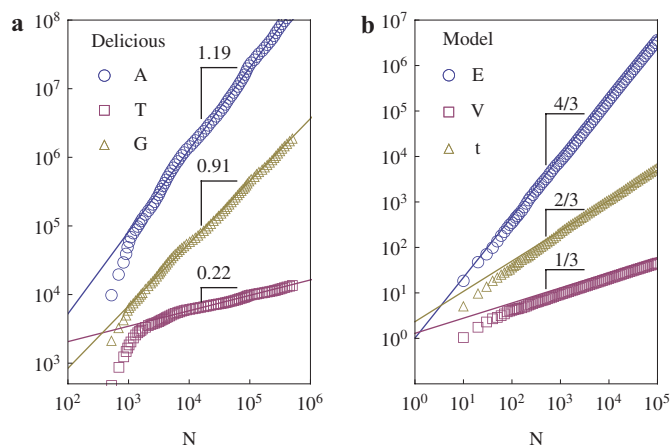
$$V(t) \sim N(t)^{\frac{d}{d+1}}. \quad (9)$$

This scaling behaviour indicates a densification effect because the growth rate of  $V(t)$  is slower than that of  $N(t)$ . As a result, the network becomes more and more compact and dense. For  $d=2$ , the scaling exponent is  $2/3$ , which is in good agreement with the empirically observed scaling between area and population in cities<sup>27,38</sup>. Meanwhile, this scaling also corresponds to the widely observed sub-linear scaling behaviour of diversity (Heap's law) in complex systems if we regard the number of distinct types of entities as the volume in the similarity space<sup>9</sup> (see Fig. 2 (a)). As a bonus, another scaling to describe the accelerating growth phenomenon can be also analytically obtained:

$$t \sim N(t)^{\frac{1}{d+1}}, \quad (10)$$

demonstrating that the growth rate of new nodes increases as the size of the network increases ( $dN(t)/dt \sim N(t)^{\frac{d}{d+1}}$ ). This phenomenon has been empirically observed in online systems (Fig. 2(a)) and the citation network (see Supplementary Information Figure S8). Our theoretical predictions are in good accordance with the numerical results, as demonstrated in Fig. 2 (b). In addition, we also offer analytical results for the node-degree distributions and for the clustering coefficients, that are determined by the distribution of triangles. (The detailed derivations are provided in Supplementary Information Section 1.4).

**Model Extensions.** Although both the super-linear and sub-linear behaviours produced by the basic model are in qualitative agreement with the empirical findings in complex networked systems, a small difference remains between the predicted super-linear scaling exponents  $\gamma$  and the real values (For example, the exponent of the super-linear scaling in cities is 1.17, slightly smaller than  $4/3$ ). The discrepancy stems from the non-negligible sizes of nodes in real scenarios, such as in cities, which prohibits the birth of an infinite number of nodes in a bounded area. Thus, we incorporate the probability of survival for new nodes even when they are properly matched with some existing nodes to better mimic the real situation. As shown later, this additional feature still allows us to derive analytical results and can produce an adjustable scaling exponent.



**Figure 2 | The scaling behaviours of the Delicious community (a) and the model (b).** In (a), the scalings represent the cumulative number of activities  $A$  (blue circles), the total time elapsed  $T$  (for convenience of comparison on the plot; filled purple squares), and the cumulative number of tags used by users  $G$  (yellow triangles) versus the cumulative number of users of the Delicious community from Feb 1, 2003 to Nov 8, 2006. The solid lines represent the best fits. In (b), the scalings represent  $E(t)$  (blue circles),  $V(t)$  (filled purple squares), and  $t$  (yellow triangles) versus  $N(t)$  of the basic model ( $d=2$ ). The solid lines are theoretical predictions of the mean-field approximations (see Supplementary Information Section 1), in all simulations  $r=1$ . We also provide the Flickr munity and the APS citation network as other empirical evidence in Supplementary Information Figure S7 and S8.

**Crowding effect.** A newly generated node can survive with probability  $\mu(\rho, \Theta, t)^{-\alpha}$  if it matches at least one neighboring node (the distance is smaller than  $r$ ), where  $\alpha$  is adjustable and captures the crowding effect. As  $\alpha$  increases, it becomes more difficult for a new node to attach in a dense area. The basic model is recovered if we set  $\alpha$  to be zero.

We provide theoretical predictions for the model that incorporates the crowding effect. Suppose that the node density at a given location is  $\mu(\rho, \Theta, t)$ . The number of nodes in  $d\sigma$  is given by

$$\mu(\rho, \Theta, t) d\sigma = \int_{\tau_\rho}^t \mu(\rho, \Theta, s)^{-\alpha} \frac{d\sigma}{L^d} ds, \quad (11)$$

where  $\tau_\rho$  is the time step when  $\mu(\rho, \Theta, t)$  becomes non-zero. Taking the derivative with respect to  $t$  on both sides of Eq. (11), we obtain a differential equation:

$$\partial \mu / \partial t = \frac{1}{L^d} \mu^{-\alpha}. \quad (12)$$

Solving this equation with the initial condition  $\mu(\rho, \Theta, t_\rho) = 0$ , we obtain

$$\mu(\rho, \Theta, t) \sim (t - \tau_\rho)^{\frac{1}{1+\alpha}}. \quad (13)$$

All scalings are produced with adjustable exponents. To be concrete, the scaling between the number of edges and nodes is

$$E(t) \sim N(t)^{1 + \frac{1}{1+(1+\alpha)d}}, \quad (14)$$

and that between the volume and the number of nodes is

$$V(t) \sim N(t)^{\frac{(1+\alpha)d}{1+(1+\alpha)d}}. \quad (15)$$

Note that if  $\alpha \rightarrow +\infty$ , the exponents in Eqs. (14) and (15) become 1, yielding a linear scaling and a homogenous  $d$ -dimensional regular spatial network. Thus the modified model that incorporates the crowding effects can generally produce scaling behaviours with arbitrary exponent values.





It is interesting to note that the exponents of super-linear scaling and sub-linear scaling in Eqs. (14) and (15) change in different directions when  $\alpha$  changes. Therefore, observing the sub-linear scaling of diversity may imply the existence of the super-linear scaling of productivity or vice versa, and knowing one exponent may predict another one. That implies diversity phenomenon and productivity in systems are two sides of the same coin. Too fast diversity increase or innovation may slow down the rate of interactions and depress the productivity, a slow and continuous innovation process can accelerate interactions and productivity.

Next, we apply our model to cities. Assume that several cities arise in a 2-dimensional space. The development of these cities is modeled with multiple seeds. We introduce a new rule to naturally mimic the sequential emergence of new seeds. To be concrete, each newly added, surviving node has a probability  $\varepsilon \in [0, 1]$  of moving to a random location in the hypercube. By contrast, all existing nodes are not allowed to move. The basic model is recovered for  $\varepsilon = 0$ . If  $\varepsilon = 1$ , the model resembles the conventional random geometric graph model<sup>32</sup>. In general, mobile nodes will become new seeds, around which nodes will aggregate, leading to the formation of densely connected local clusters centered on these seeds.

If  $\pm L$  is sufficiently large so that the clusters are isolated, each cluster can grow independently. In this case, the distribution of cluster sizes can be estimated using the Yule-Simon process. Suppose that there are  $s$  survival nodes and  $N_c$  clusters at  $t$ . We use  $s$  instead of  $t$  as a new "time" index to facilitate our analysis such that if and only if a new node is attached to existent network, the step  $s$  becomes  $s + 1$ . We denote the volume of the cluster  $i$  by  $V_i(s)$ , and the number of nodes in the cluster by  $N_i(s)$ . The probability of a newly added node that attach to cluster  $i$  is  $\Pi = \frac{V_i(s)}{\sum_j V_j(s)}$ . Because all  $N_c$  clusters evolve independently, the volume of the  $i$ th cluster obeys Eq. (15):  $V_i(s) \sim N_i(s)^\eta$ , where  $\eta = (2 + 2\alpha)/(3 + 2\alpha) < 1$ . We thus can write a master equation describing a sub-linear preferential attachment process for all these clusters<sup>39</sup> (see Supplementary Information Section 4.2):

$$s \frac{\partial P_N}{\partial s} = s \frac{1 - \varepsilon}{Z} [(N - 1)^\eta P_{N-1} - N^\eta P_N] + \delta_{N=1} - P_N. \quad (16)$$

where  $P_N$  is the fraction of clusters with  $N$  nodes at step  $s$ ,  $Z = \sum_{N=1}^{\infty} N^\eta P_N$ , and  $\delta$  is the Dirac delta function. By solving this equation, we find that the size distribution as  $s \rightarrow \infty$  is given by<sup>39</sup>

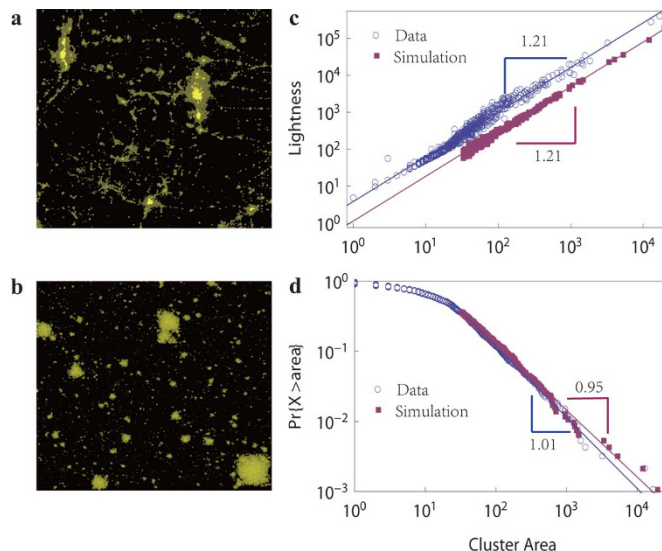
$$P_N \sim N^{-\eta} \exp\left(\frac{-\zeta}{1 - \eta} (N^{1-\eta} - 2^{1-\eta})\right), \quad (17)$$

where  $\zeta$  is the solution of the equation  $(1 - \varepsilon)/\varepsilon = \sum_{k=1}^{\infty} \Pi_{j=1}^k (1 + \zeta/j^\eta)^{-1}$ . This distribution resembles a power law if  $\eta$  is close to 1. However,  $L$  is finite in real situations, leading to certain degree of overlap among clusters and influence on the exponent. The finite-size effect has been explored in detail (see Supplementary Information Section 4.3). The isolation assumption of clusters for deriving analytical results is valid for the scale among cities.

If each cluster corresponds to a city, the model by choosing  $\alpha = 1.5$  and  $\varepsilon = 0.03$  gives rise to the optimal recovery of the empirical results of cities, with scaling exponents<sup>5</sup> and a power law distribution<sup>30</sup> that are nearly identical to the real values, as shown in Fig. 3. Moreover, the spatial pattern of aggregation clusters produced by our model is quite similar to those observed in satellite image of nighttime lights in several cities, as shown in Fig. 3 (see Supplementary Information Section 4.1 for more details).

## Discussion

In summary, we developed a growing geometric graph model to uncover the simple underlying mechanisms that account for the super-linear and sub-linear scaling behaviours that are ubiquitously observed in complex networked systems. Our basic model without free



**Figure 3 | Comparison of modified SCA model with nighttime light data.**

(a) Satellite image of a nighttime light distribution. We suppose that each connected cluster in the image is a natural city. (b) The clusters grown by the model using the modified rules ( $\varepsilon = 0.03, \alpha = 1.5$ ). (c) The scaling between the total area and the total light intensity of these clusters both for the nighttime light image (blue circles) and for the image generated by the model (purple squares). In the model, we take the total number of edges of each cluster to represent its total light intensity. (d) Area distributions of nighttime light clusters (blue circles) and the model's clusters (purple squares). In (c) and (d), only clusters with sizes larger than 33 in simulation are shown for comparison. In all simulations,  $r = 1$ .

parameter is capable of reproducing both categories of scaling behaviours in qualitative agreement with empirical findings, and the scaling exponents are determined exclusively by the dimension of Euclidean space. These results indicate that the SCA plays the primary role in the origins of the scaling behaviours in complex networked systems. However, our model is not limited to Euclidean space. Inserting our SCA mechanism into other spaces, such as hyperbolic space<sup>28</sup>, may offer more general interpretations of generalized scaling behaviours.

To better understand the evolutionary dynamics of real systems, we developed modified models by incorporating crowding effects and multiple seeds with adjustable parameters into the basic model, offering accurate predictions for a variety of empirically observed scaling exponents. For both basic and modified models, we derived analytical results for all important network properties, including the scalings, Zipf distributions, degree distributions and clustering coefficients. The theoretical predictions are in good agreement with the simulated results.

Our approach also offers new insights into the evolution of complex networked systems with respect to the accelerating growth rates of interactions and nodes. Furthermore, our approach provides a deeper understanding of the origins of scaling behaviours in complex networked systems in terms of the trade-off between costs and efficiency, where the former pertains to the interaction density associated with spatial distance and the latter may be measured in terms of topological properties. Taken together, our finding of simple underlying mechanisms that account for the common scaling behaviours in different fields will inspire further effort toward discovering universal rules in the science of complexity.

## Method

**Nighttime light data.** The global satellite image of nighttime lights used in our study is collected by the Operational Linescan System (OLS) of the US Air Force Defense Meteorological Satellite Program (DMSP) and archived at NOAA National Geophysical Data Center (NGDC). The image is 30-arc-second grided and spans from  $-180$  to  $180$  degrees longitude and from  $-65$  to  $75$  degrees latitude. The digital number ( $DN$ ) values of the nighttime lights range from 1 to 63. In addition, although



sunlit data, moonlight, glare, observations containing clouds and lighting features from the aurora are excluded from the DMSP nighttime stable lights dataset, gas flares are not. Therefore, we used the global gas flare map generated by NGDC<sup>40</sup> to identify and remove gas flares, reducing the possibility of mistaking them for urbanized areas.

The year 2009 was chosen because it was the latest product freely accessible when we first conducted our analysis. For detailed comparison between our model simulation results and nighttime light observations, we narrow our scope down to part of the south central contiguous United States. Using GIS software, the nighttime lights image was re-projected into Lambert conic conformal projection, and a 1000 pixels × 1000 pixels region was extracted from the global image. The upper left corner of the region of interest (ROI) is 113.8 W, 42.2 N, upper right 101.7 W, 43.4 N, lower left 111.7 W, 33.5 N and lower right 100.9 W, 34.5 N. In this region, two lighted pixels were considered as connected if one of them is the Moore neighborhood of the other, and all the connected pixels formed a cluster. Thus we identified 921 clusters in Fig. 3(a). For each cluster, we treated the total number of non-zero pixels as the area of the cluster, and the sum of non-zero pixels values as the total light intensity of the cluster. Then, the scaling between light intensity and area as well as the size distribution of the areas of all clusters were calculated to produce Fig. 3(c) in the main text. More details about nighttime light data can be referred to the Supplementary Information Section 4.1.

**Delicious community data.** We compare the scalings of social tagging systems including Delicious and Flickr (see Supplementary Information Section 2) communities to the results of our model. The data can be downloaded freely at <http://www.tagora-project.eu/data/>. In both systems, users visit certain online resources (pictures on Flickr) and may tag them with certain words. We can consider the semantic space of these tags as the space in which our model is established. Then, the number of distinct tags can be regarded as the total volume occupied by users. And the total number of tagging events can be viewed as interactions happened in the system.

- Brown, J. & West, G. *Scaling in Biology* (Oxford University Press, 2000).
- West, G., Brown, J. H. & Enquist, B. J. A general model for the origin of allometric scaling laws in biology. *Science* **276**, 122–126 (1997).
- Banavar, J., Maritan, A. & Rinaldo, A. Size and form in efficient transportation networks. *Nature* **399**, 130–132 (1999).
- Dreyer, O. Allometric scaling and central source systems. *Phys. Rev. Lett.* **87**, 038101 (2001).
- Bettencourt, L. M., Lobo, J., Helbing, D., Kuhnert, C. & West, G. B. Growth, innovation, scaling, and the pace of life in cities. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 7301–7306 (2007).
- Bettencourt, L. M. Invention in the city: increasing returns to scale in metropolitan patenting. *Res. Policy* **36**, 107–120 (2007).
- Bettencourt, L. M. A., Lobo, J., Strumsky, D. & West, G. B. Urban scaling and its deviations: Revealing the structure of wealth, innovation and crime across cities. *PLoS ONE* **5**, e13541 (2010).
- Bettencourt, L. & West, G. A unified theory of urban living. *Nature* **467**, 912–913 (2010).
- Cattuto, C., Barrat, A., Baldassarri, A., Schehr, G. & Loreto, V. Collective dynamics of social annotation. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 10511–10515 (2009).
- Wu, L. & Zhang, J. Accelerating growth and size-dependent distribution of human online activities. *Phys. Rev. E* **84**, 026113 (2011).
- Wu, L. F. The accelerating growth of online tagging systems. *Eur. Phys. J. B* **83**, 283–287 (2011).
- Dorogovtsev, S. N. & Mendes, J. F. F. Accelerated growth of networks. *arXiv: cond-mat/0204102* (2002).
- Schlapfer, M. *et al.* The scaling of human interactions with city size. *J. R. Soc. Interface* **11**, 20130789 (2014).
- Mattick, J. S. & Gagen, M. J. Accelerating networks. *Science* **307**, 856–858 (2005).
- Leskovec, J., Kleinberg, J. & Faloutsos, C. Graphs over time: densification laws, shrinking diameters and possible explanations. In *Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining, KDD '05*, 177–187 (ACM, New York, NY, USA, 2005).
- Barabasi, A. *et al.* Evolution of the social network of scientific collaborations. *Physica A* **311**, 590–614 (2002).
- Papadopoulos, F., Psomas, C. & Krioukov, D. Network mapping by replaying hyperbolic growth. *IEEE/ACM Transactions on Networking* **Early Access Online** (2014).
- Martinez, N. D. Scale-dependent constraints on food-web structure. *Amer. Nat.* **144**, 935–953 (1994).
- Zhang, K. & Sejnowski, T. J. A universal scaling law between gray matter and white matter of cerebral cortex. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 5621–5626 (2000).

- Heaps, H. S. *Information Retrieval: Computational and Theoretical Aspects* (Academic Press, Inc., Orlando, FL, USA, 1978).
- Cattuto, C., Loreto, V. & Pietronero, L. Semiotic dynamics and collaborative tagging. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 1461–1464 (2007).
- Enquist, B. J., Haskell, J. P. & Tiffney, B. H. General patterns of taxonomic and biomass partitioning in extant and fossil plant communities. *Nature* **419**, 610–613 (2002).
- Dodds, P. S. & Rothman, D. H. Scaling, universality, and geomorphology. *Annu. Rev. Earth Planet. Sci.* **28**, 571–610 (2000).
- Changizi, M. A. Universal scaling laws for hierarchical complexity in languages, organisms, behaviors and other combinatorial systems. *J. Theor. Bio.* **211**, 277–295 (2001).
- Arbesman, S., Kleinberg, J. M. & Strogatz, S. H. Superlinear scaling for innovation in cities. *Phys. Rev. E* **79**, 016115 (2009).
- Leskovec, J., Chakrabarti, D., Kleinberg, J., Faloutsos, C. & Ghahramani, Z. Kronecker graphs: An approach to modeling networks. *J. Mach. Learn. Res.* **11**, 985–1042 (2010).
- Bettencourt, L. M. A. The origins of scaling in cities. *Science* **340**, 1438–1441 (2013).
- Papadopoulos, F., Kitsak, M., Serrano, A., Boguna, M. & Krioukov, D. Popularity versus similarity in growing networks. *Nature* **489**, 537–540 (2012).
- Brockmann, D. & Helbing, D. The hidden geometry of complex, network-driven contagion phenomena. *Science* **342**, 1337–1342 (2013).
- Gabaix, X. Zipf's law for cities: An explanation. *Q. J. Econ.* **114**, 739–767 (1999).
- Barthelemy, M. Spatial networks. *Phys. Rep.* **499**, 1–101 (2011).
- Penrose, M. *Random Geometric Graphs* (Oxford University Press, 2003).
- Li, D., Kosmidis, K., Bunde, A. & Havlin, S. Dimension of spatially embedded networks. *Nat. Phys.* **7**, 481–484 (2011).
- Jacob, E. & Mrters, P. A spatial preferential attachment model with local clustering. In *Algorithms and Models for the Web Graph*, vol. 8305, 14–25 (Springer International Publishing, 2013).
- Williams, R. J. & Martinez, N. Simple rules yield complex food webs. *Nature* **404**, 180–183 (2000).
- Vicsek, T. *Fractal Growth Phenomena: 2nd Edition* (World Scientific Publishing Company, 1992).
- Barabasi, A.-L. & Stanley, H. E. *Fractal Concepts in Surface Growth* (Cambridge University Press, 1995).
- Nordbeck, S. Urban allometric growth. *Geogr. Ann. Ser. B.* **53**, 54–67 (1971).
- Krapivsky, P. L., Redner, S. & Leyvraz, F. Connectivity of growing random networks. *Phys. Rev. Lett.* **85**, 4629–4632 (2000).
- Elvidge, C. D. *et al.* A fifteen year record of global natural gas flaring derived from satellite data. *Energies* **2**, 595–622 (2009).

## Acknowledgments

We thank Prof. Y. G. Wang, Q. H. Chen, Z. Di, J.S. Wu, and X. S. Chen for valuable discussions. The paper is supported by the NNSFC under Grant No.611004107, No. 61174165 and No. 11105011, and supported by the Beijing Higher Education Young Elite Teacher Project under the Grant No.YETP0291.

## Author contributions

JZ conceives and implements the experiments, XL, XW, and LW collect and analyse the empirical data. WW writes the manuscript. All authors reviewed the manuscript.

## Additional information

Supplementary information accompanies this paper at <http://www.nature.com/scientificreports>

**Competing Interests:** The authors declare no competing financial interests.

**How to cite this article:** Zhang, J., Li, X., Wang, X., Wang, W.-X. & Wu, L. Scaling behaviours in the growth of networked systems and their geometric origins. *Sci. Rep.* **5**, 9767; DOI:10.1038/srep09767 (2015).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder in order to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>