

## Scale-free models for the structure of business firm networks

Maksim Kitsak,<sup>1,2</sup> Massimo Riccaboni,<sup>1,3</sup> Shlomo Havlin,<sup>1,4</sup> Fabio Pammolli,<sup>1,5</sup> and H. Eugene Stanley<sup>1</sup>

<sup>1</sup>Center for Polymer Studies, Boston University, Boston, Massachusetts 02215, USA

<sup>2</sup>Cooperative Association for Internet Data Analysis (CAIDA), University of California–San Diego (UCSD), La Jolla, California 92093, USA

<sup>3</sup>DISA, University of Trento, Trento, Italy

<sup>4</sup>Minerva Center and Department of Physics, Bar-Ilan University, Ramat Gan, Israel

<sup>5</sup>IMT Institute for Advanced Studies, Lucca, Italy

(Received 6 August 2008; revised manuscript received 26 January 2010; published 29 March 2010)

We study firm collaborations in the life sciences and the information and communication technology sectors. We propose an approach to characterize industrial leadership using  $k$ -shell decomposition, with top-ranking firms in terms of market value in higher  $k$ -shell layers. We find that the life sciences industry network consists of three distinct components: a “nucleus,” which is a small well-connected subgraph, “tendrils,” which are small subgraphs consisting of small degree nodes connected exclusively to the nucleus, and a “bulk body,” which consists of the majority of nodes. Industrial leaders, i.e., the largest companies in terms of market value, are in the highest  $k$ -shells of both networks. The nucleus of the life sciences sector is very stable: once a firm enters the nucleus, it is likely to stay there for a long time. At the same time we do not observe the above three components in the information and communication technology sector. We also conduct a systematic study of these three components in random scale-free networks. Our results suggest that the sizes of the nucleus and the tendrils in scale-free networks decrease as the exponent of the power-law degree distribution  $\lambda$  increases, and disappear for  $\lambda \geq 3$ . We compare the  $k$ -shell structure of random scale-free model networks with two real-world business firm networks in the life sciences and in the information and communication technology sectors. We argue that the observed behavior of the  $k$ -shell structure in the two industries is consistent with the coexistence of both preferential and random agreements in the evolution of industrial networks.

DOI: [10.1103/PhysRevE.81.036117](https://doi.org/10.1103/PhysRevE.81.036117)

PACS number(s): 89.75.Hc

### I. INTRODUCTION

Many real-world complex systems are often described using the representation of graphs or networks, as sets of nodes connected by links [1–7]. Networks appear in various areas of science such as physics, biology, computer sciences, economics, and sociology [3–7]. Real networks, despite their diversity, appear to have many common properties. Some real networks have been found to be “small-world” [3,8–10]: the typical distance between nodes in a network is very small, of order of  $\log(N)$  or less [3,11,12], where  $N$  is the total number of nodes. Also, some real networks are scale-free (SF) with a power-law tail in their degree distribution

$$P(q) \propto (q + c)^{-\lambda}, \quad (1)$$

where  $q$  is the number of links per node,  $\lambda$  and  $c$  are distribution parameters [13–16].

Many statistical physics and mathematical techniques have been successfully applied to study network structure including percolation [12,17–21], scaling [6,7,11], partitioning [22], box covering [23–26],  $k$ -core percolation, and  $k$ -shell decomposition [27–29].

$k$ -cores and  $k$ -shells of a network can be conveniently defined with the  $k$ -shell decomposition algorithm (see, for instance, [29]). One starts the process of the  $k$ -shell decomposition on a network by removing all nodes with degree  $q = 1$ . After the first iteration of pruning, there may appear new nodes with degrees  $q = 1$ . One keeps on pruning these nodes until only nodes with degree  $q \geq 2$  are left. The removed nodes along with the links connecting them form the  $k = 1$

shell. Next, one iterates the pruning process for nodes of degree  $q = 2$ , thereby creating the  $k = 2$  shell. The  $k$ -shell decomposition is repeated for higher values of  $q$  until all nodes of the network are removed. As a result, each node in the network is assigned a  $k$ -shell index  $k$ . The largest  $k$ -shell index is called  $k_{\max}$ , which is also the total number of  $k$ -shells in the network, provided all  $k$ -shells with indices below  $k_{\max}$  exist. The  $k$ -crust is defined as the union of all  $k$ -shells with indices smaller or equal  $k$ . Similarly, the  $k$  core is defined as the union of all nodes with indices greater or equal  $k$  [see Fig. 1 for demonstration]. The  $k$ -core analysis has been successfully employed for visualization of large-scale networks [29] and uncovering the structure of the Internet at the autonomous system level [30]. The  $k$ -core (bootstrap) percolation also has many applications in physics of magnetic materials [31] and glass-jamming transitions [32]. An analytical approach for describing  $k$ -core structure of uncorrelated treelike graphs was proposed by mathematicians [33,34]. The  $k$ -core structure and the  $k$ -core (bootstrap) per-

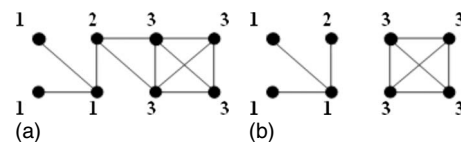


FIG. 1. The illustration of the  $k$ -shell decomposition method. (a) Original network. Nodes are marked by corresponding  $k$ -shell indices. Note that the  $k$ -shell index does not coincide with the node degree. (b) The two-crust (left) and the 3-core (right) of the original network.

colation in equilibrium of random uncorrelated complex networks was studied analytically in [35–38].

The structure and growth of business firms is of considerable interest. Recently, a proportional growth model has been introduced to explain the distribution of business-firm growth rates [39,40]. In the present work, we study dynamics and leadership stability in business firm networks by applying the  $k$ -shell decomposition technique to networks of collaborative agreements in the life sciences (LS) and the information and communication technology (ICT) sectors [26,41–43]. Industrial leaders are top-ranking companies in terms of size as measured by market value. The persistency of leadership debate has attracted a great deal of attention in the Industrial Organization literature over the past generation [44,45]. Two rival views exist. The first asserts that leadership tends to persist for a “long” time [46], while the second one emphasizes the transience of leadership positions due to radical innovation and “leapfrogging competition” [47]. The LS sector is generally cited as an archetypal example of stability of industry leaders [41], while, on the contrary, the ICT sector is widely considered as a good example of high market turnover and instability [43].

We propose an approach to characterize leadership positions of industry firms by means of the  $k$ -shell index, with firms in higher  $k$ -shell layers having higher positions in the leadership hierarchy. Indeed, industrial leaders are typically cross linked in a dense network of collaborative agreements. As a result, industrial leaders are connected to a large number of other industrial leaders and, therefore, belong to high  $k$ -shell layers. The  $k$ -shell decomposition provides an easy parameter-free way of identification of industry leaders, which is especially useful to control for the existence of potential collusive practices [48]. We verify the validity of our approach by comparing firms occupying the innermost  $k$ -shells in the LS and the ICT industries with the list of industry leaders based on business week reviews of global leading firms [49–51] [see Table I].

We also conduct a numerical analysis of model SF networks to better understand their  $k$ -shell structure. We then compare the  $k$ -shell structure of SF models with that of the LS and the ICT industry networks in order to get a better insight into the growth principles of the industry networks.

The rest of the manuscript is organized as follows: in Sec. II, we apply the  $k$ -shell decomposition to the LS and the ICT sectors and analyze the leadership structure of these sectors. In Sec. III, we conduct a systematic analysis of the  $k$ -shell properties of model SF networks. In Sec. IV, we compare the results for SF models with those observed in the two industrial networks. We conclude our manuscript in Sec. V with a discussion and summary. In the Appendix, we discuss the validity of our approach by comparing firms in the innermost  $k$ -shell layers with global firm rankings based on market value.

## II. $k$ -SHELL STRUCTURE OF THE LS AND THE ICT SECTORS

We analyze the  $k$ -shell structure of the LS and the ICT sectors in time periods between 1990 and 2002 and between

TABLE I. Industrial leaders based on Business Week Global 1000, years 2001–2. Companies who are in the highest  $k$ -shell layers of the LS ( $k_{\max}=19$ ) and ICT ( $k_{\max}=6$ ) networks, years 2002 (LS) and 2001 (ICT). For each company business week overall rank and industry specific rank for IT and Health companies (in parenthesis, \* for companies in other industries) are reported as well as market value (in millions of U.S. dollars).

ICT Leaders (2001)		
Company	Rank	Market value
Microsoft	2(1)	369102
AOL	7(2)	230475
IBM	10(3)	196858
NTT	11(*)	192578
Intel	13(4)	181453
Verizon	17(*)	148290
Cisco	19(5)	140143
AT&T	43(*)	80647
Sony	49(*)	70745
Siemens	61(*)	63731
Dell	62(8)	63611
Texas Instrument	70(9)	59086
HP	72(10)	56662
Sun Microsystems	80(11)	53033
Ericsson	89(12)	49595
QualComm	95(13)	45997
Nortel	111(15)	42249
Applied Materials	116(16)	40500
Hitachi	143(18)	34385
Motorola	151(19)	32110
Alcatel	161(21)	30773
Compaq	177(23)	27183
NEC	178(24)	27004
Lucent	182(25)	26838
Fujitsu	190(26)	25676
Infineon	221(30)	21421
Toshiba	254(34)	18287
Ebay	288(*)	16350
Analog Development	297(41)	16031
Agilent Technology	306(42)	15307
Juniper Network	350(52)	13641
Xilinx	351(53)	13612
Verisign	421(58)	11349
Matsushita	441(60)	10736
Yahoo!	466(63)	10267
AMD	536(70)	8927
Apple	681(84)	6913
LSI Logic	772(93)	5909
Symantec	840(101)	5367
Terra Network	911(114)	4818
ARM	923(115)	4749

TABLE I. (Continued.)

LS Leaders (2002)		
Company	Rank	Market value
General Electric	1( *)	309462
Pfizer+Pharmacia	6(1)	216777
Johson&Johnson	9(2)	186942
Merck	15(3)	129679
GlaxoSmithKline	16(4)	126273
Novartis	17(5)	123930
P&G	21(6)	116380
AstraZeneca	38(7)	76648
Abbott	40(8)	74129
Wyeth	42(9)	73640
Roche+Genentech	43(10)	73339
Eli Lilly	44(11)	72724
Bristol-Myers Squibb	53(12)	60330
Sanofi+Aventis	60(14)	55377
Amgen	70(15)	49907
Schering Plough	97(17)	38776
Bayer	179(23)	23745
Schering AG	373(29)	12051
Novo Nordisk	397(30)	11326
Biogen	603(47)	7408
Monsanto	638( *)	6916
Chiron	641(55)	6874
Genzyme	651(56)	6833
UCB	762(66)	5590
Biovail	840(71)	4936
Shire	910(74)	4392
Millennium	952(77)	4169

1990 and 2000, respectively. The LS network expanded linearly since the mid-1970s while the ICT network took off in the 1990s and grew exponentially for a decade [see Fig. 2(a) and the inset]. The total number of firms in the LS is  $N=6,776$  and in the ICT is  $N=7,759$ . These sizes refer to the

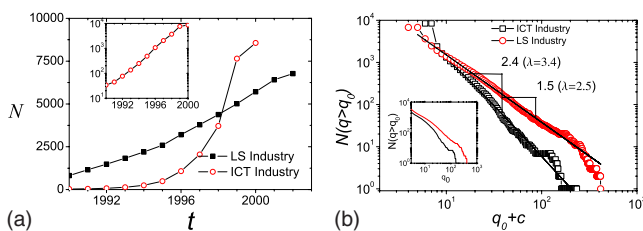


FIG. 2. (Color online) Properties of the LS and the ICT sectors. (a) The growth of the largest connected component of the LS and the ICT industries. The LS industry expands almost linearly while the ICT industry exhibits exponential growth (see inset, a semilog plot of the ICT network size as a function of time). (b) Cumulative degree distribution,  $N(q > q_0)$ , of the LS and the ICT networks. The inset displays cumulative degree distribution  $N(q > q_0)$  of the LS and the ICT network without offset  $c$ .

largest connected component of each of the networks at the last year of observation. Both industrial networks feature SF degree distributions with  $\lambda \approx 2.5$ ,  $c \approx 4$  (LS) and  $\lambda \approx 3.4$ ,  $c \approx 6$  (ICT), which we estimate with the maximum likelihood method [52,53] [see Fig. 2(b)]. We use the Kolmogorov-Smirnov test in order to examine the goodness of fit of the degree distributions. The obtained  $p$ -values for 1000 trials are 0.24 and 0.33 for the LS and the ICT industry networks, respectively.

We next apply the  $k$ -shell decomposition procedure to the LS and the ICT networks. For each  $k$ -crust we calculate  $N_0$ , the total number of nodes in the  $k$ -crust,  $N_1$ , the size of the largest connected component in the  $k$ -crust and  $N_2$ , the size of the second largest component of the  $k$ -crust. As seen in Fig. 3, the LS network consists of  $k_{\max}=19$  shells while the ICT network (which consists of comparable number of firms) has only  $k_{\max}=6$  shells. The size of the largest cluster in the  $k$ -crust starts growing rapidly after  $k=4$  and  $k=2$  in the LS and the ICT networks respectively. At these values of  $k$ , the size of the second largest cluster in the  $k$ -crust reaches a maximum. The above behavior of the  $k$ -crust components is consistent with the existence of a second order phase transition in the  $k$ -crust structure [30]. The type of the phase transition taking place at the  $k$ -shell decomposition is similar to that of targeted percolation in SF networks [21,54].

Unlike in the ICT network, the size of the largest connected component  $N_1$  in the LS network  $k$ -crust undergoes a large jump  $R(k=19) \equiv N_1(k=19) - N_1(k=18) = 745$ , while the total size  $N_0$  of the  $k$ -crust experiences only a small change  $N_0(k=19) - N_0(k=18) = 43$  [see Fig. 3(a)]. This can be explained as follows. 745 firms of the LS sector (approximately 10% of the LS network) sign collaborative agreements exclusively with the 43 firms that form the  $k$ -shell  $k=k_{\max}=19$ . Typically, each of these 745 firms sign a small number of agreements with other firms and, therefore, has small degree. Thus, the 745 firms are removed in the decomposition of the first few  $k$ -shells. However, being connected exclusively to the  $k_{\max}=19$  shell, the 745 firms do not contribute to the largest connected component of the  $k$ -crust until the last  $k_{\max}$ -shell is decomposed from the network and added to the  $k$ -crust. It is the inclusion of the 745 firms into  $N_1(k)$  at the decomposition of the  $k_{\max}$ -shell that results in the observed jump  $R(k)$ . Thus, the 43 firms in the  $k_{\max}$ -shell of the LS network connect the 745 firms with the remaining 90% of the network. This allows one to associate the structure of the LS sector with that of a jellyfish. Similar structure was observed in the internet at the autonomous system level [30] and the contact network of inpatients in Sweden [55]. For a different jellyfish model see [56]. In our paper, we follow the terminology of [30] and call the 43 firms in the  $k_{\max}$ -shell of the LS sector the *nucleus*. We refer to the 745 firms which are connected to other firms in the LS sector exclusively via nucleus as *tendrils*. We call the rest of the LS industry the *bulk body*. Interestingly, we do not observe a pronounced jump in  $N_1(k)$  in the ICT network [see Fig. 3(b)].

In both the LS and the ICT sectors  $k_{\max}$ -shells include leading companies in pharmaceuticals (Pfizer), computers (IBM), software (Microsoft), media (Time Warner), telecommunication (Verizon), household and personal equipments (Procter & Gamble), semiconductors (Intel) as well as top

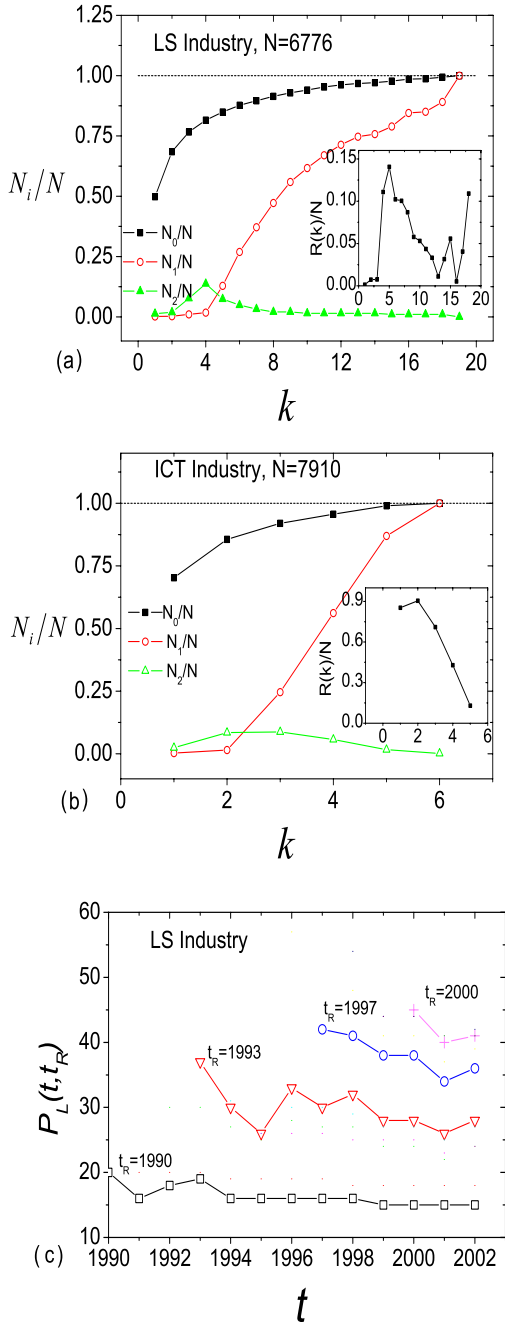


FIG. 3. (Color online) (a) The  $k$ -shell structure of the LS network and (b) of the ICT network. Plots of the size of the  $k$ -crust,  $N_0$ ; the size of the largest connected component of the  $k$ -crust,  $N_1$ ; and the size of the second largest connected component of the  $k$ -crust,  $N_2$ , as a function of the  $k$ -crust index  $k$ . The sizes are normalized with the total number of nodes in the network,  $N$ .  $N_2$  is multiplied by 100. The transition of the  $k$ -crust at  $k$ -shell  $k=18$  of the LS network reveals a jellyfish topology [30]. Since the change in  $F_1 \equiv N_1/N$  at the  $k=k_{\max}$  is smaller in the ICT network than in the previous  $k$ -shell  $k=k_{\max}-1$ , the ICT does not have a jellyfish structure. The insets of (a) and (b) show the rate  $R$  of the largest connected component change as a function of  $k$ -shell index  $k$ . (c) The leadership persistence  $P_L(t, t_R)$  as a function of years in the LS industry. We define  $P_L(t, t_R)$  as a number of firms that were in the nucleus both at time  $t_R$  and  $t$ . Note that most of the LS firms remain in the nucleus after they first enter it.

conglomerates such as General Electric, Siemens, and Samsung [see Table I for the full listing of firms constituting  $k_{\max}$ -shells of the LS and the ICT industries].

We find that the LS network tendrils are mostly composed of the new start-up firms and their university partners. On the one hand, start-up these firms preferentially attach to market leaders. On the other hand, market leaders of the LS sector compete to sign exclusive deals with new and promising start-up firms. In order to assess stability of the LS nucleus, we introduce the leadership persistence function,  $P_L(t, t_R)$ , which we define as a number of firms that were present in the nucleus of the LS industry both at time  $t$  and reference time  $t_R$ . In Fig. 3(c) we plot  $P_L(t, t_R)$  as a function of time  $t$  for different reference times  $t_R$ . We notice the remarkable stability of the LS nucleus: once a particular firm enters the nucleus it is very likely to remain there for many years. As seen in Fig. 3(c), out of 20 firms which were in the LS nucleus in 1990, 15 firms remained in the nucleus in 2002.

On the contrary, in the ICT sector there is more emphasis on how to integrate different technologies and markets [43]. Thus, there are more deals between firms of the same size and age but in different technological and market areas. As a consequence, the  $k$ -shell of the ICT network is more unstable and heterogeneous which may explain why we do not detect the emergence of a nucleus such as in the LS sector.

### III. $k$ -SHELL PROPERTIES OF SCALE-FREE NETWORKS

As seen in Fig. 2(b), both the LS and the ICT industrial networks exhibit a SF degree distribution. Hence, in order to better understand the substructures of real networks—the nucleus, the tendrils, and the bulk components—we analyze the  $k$ -shell structure of the random SF models which were generated using the configurational approach [57]. We calculate the  $k$ -shell structure of random SF networks with  $c \geq 0$ , and degree distribution exponent  $\lambda \in [2, 3]$ . For our simulations [Fig. 4] we choose networks sizes of  $N=8,000$  (which is comparable to the size of the LS and ICT industry networks) and  $N=10^6$  in order to test the influence of finite size effects.

As discussed above, one can in general calculate the size of each  $k$ -shell,  $S_k$ , from the  $k$ -shell decomposition as

$$S_k = N_0(k) - N_0(k-1). \quad (2)$$

The jump in  $N_1(k)$  at  $k=k_{\max}$  is comprised by the inclusion of tendrils  $S_t$  and the nucleus  $S_{k=k_{\max}}$ . Thus, the size of tendrils can be calculated as

$$S_t = R(k_{\max}) - S_{\max}, \quad (3)$$

where  $R(k)$  is the rate of change of the  $N_1(k)$

$$R(k) \equiv N_1(k) - N_1(k-1). \quad (4)$$

It is seen in Fig. 4 that both  $k_{\max}$  and  $R(k_{\max})$  decrease as the exponent  $\lambda$  increases which lead to the decrease of relative size of tendrils,  $S_t$ , determined by Eq. (3). In the case of SF models with  $N=8000$  and  $\lambda > 2.5$   $R(k_{\max}) < R(k_{\max}-1)$  and the relative size of the tendrils,  $S_t$ , becomes unresolved. This motivates us to introduce a quantitative criterion for the

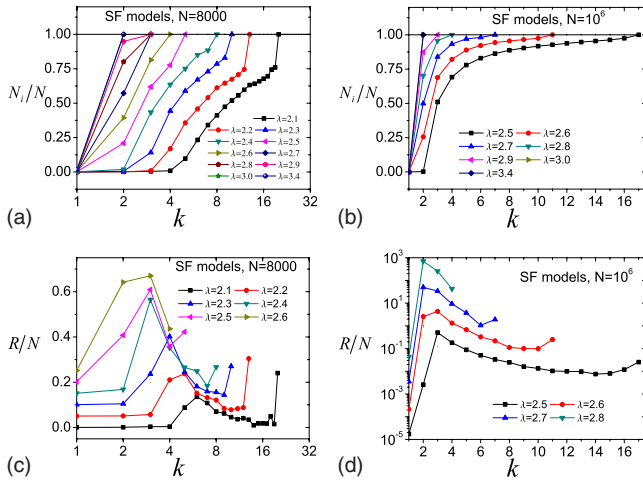


FIG. 4. (Color online) The  $k$ -shell structure of random SF models with (a)  $N=8000$  and (b)  $N=10^6$  nodes. Note that  $\lambda=3.4$  curve overlaps with that of  $\lambda=3.0$ . The rate of the largest connected component change,  $R$ , as a function of  $k$ -shell index  $k$  for (c)  $N=8000$  and (d)  $N=10^6$ . Note that in order to avoid the overlap of curves in (c) and (d) we subsequently shift the plots with respect to each other by the additive factor 0.05 in (c) and the multiplicative factor of 10 in (d). The sizes of the nucleus and the tendrils decrease as  $k$  increases. The nucleus and the tendrils disappear in (a) and (c) for  $N=8000$  for  $\lambda > 2.5$  and in (b) and (d) for  $N=10^6$  for  $\lambda > 2.7$ .

emergence of the three distinct components—nucleus, tendrils and bulk body,

$$R(k_{\max}) > R(k_{\max} - 1). \quad (5)$$

By examining  $R(k)$  plots, we observe in the case of  $N=8000$  [Fig. 4(c)], that SF models with  $\lambda > 2.5$  already do not have a nucleus and tendrils. However, for  $N=10^6$ , we do observe nucleus and tendrils in SF models for  $\lambda \leq 2.7$  [Fig. 4(d)]. These observations suggest that all SF networks with  $2 < \lambda < 3$  have a nucleus and tendrils provided  $N$  is sufficiently large. The above observations for SF models agree with the fact that we observe jellyfish topology for the LS network ( $\lambda=2.5$ ) and do not observe it for the ICT network  $\lambda=3.4$ . However, according to our simulations, SF model networks with  $N=8000$  with  $\lambda=2.5$  and  $\lambda=3.4$  have on the average  $k_{\max}=5$  and  $k_{\max}=2$ , respectively, while the measured  $k_{\max}$  in the LS and the ICT networks (which have the same  $\lambda$  values) are  $k_{\max}=19$  and  $k_{\max}=6$ . The observed difference in  $k_{\max}$  between industry networks and SF models with similar parameters suggests to further explore the  $k$ -shell structure of SF networks and consider SF model networks with  $c > 0$  values in  $P(q)$  (as found in the LS and the ICT networks).

In order to better understand how  $k$ -shell structure changes as networks grow, we calculate the sizes of the  $k$ -shells as a function of  $N$  and  $\lambda$  for model SF networks. For each pair of values  $N$  and  $\lambda$  we generate  $10^3$ – $10^4$  realizations of SF models and calculate the average number of  $k$ -shells  $k_{\max}$  constituting the network as well as their average size  $S_k$  [see Fig. 5]. For small  $k$  values,  $S_k$  is proportional to  $N$ . As the size of the network increases new  $k$ -shells start to appear. When the size increases further, the growth of new  $k$ -shells

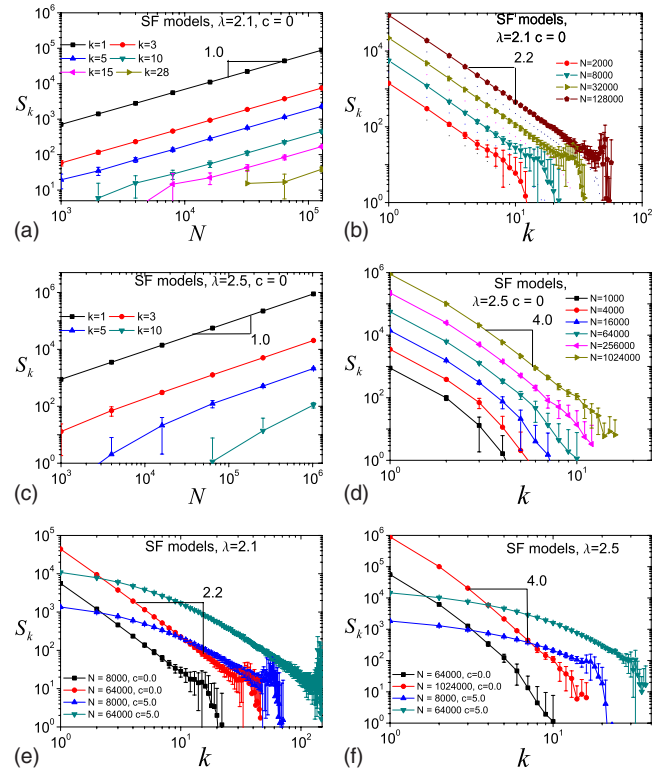


FIG. 5. (Color online) Sizes of  $k$ -shells,  $S_k$ , for SF random models with  $c=0$  (a,b)  $\lambda=2.1$  and (c,d)  $\lambda=2.5$  as a function of (a,c)  $N$  and the  $k$ -shell index (b,d)  $k$ . Note that sizes of  $k$ -shells increases proportionally to  $N$ . (e,f) Sizes of  $k$ -shells,  $S_k$ , for SF networks with  $c=5.0$  and (e)  $\lambda=2.1$ , (f)  $\lambda=2.5$ . SF models with  $c=5.0$  have significantly larger number of  $k$ -shells,  $k_{\max}$ , compared to SF models with the same  $\lambda$  and  $c=0$ .  $S_k$  in SF models with  $c=5.0$  decreases significantly slower as a function of  $k$  compared to SF models with the same  $\lambda$  and  $c=0$  in the small  $k$  region. In the large  $k$  region both types of SF models with  $c=0.0$  and  $c=5.0$  seem to have similar behavior  $S_k \sim k^{-\delta}$ , where  $\delta=2/(3-\lambda)$ .

stabilizes and becomes also proportional to  $N$  [see Figs. 5(a) and 5(c)]. This result indicates that the size of each  $k$ -shell constitutes a certain finite fraction of the network and this fraction decreases with increasing  $k$ . The analytical analysis of  $k$ -core structure [35] leads to

$$S_k \propto k^{-\delta}, \quad (6)$$

where

$$\delta = \frac{2}{3-\lambda}. \quad (7)$$

Hence  $\delta \approx 2.2$  for  $\lambda=2.1$  and  $\delta \approx 4.0$  for  $\lambda=2.5$ , which agrees with our simulations. [Figs. 5(b) and 5(d)]. The appearance of  $c > 0$  in the SF degree distribution significantly increases the average degree in the network and also  $k_{\max}$  [see Figs. 5(e) and 5(f)]. However, the dependence of  $k$ -shell sizes  $S_k$  seems to remain the same in the large  $k$  limit:  $S_k \sim k^{-\delta}$  [Figs. 5(e) and 5(f)].

One can estimate the total number of  $k$ -shells in a random SF network of size  $N$  as follows. Since every  $k$ -shell constitutes a fixed fraction of  $N$  it follows that  $S_k \propto Nk^{-\delta}$ . The last

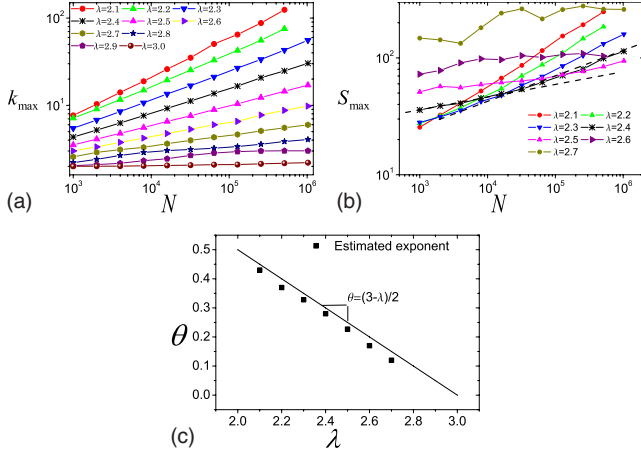


FIG. 6. (Color online) (a) The total number of  $k$ -shells,  $k_{\max}$ , in SF models as a function of  $N$ . Note that  $k_{\max} \propto N^\theta$ . (b) The size of the last  $k$ -shell  $S_{\max}$  in the SF model as a function of  $N$ . Each curve crosses over into a power law regime for  $N \geq N_c(\lambda)$ , where  $N_c(\lambda)$  increases with  $\lambda$ .  $N_c \approx 2 * 10^4$  for  $\lambda \approx 2.4$ . (c) The calculated exponent  $\theta$  as a function of  $\lambda$  (symbols). Our calculated values of  $\delta$  agree with the mean field theory result  $1/\theta \approx \delta = 2/(3-\lambda)$  (solid line).

$k$ -shell  $k_{\max}$  needs to possess at least one node  $S_{k_{\max}} \equiv S_{k=k_{\max}} \sim 1$ , which leads to

$$k_{\max} \leq N^{1/\delta}, \quad (8)$$

Indeed, the total number of  $k$ -shells,  $k_{\max}$ , seems to increase as a power law with the network size  $N$  [Fig. 6(a)]. The smaller is  $\lambda$  the faster is the growth of  $k_{\max}$ . As seen from Fig. 6(c), the estimated exponents  $1/\delta$  agree with those predicted by Eq. (8). Note that Eq. (8) together with Eq. (7) is consistent with the fact that networks with  $\lambda > 3$  do not have a  $k$ -shell structure.

The dependence of  $S_{\max}$  on  $N$  can be regarded as a crossover from the small  $N$  regime with  $N < N_c(\lambda)$ , where there is no nucleus, to the power-law regime for  $N > N_c(\lambda)$  where  $S_{\max} \sim N^{\tau(\lambda)}$  [Fig. 6(b)]. We relate the observed crossover with the emergence of the nucleus and tendrils in SF networks for  $N > N_c(\lambda)$ . The critical size of SF networks,  $N_c(\lambda)$ , corresponding to the emergence of the nucleus and tendrils in SF networks seems to increase as  $\lambda$  increases. In the  $N < N_c$  regime the size of the last  $k$ -shell  $S_{\max}$  increases with  $\lambda$ , which can be explained by the fact that SF networks with higher  $\lambda$  have fewer  $k$ -shells. On the other hand, in the power-law regime, the size of the last  $k$ -shell,  $S_{\max}$ , (which now becomes the nucleus  $S_n$ ) is smaller for larger values of  $\lambda$ .

#### IV. EVOLUTION AND STRUCTURE OF THE LS AND THE ICT INDUSTRY NETWORKS

We further analyze the  $k$ -shell structure of the LS and the ICT industry networks. The number of  $k$ -shells,  $k_{\max}$ , in the LS industry grows as a power-law function of its size,

$$k_{\max} \sim N^\theta, \quad (9)$$

and reaches  $k_{\max} = 19$  in 2002. Our estimates yield  $\theta \approx 0.6$  [Fig. 7(a)]. We find that the number of  $k$ -shells in the ICT

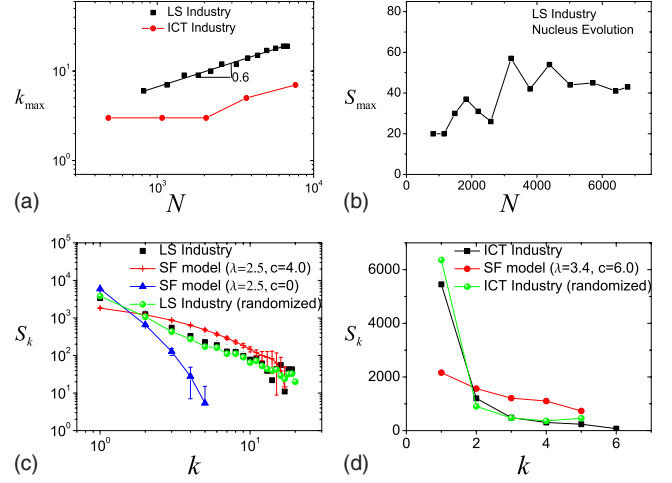


FIG. 7. (Color online) (a) The total number of  $k$ -shells,  $k_{\max}$ , of the LS and the ICT networks as a function of  $N$  calculated for different years. The number of  $k$ -shells  $k_{\max}$  of the LS network increases approximately as a power-law function of the network size  $k_{\max} \sim N^\theta$ , where  $\theta \approx 0.6$ . (b) Size of the nucleus of the LS network,  $S_n$ , as a function of  $N$ .  $S_n$  exhibits fluctuations as the LS network grows. Unlike in the analyzed SF models,  $S_{\max}$  becomes stable for  $N > 5000$ . (c)  $S_k$  as a function of the  $k$ -shell index  $k$  for the LS network (squares).  $k$ -shell sizes  $S_k$  as a function of  $k$  decrease significantly slower than  $S_k \sim k^{-4}$ , which is expected for a random SF model with the same  $\lambda$  and  $c=0$  (triangles). However, the  $c=4.0$  offset in the SF degree distribution,  $P(q)$ , mimics the  $k$ -shell structure of the LS network (pluses).  $k$ -shell sizes,  $S_k$ , of the randomized LS network (circles) almost exactly match  $S_k$  of the original LS network. (d)  $S_k$  as a function of  $k$ -shell index  $k$  for the ICT network (squares). SF model with  $\lambda=3.4$  and  $c=0$  does not possess a  $k$ -shell structure. However, the introduction of  $c=6.0$  in the SF degree distribution yields similar  $k$ -shell structure (circles), but we attribute it only to finite size effect.

sector,  $k_{\max}$ , also grows and reaches  $k_{\max} = 6$  in 2000. Note that  $S_n$  for the LS network exhibits fluctuations for  $N \in [2000, 5000]$  and stabilizes for  $N > 5000$  [See Fig. 7(b)]. As seen in Fig. 7(c) and Fig. 7(d), the sizes of  $k$ -shells  $S_k$  decrease as a function of their index  $k$  in both networks. As we notice in Sec. II, the observed  $k$ -shell sizes  $S_k$  as well as the number of  $k$ -shells  $k_{\max}$ , measured for the LS and the ICT networks, deviate from those obtained for SF models with  $c=0$ . We expect for random SF models with  $c=0$ ,  $\lambda=2.5$  and  $\lambda=3.4$  and similar sizes as the LS and the ICT to find  $k_{\max} = 5$  and  $k_{\max} = 2$  respectively. Also, a SF network with  $\lambda = 2.5$  is expected to have  $S_k \sim k^{-4.0}$  and  $k_{\max} \sim N^{0.25}$  which deviates from the observed  $k$ -shell sizes.

The observed differences in  $S_k$  can be explained by taking into account the offset  $c$  in SF degree distribution of both networks. The adjustment of the degree distribution of random SF models with  $c=4.0$  and  $c=6.0$  allows one to obtain similar patterns  $S_k$  which are in fair agreement with the industry networks, as seen in Figs. 7(c) and 7(d). We note, however, that  $S_k$  of the SF models do not fit precisely  $S_k$  of the corresponding industry networks. In order to achieve a good  $S_k$  fit we construct a random model network with degree sequence identical to that of the LS network. This is achieved by randomly reconnecting links of the industry net-

works while preserving individual degrees of all nodes. As a result, the  $S_k$  of the randomized LS network makes a precise fit to  $S_k$  of the LS network. (The Kolmogorov-Smirnov goodness of fit test in this case yields the  $p$  value of 0.17). Thus, the same shape of the degree distribution in model networks is sufficient to obtain qualitatively similar  $k$ -shell size distribution as in industry networks. However, in order to fine tune the  $S_k$  of the industry networks one needs to preserve not only the shape of degree distribution but also individual degrees of nodes.

Taken together, our results suggest that the offset  $c > 0$  in the SF degree distribution plays a crucial role in the formation of the  $k$ -shell structure of industry networks. A possible reason for the emergence of  $c > 0$  in growing networks is the combination of preferential attachment with random attachment in network evolution [58]. The coexistence of preferential attachment regime with random collaborative agreements was suggested to take place in industry networks [41]. The random component is caused by the fact that sometimes firms choose exclusive relationships and novelty, and do not prefer to make deals with hub firms. Even though the coexistence of both preferential and random regimes seems to be crucial in the formation of the industry networks, it does not fully reproduce the  $k$ -shell structure of the LS and the ICT industries. We believe that a better understanding of the evolution of the LS and the ICT networks may be achieved by further improvements of the modeling.

## V. DISCUSSION AND SUMMARY

We use the  $k$ -shell decomposition to analyze the structure of the LS and the ICT industry networks. We find that the firms in the LS industry can be naturally divided into three components: the nucleus, the tendrils and the bulk body. The nucleus of the LS industry consists mostly of big companies while the tendrils are typically comprised of small start-up firms that preferentially make deals with leading firms which are in the nucleus. We show that the nucleus of the LS industry exhibits remarkable stability in time. In contrast, the ICT industry does not have a nucleus. We also analyzed the dependence of the  $k$ -shell structure of SF model networks on  $N$ ,  $\lambda$  and  $c$ . We observed the formation of the nucleus and the tendrils in SF networks only for  $\lambda < 3$ . The number of  $k$ -shells  $k_{\max}$  and the size of the nucleus  $S_n$  are larger for SF networks with  $c > 0$  compared to those with  $c = 0$ . Our results can partly explain the  $k$ -shell structure of the LS and the ICT industry networks. The coexistence of preferential and random attachment leads to the appearance of the offset  $c > 0$  in the SF degree distribution  $P(q) \sim (q+c)^{-\lambda}$  [58]. The appearance of  $c > 0$  in the degree distribution of the LS and the ICT networks might be explained by the interplay of random and preferential agreements among firms in the industries [41].

## ACKNOWLEDGMENTS

We thank ONR, European project EPIWORK, Merck Foundation (EPRIS project) and Israel Science Foundation for financial support. We thank L. Braunstein, S. Carmi and L. K. Gallos for valuable discussions.

## APPENDIX: INDUSTRY LEADERSHIP AND NETWORK K-CORE MEMBERSHIP

Industrial leadership is measured in terms of market value. We refer to the Business Week 1000 global rankings for year 2001 (ICT) and 2002 (LS) as comprehensive lists of the leading companies worldwide. Due to the dot-com bubble burst in March 2000, we use 2001 data to control for industry turnover. We noticed that the most important companies in pharmaceuticals (Pfizer), computers (IBM), software (Microsoft), media (AOL Time Warner), telecommunication (Verizon), household, and personal equipments (Procter & Gamble), semiconductors (Intel) as well as top conglomerates such as General Electric and Siemens are at the  $k_{\max}$ -shells of the LS and ICT networks of collaboration. In particular 15 out of the top 20 pharmaceutical and IT companies are in the LS  $k_{\max}$ -shell and the ICT  $k_{\max}$ -shell, respectively. Taken together, the companies in the core of the LS network control the 57% of the global pharmaceutical market in terms of market share, and same is true for the global PC and semiconductor industries (ICT network). About 70% of the companies in the  $k_{\max}$ -shells are industrial leaders according to Business Week. We outline firms in the  $k_{\max}$ -shells of the LS and the ICT network in Table I. Other companies in the ICT highest  $k$ -shell not listed in Business Week 2000–2001 are: Akamai, 3Com, Wind River Systems, Netscape, Infospace, PSINet, Inktomi, Virata, SanDisk, Real Networks, Allaire, @home, Net2Phone, Virange, Tut Systems, Concentric Network, Selectica, Earthlink, Palm, Chartered Semiconductor, OmniSky, Stamps.com, Northpoint, MindSpring, Linuxcare, Rambus, AvantGo, Cirrus Logic, 800.com, NeoMagic, INTO Networks, and IteX. Other companies in the LS highest  $k$ -shell not listed in Business Week 2002 are: Elan, Ivax, Incyte, Applera, Pharmacopeia, Lion, Lexicon, Watson, Trega, Vertex and Affymetrix. Boehringer Ingelheim (LS), NIH (LS), Check Point Software (ICT), Samsung (ICT), and UMC (ICT) are leading organizations in the highest  $k$ -shell not covered by Business Week (they have been omitted in the odd ratio tests, as shown below).

We use the odds ratio test to evaluate the efficiency of predicting top firms (according to the Business Week listings) based on the firm's placement in the  $k_{\max}$ -shell. In the

TABLE II. Contingency tables for (a) ICT and (b) LS.

Core	Leader	
	Yes	No
(a) Contingency Table ICT		
Yes	41	32
No	71	7615
Log (odd ratio), 2001:	4.92 ± 0.52	
(b) Contingency Table LS		
Yes	30	11
No	25	6715
Log (odd ratio), 2002:	6.60 ± 0.79	

ICT industry there are  $n_{11}=41$  firms in the  $k_{\max}$ -shell, which are also listed in the Business Week magazine as top firms. On the other hand, there are  $n_{10}=32$  ICT firms which are listed in the magazine that do not appear in the  $k_{\max}$ -shell of the ICT networks. We also count the total of ICT  $n_{01}=71$  firms which are present in the Business Week listings and are not in the  $k_{\max}$ -shell. There are ICT  $n_{00}=7615$  firms which are neither leaders nor in the  $k_{\max}$ -shell. The odds of a firm being a leader given it is in the  $k_{\max}$ -shell are  $p_1=n_{11}/n_{10}$ , while the odds the firm is a leader given it is not in the

$k_{\max}$ -shell are  $p_2=n_{01}/n_{00}$ . The logarithm of the ratio of odds for the ICT industry is  $\ln(p_1/p_2)=\ln\left[\frac{n_{11}n_{00}}{n_{10}n_{01}}\right]\approx 4.92$ . Similarly, the logarithm of the ratio of odds for the LS industry is 6.60. We outline the odds ratio calculations in Table II. The odd ratio tests demonstrate that the presence in the highest  $k$ -shell is strongly and significantly correlated with industrial leadership. The correlation is higher in the LS industry than in the ICT one, especially after the burst of the dot-com bubble. We cross check our results by means of Forbes Global 2000 data [59] without noticing any significant difference.

- 
- [1] P. Erdős and A. Rényi, *Publ. Math. Inst. Hung. Acad. Sci.* **6**, 290 (1959).
- [2] P. Erdős and A. Rényi, *Publ. Math. Inst. Hung. Acad. Sci.* **5**, 17 (1960).
- [3] B. Bollobas, *Random Graphs* (Cambridge University, Cambridge, England, 2001).
- [4] S. Wasserman and K. Faust, *Social Network Analysis* (Cambridge University, Cambridge, England, 1999).
- [5] S. N. Dorogovtsev and J. F. F. Mendes, *Evolution of Networks* (Oxford University, New York, 2003).
- [6] M. Newman, A. L. Barabási, and D. J. Watts, *The Structure and Dynamics of Networks* (Princeton University, Princeton, NJ, 2006).
- [7] S. Bornholdt and H. G. Schuster, *Handbook of Graphs and Networks* (WILEY-VCH, New York, 2001).
- [8] S. Milgram, *Psychol. Today* **2**, 60 (1967).
- [9] D. J. Watts and S. H. Strogatz, *Nature (London)* **393**, 440 (1998).
- [10] R. Albert, H. Jeong, and A.-L. Barabási, *Nature (London)* **401**, 130 (1999).
- [11] R. Cohen and S. Havlin, *Phys. Rev. Lett.* **90**, 058701 (2003).
- [12] R. Cohen and S. Havlin, *Complex Networks: Structure, Stability and Function* (Cambridge University, Cambridge, 2008).
- [13] H. A. Simon, *Biometrika* **42**, 425 (1955).
- [14] D. de S. Price, *J. Am. Soc. Inf. Sci.* **27**, 292 (1976).
- [15] A.-L. Barabási and R. Albert, *Science* **286**, 509 (1999).
- [16] R. Albert and A.-L. Barabási, *Rev. Mod. Phys.* **74**, 47 (2002).
- [17] S. Kirkpatrick, *Rev. Mod. Phys.* **45**, 574 (1973).
- [18] D. Stauffer and A. Aharony, *Introduction to Percolation Theory*, 2nd ed. (Taylor and Francis, New York, 2004).
- [19] A. Bunde and S. Havlin, *Fractals and Disordered Systems*, 2nd ed. (Springer, Berlin, 1996).
- [20] R. Cohen, K. Erez, D. ben-Avraham, and S. Havlin, *Phys. Rev. Lett.* **85**, 4626 (2000).
- [21] D. S. Callaway, M. E. J. Newman, S. H. Strogatz, and D. J. Watts, *Phys. Rev. Lett.* **85**, 5468 (2000).
- [22] G. Paul, R. Cohen, S. Sreenivasan, S. Havlin, and H. E. Stanley, *Phys. Rev. Lett.* **99**, 115701 (2007); Y. Chen, E. Lopez, S. Havlin, and H. E. Stanley, *ibid.* **96**, 068702 (2006).
- [23] C. Song, S. Havlin, and H. Makse, *Nature (London)* **433**, 392 (2005).
- [24] C. Song, S. Havlin, and H. Makse, *Nat. Phys.* **2**, 275 (2006).
- [25] K. I. Goh, G. Salvi, B. Kahng, and D. Kim, *Phys. Rev. Lett.* **96**, 018701 (2006).
- [26] M. Kitsack, S. Havlin, G. Paul, M. Riccaboni, F. Pammolli, and H. E. Stanley, *Phys. Rev. E* **75**, 056115 (2007).
- [27] B. Bollobas, *Graph theory and combinatorics 1988. Proceedings of the Cambridge Combinatorial Conference in Honor of P. Erdős* (Cambridge, UK, 1988).
- [28] S. B. Seidman, *Soc. Networks* **5**, 269 (1983).
- [29] J. I. Alvarez-Hamelin, L. Dall'Asta, A. Barrat, and A. Vespignani, *Networks Heterog. Media* **3**, 371 (2008).
- [30] S. Carmi, S. Havlin, S. Kirkpatrick, Y. Shavitt, and E. Shir, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 11150 (2007).
- [31] J. Chalupa, P. L. Leath, and G. R. Reich, *J. Phys. C* **12**, L31 (1979).
- [32] C. Toninelli and G. Biroli, *J. Stat. Phys.* **126**, 731 (2007).
- [33] B. Pittel, J. Spencer, and N. Wormald, *J. Combin. Theory B* **67**, 111 (1996).
- [34] D. Fernholz and V. Ramachandran, UTCS Technical Report No. TR04-13, 2004 (unpublished).
- [35] S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes, *Phys. Rev. Lett.* **96**, 040601 (2006).
- [36] S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes, *Physica D* **224**, 7 (2006).
- [37] A. V. Goltsev, S. N. Dorogovtsev, and J. F. F. Mendes, *Phys. Rev. E* **73**, 056101 (2006).
- [38] S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes, *Rev. Mod. Phys.* **80**, 1275 (2008).
- [39] D. Fu, F. Pammolli, S. V. Buldyrev, M. Riccaboni, K. Matia, K. Yamasaki, and H. E. Stanley, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 18801 (2005).
- [40] M. Riccaboni, F. Pammolli, S. V. Buldyrev, L. Ponta, and H. E. Stanley, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 19595 (2008).
- [41] L. Orsenigo, F. Pammolli, and M. Riccaboni, *Res. Policy* **30**, 485 (2001).
- [42] F. Pammolli and M. Riccaboni, *Small Bus. Econ.* **19**, 205 (2002).
- [43] M. Riccaboni and F. Pammolli, *Res. Policy* **31**, 1405 (2002).
- [44] J. Sutton, *Am. Econ. Rev.* **97**, 222 (2007).
- [45] R. R. Nelson and S. G. Winter, *An Evolutionary Theory of Economic Change* (Belknap Press of Harvard University, Cambridge, MA, London, England, 1982).
- [46] J. A. Schumpeter, *Capitalism, Socialism, and Democracy* (Harper and Brothers, New York, 2008).
- [47] J. A. Schumpeter, *The Theory of Economic Development* (Harvard Economic Studies, Cambridge, MA, 1982).
- [48] M. Motta, *Competition Policy* (Cambridge University Press, Cambridge, England, 2008).
- [49] Business Week Global 1000, year 2000, <http://>



- [www.businessweek.com/2000/00\\_28/b3689014.htm](http://www.businessweek.com/2000/00_28/b3689014.htm)  
[http://www.businessweek.com/2000/00\\_28/b3689014.htm](http://www.businessweek.com/2000/00_28/b3689014.htm).
- [50] Business Week Global 1000, year 2001, [http://www.businessweek.com/magazine/content/01\\_28/b3740010.htm](http://www.businessweek.com/magazine/content/01_28/b3740010.htm)  
[http://www.businessweek.com/magazine/content/01\\_28/b3740010.htm](http://www.businessweek.com/magazine/content/01_28/b3740010.htm).
- [51] Business Week Global 1000, year 2002, [http://www.businessweek.com/magazine/content/02\\_28/b3791008.htm](http://www.businessweek.com/magazine/content/02_28/b3791008.htm)  
[http://www.businessweek.com/magazine/content/02\\_28/b3791008.htm](http://www.businessweek.com/magazine/content/02_28/b3791008.htm).
- [52] F. Y. Edgeworth, *J. R. Stat. Soc.* **71**, 499 (1908).
- [53] M. E. J. Newman, *Contemp. Phys.* **46**, 323 (2005).
- [54] R. Cohen, K. Erez, D. ben-Avraham, and S. Havlin, *Phys. Rev. Lett.* **86**, 3682 (2001).
- [55] Y. Chen, G. Paul, R. Cohen, S. Havlin, S. P. Borgatti, F. Liljeros, and H. E. Stanley, *Phys. Rev. E* **75**, 046107 (2007).
- [56] L. Tauro, C. Palmer, G. Siganos, and M. Faloutsos, *Global Internet* (Nov. 2001).
- [57] M. Molloy and B. Reed, *Random Struct. Algorithms* **6**, 161 (1995).
- [58] R. Albert and A. L. Barabási, *Phys. Rev. Lett.* **85**, 5234 (2000).
- [59] Forbes magazine, <http://www.forbes.com>.