New Journal of Physics
The open-access journal for physics

Structure and growth of weighted networks

Massimo Riccaboni[1,2,5] **and Stefano Schiavo**[3,4,5]
[1] Department of Computer and Management Sciences, University of Trento, Trento, Italy
[2] Department of Physics, Boston University, Boston, MA, USA
[3] Department of Economics, University of Trento, Trento, Italy
[4] OFCE, Paris, France
E-mail: massimo.riccaboni@unitn.it and stefano.schiavo@unitn.it

New Journal of Physics **12** (2010) 023003 (14pp)
Received 15 October 2009
Published 3 February 2010
Online at http://www.njp.org/
doi:10.1088/1367-2630/12/2/023003

Abstract. We develop a simple theoretical framework for the evolution of weighted networks that is consistent with a number of stylized features of real-world data. In our framework, the Barabási–Albert model of network evolution is extended by assuming that link weights evolve according to a geometric Brownian motion. Our model is verified by means of simulations and real-world trade data. We show that the model correctly predicts the intensity and growth distribution of links, the size–variance relationship of the growth of link weights, the relationship between the degree and strength of nodes, and the scale-free structure of the network.

Contents

[5] Authors to whom any correspondence should be addressed.

1367-2630/10/023003+14$30.00

1. Introduction

Graph theory has been used to describe a vast array of real-world phenomena, but only recently has attention shifted from binary to weighted graphs, from both an empirical perspective [1]–[6] and a theoretical perspective [7]–[9]. The empirical literature has a number of robust stylized facts that apply to a wide range of phenomena as different as Internet traffic, airport connections and international trade. In particular, it has been demonstrated that weighted graphs display (i) a power-law connectivity distribution $P(K)$, with finite size truncation [3, 10]; (ii) a skewed distribution of link weights $P(w)$ and node strengths measured as the sum of the weights of the links of a given node $P(W)$ [11, 12]; and (iii) a power-law relation between node strength $W$and node degree K: $W = K^{\theta}$, with θ ranging between 1.3 and 1.5 [10, 13].

In this paper, we present a simple stochastic model of proportionate growth of both the number and the weight of links to describe the structure and evolution of weighted networks and account for the above-mentioned regularities. In our set-up, we extend the Barabási and Albert (BA) model [14] to accommodate *weighted* network dynamics. This is done by exploiting the theoretical framework recently put forward by Stanley and co-authors to explain the scaling distribution of fluctuations in complex systems [15]–[17].

We test our model using data on the network of international trade flows, which is a prototypical example of a real-world network that is inherently weighted. International trade flows have traditionally been analyzed in the context of the so-called gravity model [18] that relates bilateral flows to countries' sizes and the distance between them. However, one of the main limits of this approach is its inability to capture the large fraction of zeros existing in the matrix of bilateral links. Although this has recently been addressed in the context of standard economic theory [19], graph theory has been applied to accommodate this feature of the data naturally.

We selected the international trade network (ITN) as a test bed for our model based on the following considerations. Firstly, the ITN has already been extensively investigated [5, 6], [20]–[24], and previous works on the ITN provide us with a rich set of empirical regularities. Thus, we know that the link weight distribution assumes a log-normal form in the case of the ITN [23, 24], whereas their growth rates display fat tails [24]. Secondly, the relationship between node strength and degree is crucial in the economic literature about the ITN since it is related to the interplay between intensive and extensive margins of trade, which is a key to explaining trade flows [25][6]. Thirdly, despite the structural inertia of the ITN, the huge volatility of trade flows after the 2008 global financial crisis has recently attracted a great deal of attention. Our theoretical framework provides an explanation for the relationship between node centrality and the variance of network flows.

This paper is organized as follows. Section 2 presents the model and its most important predictions. We then test our model using data on the ITN (section 3) and simulations (section 4). Finally, in the last section we lay down some conclusions and outline possible patterns for future research.

[6] The extensive margin consists of the number of trading partners and the number of products exported K, whereas the intensive margin represents the amount shipped per product per country w.

2. The model

Barabási and Albert [14] proposed a simple stochastic model of network growth based on preferential attachment, which accounts for many of the stylized facts observed in real-world networks. Increasing interest in the study of *weighted* versions of networks calls for an extension of the original BA model to account for the large degree of heterogeneity across link weights [7, 9]. The route we take here exploits the theoretical framework recently put forward by Stanley and co-authors [16] to deal with the growth dynamics of complex systems. We prove that our model is capable of accurately matching the structural properties that characterize a number of real-world weighted networks.

We therefore propose a generalized version of the BA model to describe the dynamics and growth of weighted networks, by modeling them as a set of links of different weights occurring among nodes. In particular, we assume that the weight of links grows according to a geometric Brownian motion (also known as Gibrat's *law of proportionate effects* [26]), so that the expected value of the growth rate of link weights is independent of their current level.

The key sets of assumptions in the model are the following [14, 16, 27]:

1. The network begins at time $t = 0$ with N_0 nodes, each with a self-loop. At each time step $t = \{1, \ldots, M\}$, a new link among two nodes arises; thus the number of links (excluding self-loops that are used only for initialization) existing at time t is $m_t = t$. We write $K_i(t)$ for the number of links of node i at time t (node degree). To identify the nodes connected by the newly formed link at time t, we adopt the following procedure: with probability a the new link is assigned to a new source node, whereas with probability $1 - a$ it is allocated to an existing node i. In the latter case, the probability of choosing node i is given by $p_i(t) = K_i(t-1)/2t$. Edge endpoints i and j of the new link are chosen symmetrically with $i \neq j$. Thus with probability a the new link is assigned to a new target node, whereas with probability $1 - a$ it is allocated to an existing node with probability $p_j(t) = K_j(t-1)/(2t - K_i(t-1))$ if $j \neq i$ and $p_j(t) = 0$ otherwise. Hence, at each time t, this rule identifies the pair of (distinct) nodes to be linked.
2. At time t, each existing link between nodes i and j has weight $w_{ij}(t) > 0$, where K_i, K_j and w_{ij} are independent random variables. At time $t + 1$, the weight of each link is increased or decreased by a random factor $x_{ij}(t)$, so that $w_{ij}(t+1) = w_{ij}(t)x_{ij}(t)$. The shocks and initial link weights are taken from a distribution with finite mean and standard deviation.

Thus, we assume that each link weight grows in time according to a random process. Moreover, the two processes governing link formation and weight growth are assumed to be independent. We therefore combine a preferential attachment mechanism (assumption 1), with an independent geometric Brownian motion of link weights (assumption 2). In this way, we obtain a generalization of the BA set-up capable of accounting for the growth of weighted networks.

Based on the first assumption, we derive the degree distribution $P(K)$ [14, 28]. In the absence of the entry of new nodes ($a = 0$), the probability distribution of the number of links at large t, i.e. the distribution $P(K)$, is exponential:

$$P(K) \approx \frac{1}{\bar{K}} \exp(-K/\bar{K}), \tag{1}$$

where $\bar{K} = 2t/N_0$ is the average number of links per node, which linearly grows with time[7].

If $a > 0$, $P(K)$ becomes a Yule distribution that behaves as a power law for small K,

$$P(K) \sim K^{-\varphi}, \tag{2}$$

where $\varphi = 2 + a/(1-a) \geqslant 2$, followed by the exponential decay of equation (1) for large K with $\bar{K} = (1 + 2t/N_0)^{1-a} - 1$ [15].

Hence, in the limit of large t when $a = 0$ (no entry), the distribution of $P(K)$ converges to an exponential; conversely, when $a > 0$ and small, the connectivity distribution at large t converges to a power law with an exponential cut-off [15].

Using the second assumption, we can compute the growth rate of the strength of nodes. The strength of node i is given by $W_i = \sum_{K_i} w_{ij}$. The growth rate is measured as $g = \ln(W(t+1)/W(t))$. Thus, the resulting distribution of the growth rates of node strength $P(g)$ is determined by

$$P(g) \equiv \sum_{K=1}^{\infty} P(K) P(g|K), \tag{3}$$

where $P(K)$ is the connectivity distribution, computed in the previous stage of the model, and $P(g|K)$ is the conditional distribution of the growth rates of nodes with a given number of links determined by the distribution $P(w)$ and $P(x)$.

Fu *et al* [16] found an analytical solution for the distribution of the growth rates of the weights of links $P(g)$ for the case when $a \to 0$ and $t \to \infty$,

$$P(g) \approx \frac{2V_g}{\sqrt{g^2 + 2V_g}\,(|g| + \sqrt{g^2 + 2V_g})^2}. \tag{4}$$

$P(g)$ has similar behavior to the Laplace distribution for small g, i.e. $P(g) \approx \exp(-\sqrt{2}|g|/\sqrt{V_g})/\sqrt{2V_g}$, whereas for large g, $P(g)$ has power-law tails, $P(g) \sim g^{-3}$, which are eventually truncated for $g \to \infty$ by the distribution $P(x)$ of the growth rate of a single link.

A further implication of the model that can be derived from the second assumption concerns the distribution of link weights $P(w)$. The proportional growth process (assumption 2) implies that the distribution of weights $P(w)$ converges to a log-normal. Thus node strength W is given by the sum of K log-normally distributed stochastic values. Since the log-normal distribution is not stable on aggregation, the distribution of node strength $P(W)$ is multiplied by a stretching factor that, depending on the distribution of the number of links $P(K)$, could lead to a Pareto upper tail [29, 30].

Moreover, a negative relationship exists among the weight of links and the variance of their growth rate. Our model implies an approximate power-law behavior for the variance of growth rates of the form $\sigma(g) = W^{-\beta(W)}$, where $\beta(W)$ is an exponent that weakly depends on the strength W. In particular, $\beta = 0$ for small values of W, $\beta = 1/2$ for $W \to \infty$, and it is well approximated by $\beta \approx 0.2$ for a wide range of intermediate values of W [17].

[7] $\bar{K}$ does not include initial self-loops.

Finally, the model also yields a prediction on the relation between degree K and strength W of each node. In section 4, we show that since the weight of each link is sampled from a log-normal distribution (w are log-normally distributed), and given the skewness of such a density function, the law of large numbers does not work effectively. In other words, the probability to draw a large value for a link weight increases with the number of draws, thus generating a positive power-law relationship between W, and K, for small K.

3. Empirical evidence

To test our model, we use the NBER–United Nations Trade Data [31] available through the Center for International Data at UC Davis. This database provides bilateral trade flows between countries over 1962–2000, disaggregated at the level of commodity groups (four-digit level of the Standard International Trade Classification, SITC). Data are in thousands of US dollars and, for product-level flows, there is a lower threshold at 100 000 dollars, below which transactions are not recorded. One point to note is that disaggregated data are not always consistent with country trade flows; in a number of cases we do not observe any four-digit transaction recorded between two countries, but nevertheless find a positive total trade, and vice versa. Since we take the number of products traded among any pair of countries as the empirical counterpart of the number of transactions, to avoid inconsistency we compute the total trade by aggregating commodity-level data.

In this section, we test the predictions of our model, while in the following section we use the data to calibrate the simulations and check for the ability of the model to replicate real-world phenomena by comparing simulated and actual trade flows. We already know from previous work [24] that the main features of the ITN are broadly consistent with our model. Here, we look in more detail at some specific characteristics of the ITN.

Figure 1 shows that the distribution $P(K)$, which is the number of four-digit SITC products traded by countries, is power-law distributed with an exponential cut-off. The main plot displays the probability distribution in log–log scale, where the power law is the straight line body, and the exponential cut-off is represented by the right tail. The inset presents the same distribution in semi-log scale; this time it is the exponential part of the distribution that becomes a straight line, so that we can magnify what happens to the probability distribution as K grows large. As discussed in section 2, the power-law distribution of K hints at the existence of moderate entry of new nodes into the network. Indeed, 17 new countries enter into the ITN during the observed time frame, mostly due to the collapse of the Soviet Union and Yugoslavia.

Moving to the weighted version of the network, one can look at the distribution of positive link weights as measured by bilateral trade flows at the commodity level, $P(w)$, as well as the total value of country trade or node strength $P(W)$. Figure 2 shows the complementary cumulative probability distribution of trade flows in log–log scale, both for product-level transactions and for aggregate flows. Figure 2 refers to 1997 data (other years display the same behavior). We observe that both distributions show the parabolic shape typical of the log-normal distribution, thus conforming to previous findings [23, 24]. As predicted, on aggregation the power-law behavior of the upper tail becomes more pronounced [29]. However, this departure from log-normality concerns a very small number of observations (0.16% in the case of commodities flows, 2.21% for aggregate flows) since only a few new nodes (countries) enter the network over time.

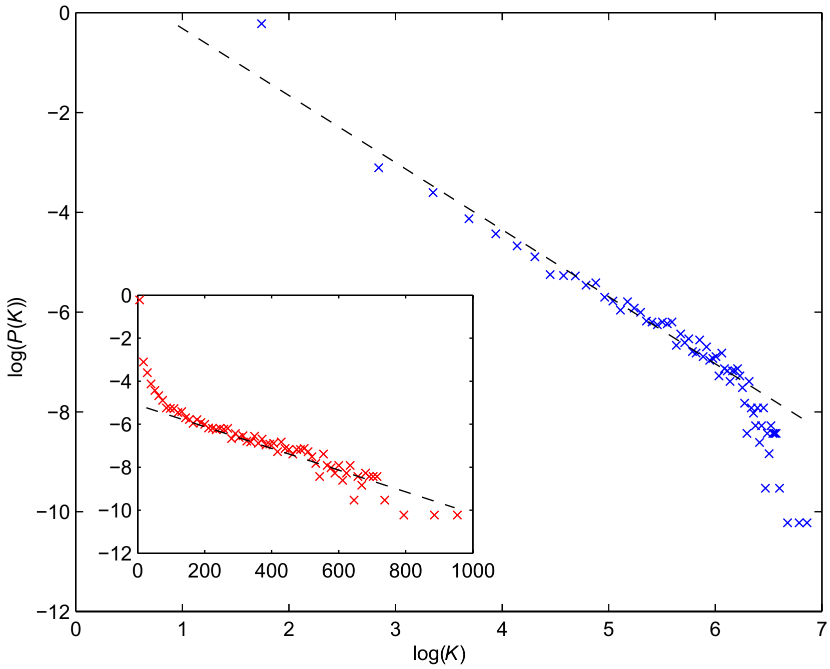

Figure 1. Distribution of the number of products traded, 1997. Double logarithmic scale (main plot) and semi-logarithmic scale (inset).

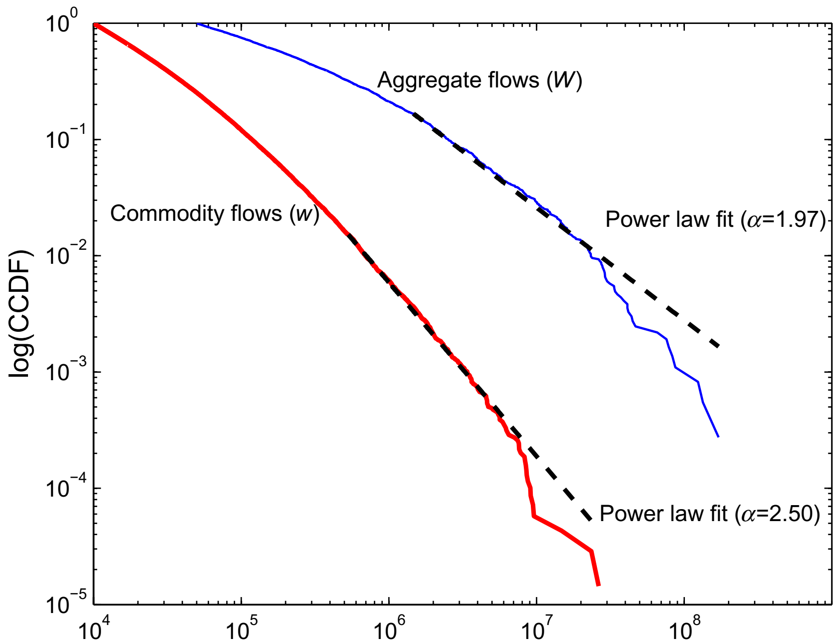

Figure 2. Distribution of the link weights and node strength in the year 1997. Complementary cumulative distribution of the strength distribution $P(W)$ (aggregate flows) and link weights $P(w)$ (commodity flows) and their power-law fits (dashed lines) [34].

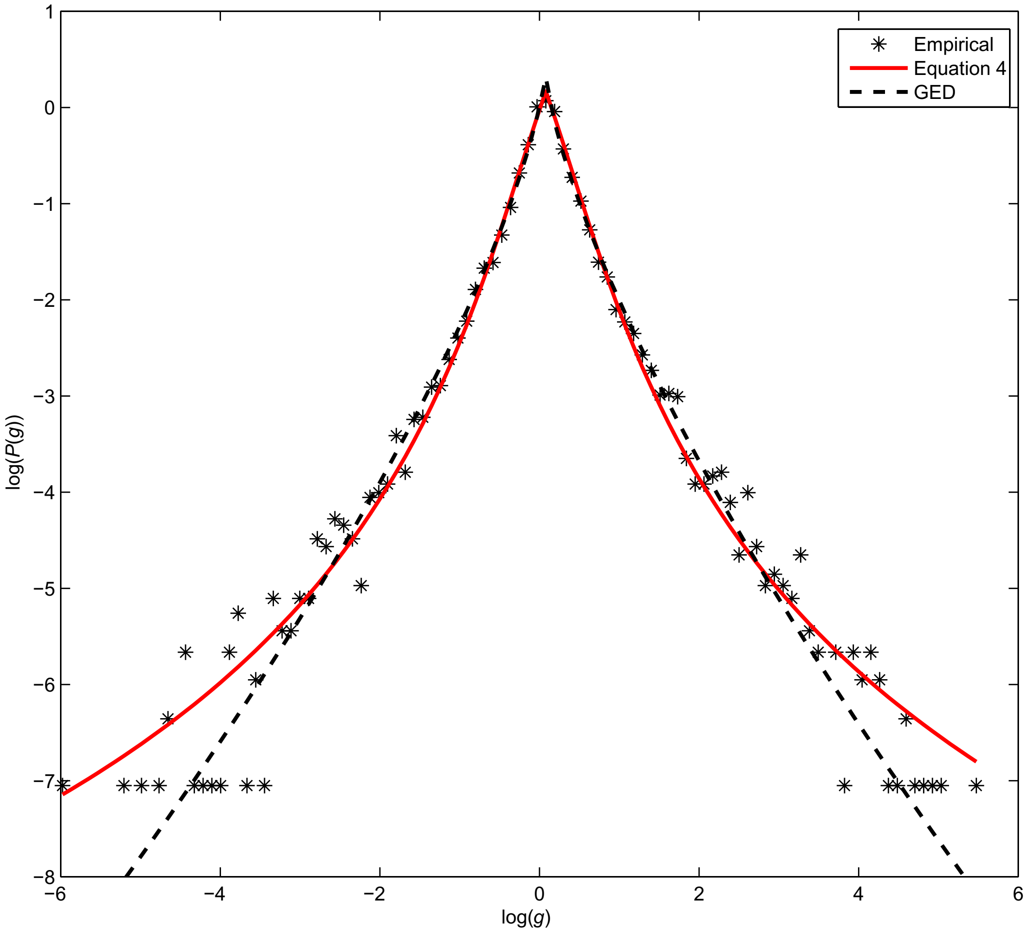

Figure 3. Distribution of the growth rates of aggregate trade flows $P(g)$.

As for the growth of trade flows, figure 3 shows the empirical distribution $P(g)$, together with the maximum likelihood fit of equation (4) and also the generalized exponential distribution (GED, with shape parameter 0.7224).

Goodness-of-fit tests, reported in table 1, show that $P(g)$ is neither Gaussian nor Laplace, whereas the distribution in equation (4) performs much better in terms of KS and AD tests[8]. Hence, the growth of node centrality, as measured by strength W, follows the same law of the fluctuations of the size of complex systems [16, 35]. This is not surprising, since the size of an airport can be measured by the number of passengers who travel through it, and the size of a firm in terms of sales is given by the sum of the value of each product it sells. Thus, the theoretical framework of Stanley and co-workers [16] complements and completes the BA proportional growth model in the case of weighted networks.

As discussed in section 2, our model implies a negative relationship between node strength and the variance of its growth rate. Figure 4 reports the standard deviation of the annual growth

[8] KS and AD are non-parametric tests used to evaluate whether a sample comes from a population with a specific distribution. Both KS and AD tests quantify a distance between the empirical distribution function of the sample and the cumulative distribution function of the reference distribution. The AD test gives more weight to the tails than the KS test. More detailed information is available in [32, 33].

Table 1. Kolmogorov–Smirnov (KS) and Anderson–Darling (AD) goodness-of-fit tests for the distribution of growth rates of trade flows $P(g)$.

Distribution	Mean	Variance	KS	AD
Gauss	0.0333	0.8417	10.8305	11.7329
Laplace	0.0040	0.5338	2.8414	1.4107
GED (shape parameter 0.72)	0.0444	0.2899	1.0915	0.0314
Equation (4)	0.0651	0.3658	0.8214	0.0477

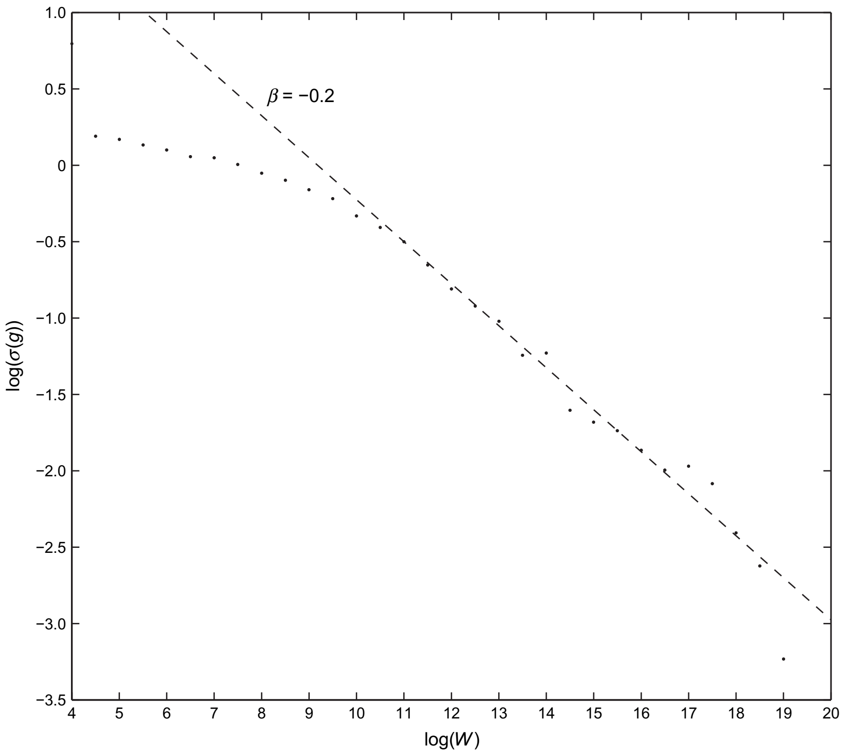

Figure 4. Size–variance relationship between node strength W (trade values) and the standard deviation of its growth rate $\sigma(g)$, double logarithmic scale.

rates of node strength $\sigma(g)$ and their initial magnitude (W). The standard deviation of the growth rate of link weights exhibits a power-law relationship $\sigma(g) = W^{-\beta}$ with $\beta \approx 0.2$, as predicted by the model [17]. This implies that the fluctuations of the most intense trade relationships are more volatile than expected based on the central limit theorem.

All in all, our model accurately predicts the growth and weight distribution of trade flows, the number of commodities traded and the size–variance relationship of trade flows. Thus, we can conclude that a stochastic model that assumes a proportional growth of the number of links combined with an independent proportional growth process of link weights can reproduce most of the observed structural features of the world trade web and should be taken as a valid stochastic benchmark to test the explanatory power of alternative theories of the evolution of international trade and weighted networks in general. In the next section we compare the structure of random networks generated according to our model with the real-world trade network.

4. Simulation results

Based on the assumptions in section 2, we generate a set of random networks and fit them with real-world data in order to test the predictive capability of our theoretical framework. We proceed in two steps. First, we generate the unweighted network according to the first set of assumptions. Next, we assign the value of weights based on a random sampling of K values from a log-normal distribution $P(w)$ whose parameters are obtained through a maximum likelihood fit of the real-world data.

We model a system where at every time t a new link is added, which represents the possibility to exchange one product with a trading partner. We slightly modify the original setting in order to account for the possibility that the new links could be assigned randomly rather than proportionally to node connectivity. Thus, in our simulations, parameter a governs the entry of new nodes according to assumption 1, whereas parameter b is the probability that a new link is assigned randomly. Thus with probability a the new link is assigned to a new source node, whereas with probability $1-a$ it is allocated to an existing node i. In the latter case, the probability of choosing node i is now given by $p_i(t) = (1-b)K_i(t-1)/2t + b/N_{t-1}$, where N_{t-1} is the number of nodes at time $t-1$. The target of the new link is chosen symmetrically with $i \neq j$.

Tuning the two model parameters a and b, we generate different networks in terms of the connectivity distribution of trade links $P(K)$. In particular, without entry $(a=0)$ and completely random allocation of opportunities $(b=1)$, one obtains a random graph characterized by a Poisson connectivity distribution [36], whereas allowing entry $(a>0)$, $P(K)$ is exponentially distributed. Keeping a positive entry rate, but assigning opportunities according to a preferential attachment model $(b=0)$, the model leads to a power-law connectivity distribution with an exponential cut-off, which is more pronounced, the higher the number of initial nodes N_0. In the limit case in which entry of new nodes is ruled out $(a=0)$, the connectivity distribution tends toward a Bose–Einstein geometric distribution.

We compare the structure of random scale-free model networks with the real-world trade network in 1997. Since the structure of the network is highly stable over time, results do not change substantially if we compare simulations with the structure of the real-world network in different years. In the first stage, we generate one million networks, with a and b both ranging from 0 to 1. We simulate random networks of 166 nodes (countries) and 1 079 398 links (number of different commodities traded by two countries). The number of commodities traded is taken as a proxy of the number of transactions. Next, we select the random networks that better fit the real-world pattern in terms of correlation, as measured by the Mantel r test, and connectivity distribution[9].

[9] The Mantel test is a non-parametric statistical test of the correlation between two matrices [37]. The test is based on the distance or dissimilarity matrices that, in the present case, summarize the number of links between two nodes in the simulated and real networks. A typical use of the test entails comparing an observed connectivity matrix with one posed by a model. The significance of a correlation is evaluated via permutations, whereby the rows and columns of the matrices are randomly rearranged.

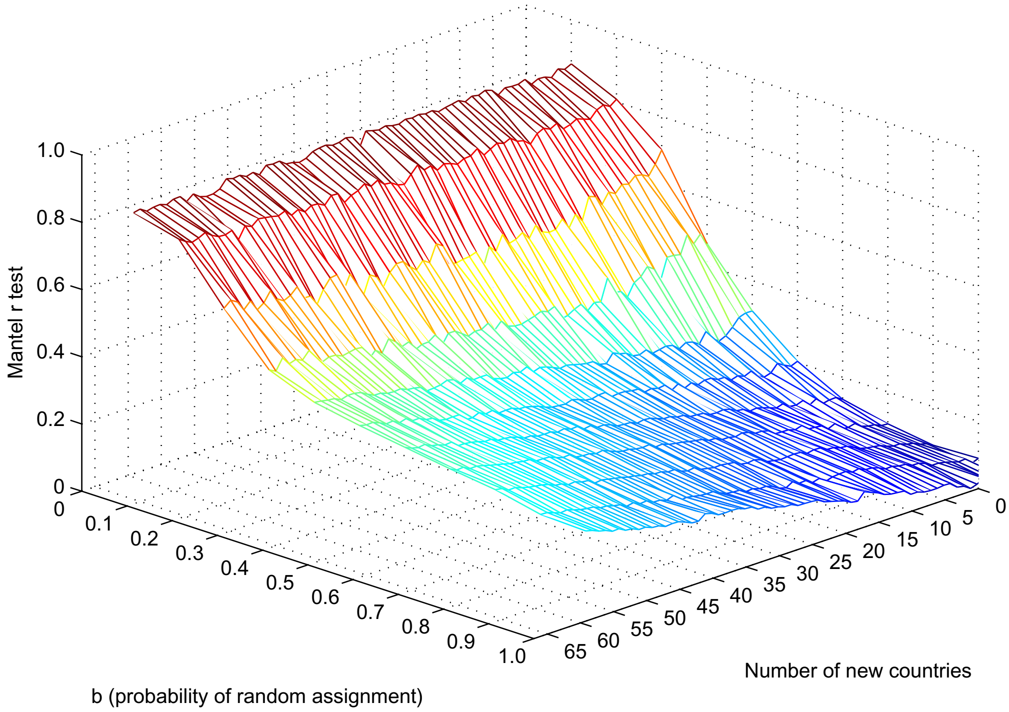

Figure 5. Mantel test comparing simulated and real networks.

Figure 5 reports the value of the Mantel test for networks with $0 \leqslant b \leqslant 1$ and an entry rate a, which implies the entry of 0 to 66 countries. The Mantel correlation statistics reach a peak of 0.88 (p-value < 0.01) in the case of pure preferential attachment regimes ($b = 0$). However, the Mantel test does not discriminate between different entry regimes. We next compare the connectivity distribution of simulated networks with the real-world distribution of the number of traded commodities $P(K)$ by means of the KS goodness-of-fit test. Figure 6 confirms that the best fit is obtained in the case of a purely preferential attachment network ($b = 0$). However, the KS tests provide additional information on the most likely value of a (entry rate of new nodes).

Figure 7 shows that our model can better reproduce the connectivity distribution with an entry rate $a > 0$, which implies the entry of 14–18 countries. This closely corresponds to the empirically observed number of new countries. Thus, we can conclude that a simple proportional growth model with mild entry can account for the distribution of the number of commodities traded by each pair of countries.

Introducing the value of transactions, we can show that the model generates the observed relationship between intensive and extensive margins of trade. Figure 8 depicts the relationship between total trade flows (W) and the number of trade links maintained by each country (K). Empirically, we proxy the number of transactions by means of the number of products traded by each country. Figure 8 displays the relationship that emerges from 1997 trade data, and confirms that there exists a positive correlation between the two variables. The slope of the interpolating line (1.33) in double logarithmic scale reveals a positive relationship between the number of commodities and their average value of the kind $W = K^{\theta}$ with $\theta \approx 1.33$.

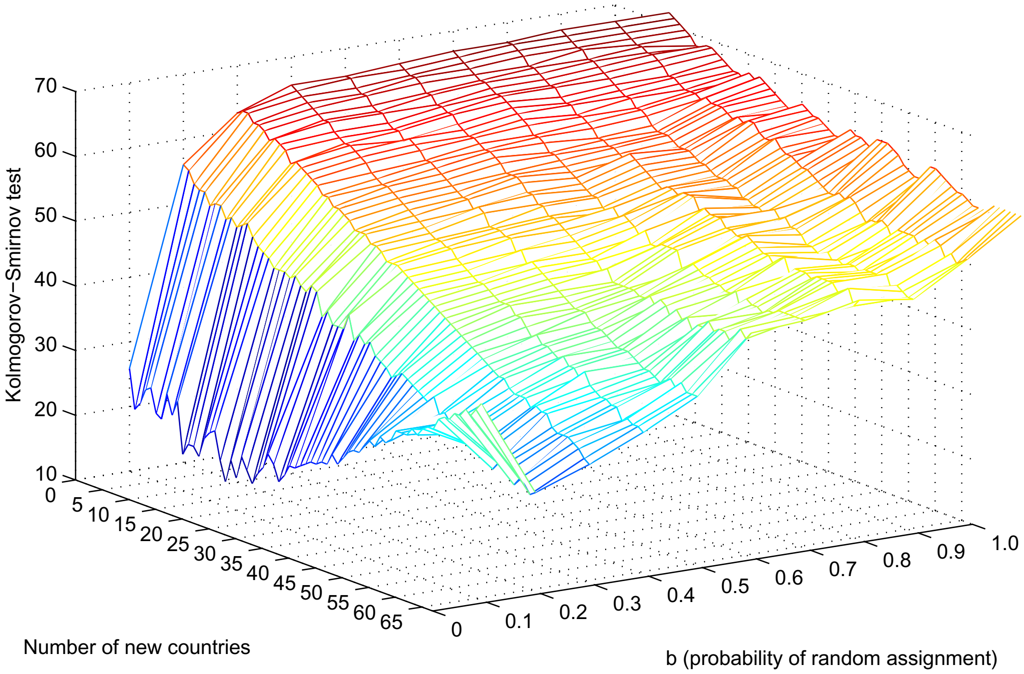

Figure 6. KS goodness-of-fit test for different entry rates and probabilities of random assignment.

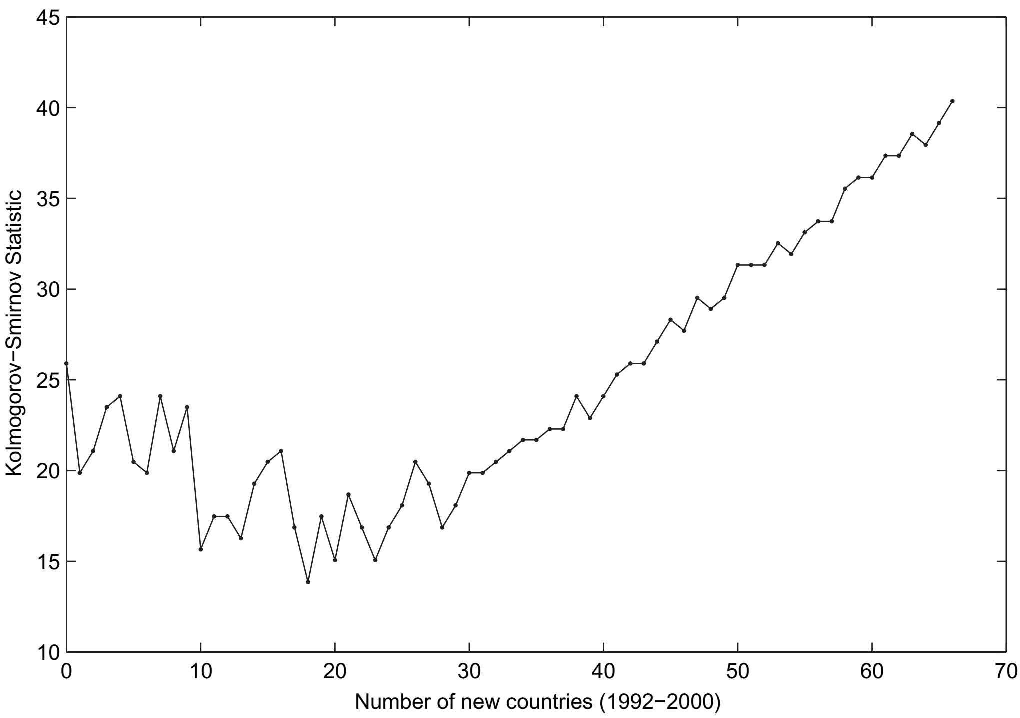

Figure 7. KS goodness-of-fit test for different entry rates in a pure preferential attachment regime ($b = 0$).

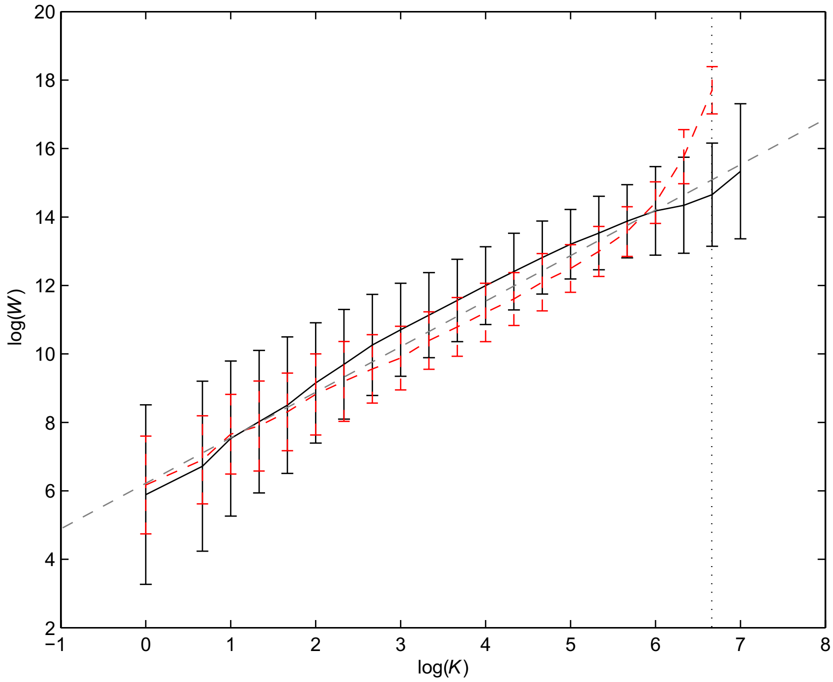

Figure 8. Relationship between the number of products traded and trade value. Double logarithmic scale. Simulated (black) and real-world (red) data, mean and one standard deviation in each direction. The dashed line represents the reference line $W = K^{\theta}$ with $\theta \approx 1.33$.

The curve displays an upward departure in the upper tail. This can be explained by noting that the product classification used imposes a ceiling on the number of products a country can trade as there are only around 1300 four-digit categories (vertical dotted line)[10].

Apart from the upper decile of the distribution, the simulated version of the network shows exactly the same dependence among the magnitude and the number of transactions. This seems surprising, considering that the model assumes two independent growth processes for the number of transactions K and their values w. However, it should be noted that the law of large numbers does not work properly in the case of skew distributions such as the log-normal. Given a random number of transactions with a finite expected value, if its values are repeatedly sampled from a log-normal, as the number of links increases the average link weight will tend to approach and stay close to the expected value (the average for the population). However, this is true only for large K, while according to the distribution $P(K)$, the vast majority of nodes have few links (small K). The higher the variance of the growth process of link weights, the larger K has to be to start observing convergence toward $W = wK^{\theta}$, with $\theta = 1$ predicted by the law of large numbers. Thus, only the largest countries approach the critical threshold. In

[10] Another possible explanation is that for large enough K, some scale effects kick in establishing a correlation between K and W. This could be tested in other real networks with larger K and no cut-off. We are aware that our modeling strategy to assume away any relationship between the mechanisms governing the binary structure of the network and the one assigning link weights is as extreme as other strategies that simply assume a single process governing the two parts. Yet, we consider the positive relationship between K and W as an interesting emerging property of the model.

sum, our simulations demonstrate that our model can account for the relationship between K and W, which has been observed in many real-world weighted networks [10, 13].

5. Discussion and conclusions

Using a simple model of proportionate growth and preferential attachment, we are able to replicate some of the main topological properties of real-world weighted networks. In particular, we provide an explanation for the power-law distribution of connectivity, as well as for the fat tails displayed by the distribution of the growth rates of link weights and node strength. Additionally, the model matches the log-normal distribution of positive link weights (trade flows in the present context) and the negative relationship between node strength and variance of growth fluctuations $\sigma(g) = W^{-\beta}$ with $\beta \approx 0.2$.

The main contribution of the paper is to offer an extension of the BA model for weighted networks. We also provide further evidence that such a unifying stochastic framework is able to capture the dynamics of a vast array of phenomena concerning complex system dynamics [16].

Further refinements of our model entail investigating its ability to match other topological properties of the networks such as assortativity and clustering.

Acknowledgments

We acknowledge Gene Stanley, Sergey Buldyrev, Fabio Pammolli, Jakub Growiec, Dongfeng Fu, Kazuko Yamasaki, Kaushik Matia, Lidia Ponta, Giorgio Fagiolo and Javier Reyes for their previous work on which this contribution builds.

References

[1] Newman M 2001 The structure of scientific collaboration networks *Proc. Natl Acad. Sci. USA* **98** 404–9

[2] Pastor-Satorras R and Vespignani A 2004 *Evolution and Structure of the Internet: a Statistical Physics Approach* (Cambridge: Cambridge University Press)

[3] Guimerà R, Mossa S, Turtschi A and Amaral L 2005 The worldwide air transportation network: anomalous centrality, community structure and cities' global roles *Proc. Natl Acad. Sci. USA* **102** 7794–9

[4] Garlaschelli D, Battiston S, Castri M, Servedio V and Caldarelli G 2005 The scale-free topology of market investments *Physica* A **350** 491–9

[5] Fagiolo G, Reyes J and Schiavo S 2008 On the topological properties of the world trade web: a weighted network analysis *Physica* A **387** 3868–73

[6] Bhattacharya K, Mukherjee G, Saramaki J, Kaski K and Manna S 2008 The International Trade Network: weighted network analysis and modelling *J. Stat. Mech.: Theor. Exp.* P02002

[7] Yook S, Jeong H, Barabási A-L and Tu Y 2001 Weighted evolving networks *Phys. Rev. Lett.* **86** 5835

[8] Zheng D, Trimper S, Zheng B and Hui P 2003 Weighted scale-free networks with stochastic weight assignments *Phys. Rev.* E **67** 040102

[9] Barrat A, Barthélemy M and Vespignani A 2004 Modeling the evolution of weighted networks *Phys. Rev.* E **70** 066149

[10] Barrat A, Barthélemy M, Pastor-Satorras R and Vespignani A 2004 The architecture of complex weighted networks *Proc. Natl Acad. Sci. USA* **101** 3747–52

[11] Ramasco J and Gonçalves B 2007 Transport on weighted networks: when correlations are independent of degree *Phys. Rev.* E **76** 066106

[12] Serrano M, Boguná M and Vespignani A 2009 Extracting the multiscale backbone of complex weighted networks *Proc. Natl Acad. Sci. USA* **106** 6483–8

[13] Eom Y-H, Jeon C, Jeong H and Kahng B 2008 Evolution of weighted scale-free networks in empirical data *Phys. Rev.* E **77** 056105

[14] Barabási A and Albert R 1999 Emergence of scaling in random networks *Science* **286** 509

[15] Yamasaki K, Matia K, Buldyrev S, Fu D, Pammolli F, Riccaboni M and Stanley H 2006 Preferential attachment and growth dynamics in complex systems *Phys. Rev.* E **74** 035103

[16] Fu D, Pammolli F, Buldyrev S, Riccaboni M, Matia K, Yamasaki K and Stanley H 2005 The growth of business firms: theoretical framework and empirical evidence *Proc. Natl Acad. Sci. USA* **102** 18801–6

[17] Riccaboni M, Pammolli F, Buldyrev S, Ponta L and Stanley H 2008 The size variance relationship of business firm growth rates *Proc. Natl Acad. Sci. USA* **105** 19595

[18] Tinbergen J 1962 *Shaping the World Economy: Suggestions for an International Economic Policy* (New York: The Twentieth Century Fund)

[19] Helpman E, Melitz M and Rubinstein Y 2008 Estimating trade flows: trading partners and trading volumes *Q. J. Econ.* **123** 441–87

[20] Serrano M and Boguná M 2003 Topology of the world trade web *Phys. Rev.* E **68** 15101

[21] Garlaschelli D and Loffredo M 2004 Fitness-dependent topological properties of the world trade web *Phys. Rev. Lett.* **93** 188701

[22] Garlaschelli D and Loffredo M 2005 Structure and evolution of the world trade network *Physica* A **355** 138–44

[23] Bhattacharya K, Mukherjee G and Manna S 2007 The International Trade Network. arXiv:0707.4347

[24] Fagiolo G, Reyes J and Schiavo S 2009 World-trade web: topological properties, dynamics and evolution *Phys. Rev.* E **79** 36115

[25] Chaney T 2008 Distorted gravity: the intensive and extensive margins of international trade *Am. Econ. Rev.* **98** 1707–21

[26] Gibrat R 1931 *Les Inegalites Economiques* (Paris: Sirey)

[27] Bollobás B and Riordan O 2003 Mathematical results on scale-free random graphs *Handbook of Graphs and Networks* ed S Bornholdt and H G Schuster (New York: Wiley–VCH), pp. 1–34

[28] Buldyrev S, Growiec J, Pammolli F, Riccaboni M and Stanley H 2007 The growth of business firms: facts and theory *J. Eur. Econ. Assoc.* **5** 574–84

[29] Growiec J, Pammolli F, Riccaboni M and Stanley H 2008 On the size distribution of business firms *Econ. Lett.* **98** 207–12

[30] De Fabritiis G, Pammolli F and Riccaboni M 2003 On size and growth of business firms *Physica* A **324** 38–4

[31] Feenstra R, Lipsey R, Deng H, Ma A, Mo H and Drive O 2005 World trade flows: 1962–2000 NBER working paper

[32] Chakravarti I, Laha R and Roy J 1967 *Handbook of Methods of Applied Statistics* Vol I (New York: Wiley), pp. 392–4

[33] Stephens M 1974 EDF statistics for goodness of fit and some comparisons *J. Am. Stat. Assoc.* **69** 730–7

[34] Clauset A, Shalizi C and Newman M 2009 Power-law distributions in empirical data *SIAM Rev.* **51** 661–703

[35] Fagiolo G, Napoletano M and Roventini A 2008 Are output growth-rate distributions fat-tailed? Some evidence from OECD countries *J. Appl. Econ.* **23** 639–9

[36] Erdos P and Renyi A 1959 On random graphs *Publ. Math. Debrecen* **6** 156

[37] Mantel N 1967 The detection of disease clustering and a generalized regression approach *Cancer Res.* **27** 209–20