Nat Comput (2014) 13:379–390
DOI 10.1007/s11047-014-9444-7

A cognitive-inspired algorithm for growing networks

Emanuele Massaro · Franco Bagnoli · Andrea Guazzini · Henrik Olsson

Published online: 11 July 2014

Abstract We present models for generating different classes of networks by adopting simple local strategies and an original model of the evolutionary dynamics and growth of on-line social networks. The model emulates people's strategies for acquiring information in social networks, emphasising the local subjective view of an individual and what kind of information the individual can acquire when arriving in a new social context. We assume that the strategy proceeds through two phases: (a) a discovery phase, in which the individual becomes aware of the surrounding world and (b) an elaboration phase, in which the individual elaborates locally the information trough a cognitive-inspired algorithm. Model generated networks reproduce the main features of both theoretical and real-world networks, such as high clustering coefficient, low characteristic path length, strong division in communities, and variability of degree distributions.

Keywords Complex networks · Computational modelling · Growing networks

1 Introduction

The emergence and the global adaptation of social networks has influenced human interaction on individual, community, and larger social levels. Most notably, perhaps, is the rise of Facebook, which in October 2011 reached more than half (55 %) of the world's global audiences, catching 835.6 millions of users in 2012 (Internet World Stats 2013). Understanding the growth and development of social networks is a task of great

E. Massaro (✉)
Department of Information Engineering and Center for the Study of Complex Systems, University of Florence, Florence, Italy
e-mail: ema.massaro@gmail.com

Present Address:
E. Massaro
Risk and Decision Science Team, US Army Engineer Research Development Center, 696 Virginia Rd, Concord, MA, USA

E. Massaro
Department of Civil and Environmental Engineering, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA, USA

F. Bagnoli
Department of Physics and Astronomy, Center for the Study of Complex Systems, University of Florence, Florence, Italy
e-mail: franco.bagnoli@unifi.it

F. Bagnoli
INFN, Florence, Italy

A. Guazzini
Department of Science of Education and Psychology, Center for the Study of Complex Systems, University of Florence, Florence, Italy
e-mail: andrea.guazzini@unifi.it

H. Olsson
Center for Adaptive Behavior and Cognition, Max Planck Institute for Human Development, Berlin, Germany
e-mail: h.olsson@warwick.ac.uk

Present Address:
H. Olsson
Department of Psychology, University of Warwick, Coventry CV4 7AL, UK

importance in many disciplines, such as sociology, biology, and computer science (Wasserman and Faust 1994; Scott 2000; Dorogovtesev and Mendes 2003; Strogatz 2001; Reka and Barabási 2002), where systems are often represented as graphs. A large number of models have been proposed that aim at exploring and explaining how local mechanisms of network formation produce global network structure. In the context of social networks it is important to understand why and how people decide to make connections and how they change or modify their own local structure. For this reason it is essential to understand some aspects of how humans behave in social networks: how do people acquire information in on-line social networks and what are the mechanisms that lead people to join together or to visit a specific website?

Graph theory was officially founded in the late eighteenth century with the solution of the famous problem of the bridges of Königsberg by the Swiss mathematician (Euler 1741). The technique used by him proved to be of much greater utility than simple puzzle solving. The German physicist Gustav Kirchoff analyzed electric circuits in terms of graphs and chemicals found a natural correspondence between graphs and structures of atoms and molecules. A graph also describes a transportation network, a neural network of the brain of a living being or an economy in which companies are the nodes and the transactions between them are edges. In the twentieth century the theory has assumed a predominantly statistical and algorithmic predisposition (König 1936; Bollobas 2001). Given the representation of the network as a graph, all deductions, measures and indicators of the theory can be applied to this case. The terms graph and network are synonymous in all respects. One of the the most well known mechanism that is used in growing networks is preferential attachment, where new connections are established preferentially to more popular nodes in a network, giving rise to a scale-free network (Barabási and Albert 1999). Moreover, users in on-line social networks tend to form groups, called communities: given a graph, a community is a group of vertices "more linked" among them than between the group and the rest of the graph (Girvan and Newman 2002). This is clearly a poor definition, and indeed, on a connected graph, there is no clear distinction between a community and the rest of the graph. In general, there is a continuum of nested communities whose boundaries are somewhat arbitrary: the structure of communities can be seen as a hierarchical dendrogram (Newman and Girvan 2004). Our communities are large and varied, and we recognize several levels of grouping, sometimes dependent on the context. In recent work we have shown that using information dynamics algorithms where nodes elaborate information locally, we are able to detect such communities in complex networks (Massaro et al. 2012; Bagnoli et al. 2012).

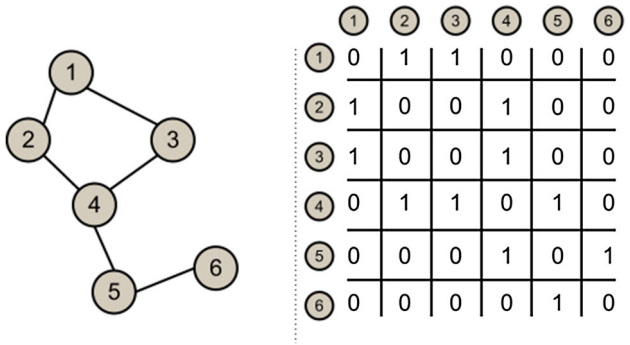
Recently, Papadopoulos et al. (2012) explored the trade-off between popularity and similarity in growing networks. Nodes in growing networks tend to link not only to the most popular nodes (as in preferential attachment, Barabási and Albert 1999) but also to the closest nodes in terms of affinity. Comparing their results with real-world complex networks, the authors showed that they were able to predict the probability of forming new links with remarkable precision.

In this paper we develop models for growing complex networks based on simple local strategies and a cognitive-inspired model that emulates the growing of a social network, starting from psychological assumptions that allow us to simulate how people acquire and elaborate information in social networks. We demonstrate the concept of similarity and popularity in growing networks, not by a geometric approach as in Papadopoulos et al. (2012), but by using a simple mechanism that explain users' behaviour in on-line social networks.

The rest of this paper is organized as follows: we start by introducing our models of growing networks in Sect. 2 with different structures: regular (Sect. 2.1), random (Sect. 2.2) and scale-free (Sect. 2.3) networks. Then we describe models for generating networks with community structure (Sect. 2.4) and finally in Sect. 2.5 we describe our cognitive-inspired model, which uses a local algorithm where an agent is modeled with a memory and a set of connections to other individuals. In the first step the new agent explore the local structure of the network, where it receives information about the neighborhood. The learning (nonlinear) phase is modeled after competition in the chemical/ecological world, where agents compete with each other. This method allow us to generate different kinds of networks by changing model's parameters and to generate real-world networks. Finally, we discuss our results in the Conclusions.

2 Network models

In this section we describe some models for growing complex networks: our goal is to develop simple methods for generating different kind of networks that could be used both for benchmark graphs testing and for reproducing real world scenarios. In particular we are going to show different models for growing three simple classes of complex networks namely, *regular*, *random* and *scale-free*. Then, by adopting different strategies, we present two models for generating hierarchical networks with community structure considering both random and scale-free networks. Finally we show a model for reproducing real-world networks. Formally, a graph $G = (V, E)$ is defined as a pair of sets: the set V of nodes, or vertices, and the set E of the connections between them (also known as edges, links), and each arc is drawn as a line connects two nodes. From the



	1	2	3	4	5	6
1	0	1	1	0	0	0
2	1	0	0	1	0	0
3	1	0	0	1	0	0
4	0	1	1	0	1	0
5	0	0	0	1	0	1
6	0	0	0	0	1	0

Fig. 1 Example of an undirected graph (network) and its mapping on the adjacency matrix

mathematical point of view a graph $G = (V, E)$ can be represented by its adjacency matrix A, as shown in Fig. 1, having nodes as elements of row or column and contains a value other than zero (if different from 1 it represents a "weight" of the connection: cost, speed, energy, etc.) intersection if the two nodes are connected. Then the adjacency matrix for a graph with N vertices is a $N \times N$ matrix whose (i, j) entry is greater than 0 if the i and j are connected and zero otherwise.

In the next subsections we describe three simple mechanisms for generating three-well known classes of networks: *regular* (1-dimensional lattices with periodic boundary conditions), *random* (with normal connectivity degree distribution) and *scale-free* networks. For generating networks with these kind of characteristics we form a fully-connected network with $m + 1$ nodes.

ALGORITHM 1: Algorithm for generating Regular Networks

```
Define m + 1 nodes
for i = 1 to m (At time T = 0 fully connected network) do
    for j = i + 1 to m + 1 do
        Targed node j
        Add Link (i, j)
    end
end
for T = 1 to N − m do
    i = m + i
    for j = 1 to m do
        Incoming node i
        Target node t_n = i − j
        Add Link (i, t_n)
    end
end
for i = 1 to m (for closing the ring) do
    Condition v = m − i + 1
    Selected node i
    for j = 1 to v do
        Target node t_n = N − j + 1
        Add Link (i, t_n)
    end
end
```

2.1 Regular networks

In the case of regular networks at time $t + 1$ the incoming nodes will link to the closest m nodes. For instance if we define $m = 2$, at time $t = 0$ the network is composed by nodes $1, 2, 3$ fully connected, then at time $t + 1$ the new node 4 will link to nodes 2 and 3 and so on until the closing of the circle where the node 1 will receive two more links from nodes going from $N - m$ to N (Algorithm 1). In this way the resulting graph is a regular network with connectivity $k = 2m$.

ALGORITHM 2: Algorithm for generating Random Networks

```
Define m + 1 nodes
for i = 1 to m (At time T = 0 fully connected network) do
    for j = i + 1 to m + 1 do
        Targed node j
        Add Link (i, j)
    end
end
for T = 1 to N − m do
    i = m + i
    for j = 1 to m do
        while i == j or i and j are already connected do
            Target Node t_n = choose a random node from i − 1 nodes
        end
        Add Link (i, j)
    end
end
```

2.2 Random networks

At $t + 1$ the incoming node will link with an uniform probability p to m different nodes. The algorithm (Algorithm 2) is developed for avoiding self-loops and multiple links, so the total number of links in the networks will be $L = 2 * m * N$ and the average connectivity degree is defined by:

$$\langle k \rangle = \frac{1}{N} \sum_{i}^{N} k_i = \frac{1}{N} 2mN = 2m. \tag{1}$$

The generated random networks have a normal distribution of connectivity degrees $P(k)$, where $P(k) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(k-z)^2}{2\sigma^2}}$ and $z = \langle k \rangle$ as shown in Fig. 2. Now we can turn the problem of generating random networks to one of controlling the clustering coefficient. Remember that the clustering coefficient is a function of the number of local triples. Following the definition of Barrat et al. (2004), the clustering coefficient C is defined as the average of the local clustering coefficients c_i:

$$C = \frac{1}{N} \sum_{i}^{N} c_i = \frac{1}{N} \sum_{i}^{N} \left(\frac{1}{k_i(k_i - 1)} \sum_{j,h} a_{ij} a_{ih} a_{jh} \right), \tag{2}$$

where k_i is the connectivity degree of node i. Then in order to increase the number of triples (and then the clustering coefficient) we simply define a *transitivity parameter* p_t which is the probability to link to the adjacent nodes of the first selected nodes instead of another random node. In Fig. 3 we show that the clustering coefficient increases decreasing the average connectivity degree $\langle k \rangle$.

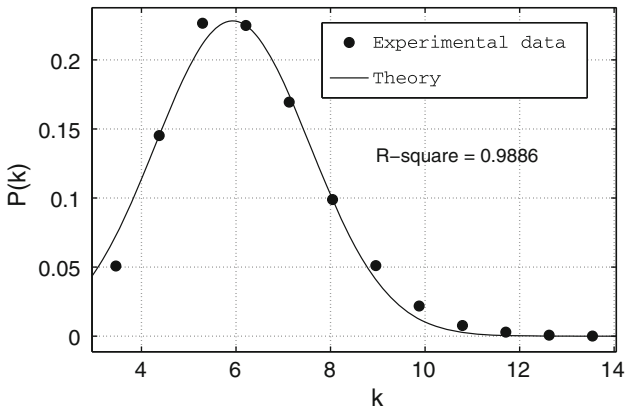


Fig. 2 Frequency distribution of connectivity degree k for a random network. The *points* correspond to the real connectivity distribution of a random network with $N = 10^4$ nodes, $m = 3$ and $\langle k \rangle = 2m = 6$, while the *solid line* correspond to a normal distribution with $\mu = 6$ and $\sigma = 1.73$

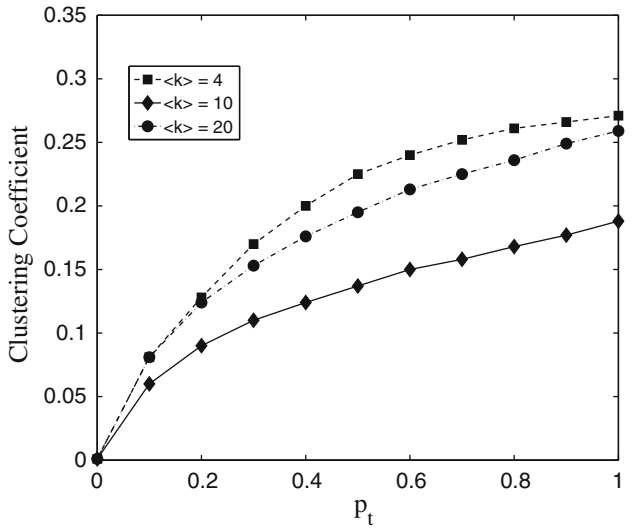


Fig. 3 Clustering coefficient (y-axis) as function of *transitivity parameter* p_t. The three different lines correspond to three random networks composed by $N = 10{,}000$ nodes and average connectivity degree $\langle k \rangle = 4, 10, 20$ as shown in the legend

2.3 Scale-free networks

Also here we start with a network composed by m connected nodes, then at each time step each node the new node establishes m links according to the following mechanism:

1. pick a random node in a network (r_n);
2. link to a random adjacent node of r_n (if not already connected);
3. repeat 1, 2 until it makes m links.

this mechanism allow us to generate scale-free networks ($P \sim k^{\gamma}$) with exponent $\gamma = -3$. Also in this case, the total

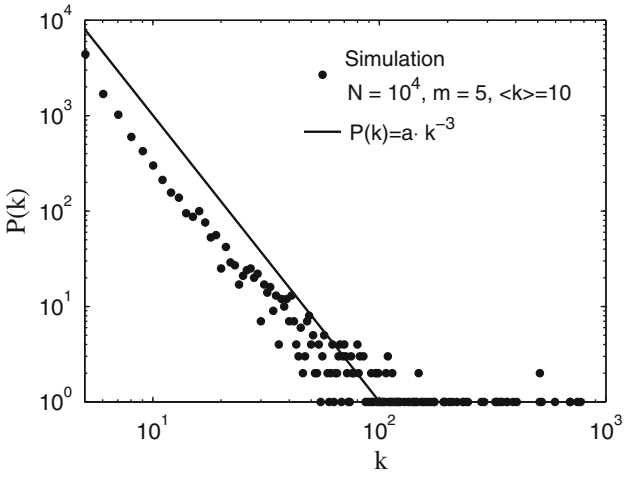


Fig. 4 Frequency distribution of connectivity degree k for a scale-free network. The *points* correspond to the real connectivity distribution of a random network with $N = 10^3$ nodes, $m = 5$ and $\langle k \rangle = 2m = 10$, while the *solid line* correspond to a power-law with $\gamma = -3$

number of links is $L = 2mN$ and so $\langle k \rangle = 2m$ as reported in Fig. 4.

2.4 Networks with community structure

2.4.1 Hierarchical networks

We start with a symmetric adjacency matrix with no self loops, i.e., $A(i,j) = A(j,i)$ and $A(i,j) = 0$. For evaluating the performance of a community detection algorithm (Massaro et al. 2012; Bagnoli et al. 2012; Massaro and Bagnoli 2014) we have developed a method for generating hierarchical benchmark networks with community structure as shown in Fig. 5. In particular we start by defining the number of vertices: then we choose the number of sublevels. Once we have defined it we define the probability of creating a link in each sub level starting from the bottom one. In the example we show a 3-levels network with 120 nodes. It has been noticed that the hierarchical structure of the network, which is evident in Fig. 5, is generally obfuscated by the arbitrary labeling of nodes. If one changes the ordering of the nodes, the corresponding adjacency matrix will no more easily reveal this structure.

2.4.2 Random and scale-free networks

There are n_c different communities with s_1 vertices (here we consider only undirected and unweighted graphs); we assume that the probability to have a link between the vertexes in the same community is p_1, while p_2 is the probability to have a link between two nodes belonging to different communities. For instance, with $p_1 = 1$ and

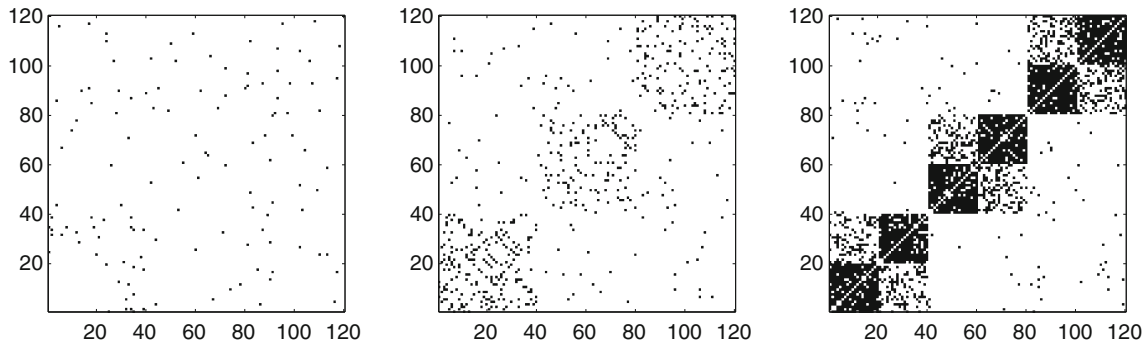


Fig. 5 From *left to right* we can observe the evolution of the creation of hierarchical networks with community structure. This a 3 level network with 120 nodes where the bottom probability $p_1 = 0.01$, the intermediate level probability is $p_2 = 0.3$ and the higher one is $p_3 = 0.9$

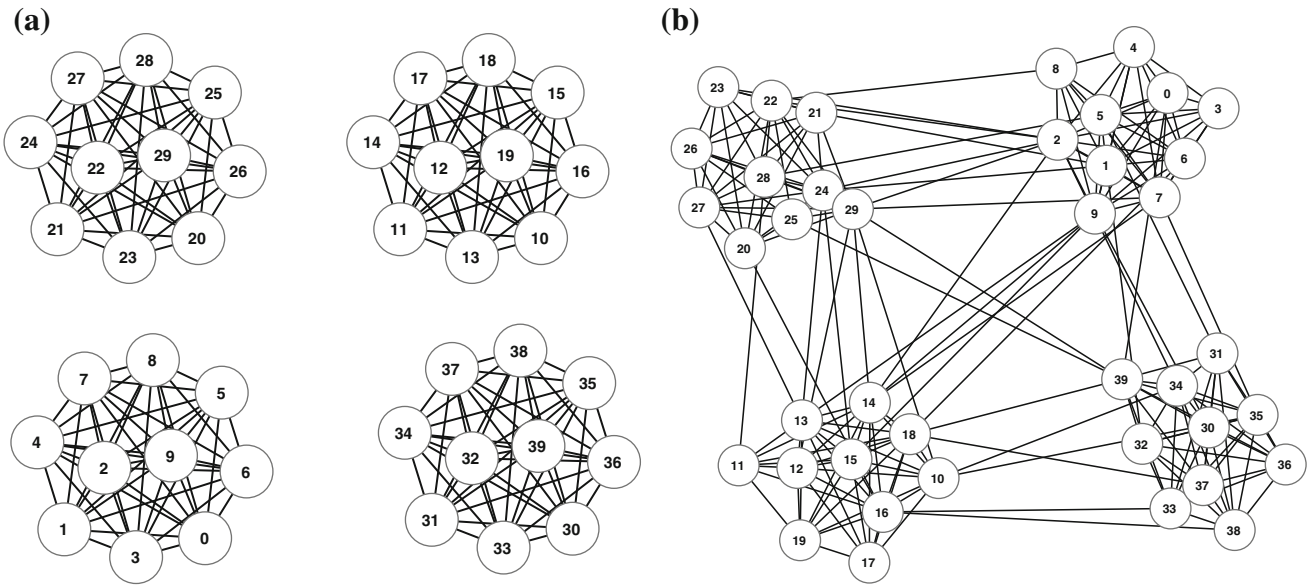


Fig. 6 **a** An example network with 4 different communities composed by 10 vertices: in this case, considering $p_1 = 1$ and $p_2 = 0$, we generate 4 non-interconnected fully connected networks. **b** The same 4 communities with parameters $p_1 = 0.95$ and $p_2 = 0.05$

$p_2 = 0$, we generate n_c fully connected graphs, with no connections among them as shown in Fig. 6a. It is possible to use the parameters p_1 and p_2 to control the interaction among different communities, as shown in Fig. 6b. The algorithm for generating this kind of networks can be summarized as:

1. Define s_1 as number of vertexes in the communities;
2. Define n_c as number of communities;
3. For all the n_c communities create a link between the vertexes on them with probability p_1;
4. With probability p_2, for all the $N = s_1 n_c$ vertexes generate a link with a random node of another community.

Given the condition $p_1 = 1 - p_2$, we can reduce the free parameters to just one. The connectivity degree itself depends on the size of the network and on the probabilities p_1 and p_2. In particular, the connectivity function $f(k)$ has a normal distribution from which we can define the mean connectivity $\langle k \rangle$ as

$$\langle k \rangle = (s_1 - 1)p_1 + (n_c - 1)s_1 p_2 \tag{3}$$

with variance $\sigma^2(k) = (s_1 - 1)p_1(1 - p_1) + (n_c - 1)s_1 p_2(1 - p_2)$.

In Fig. 7a we show the frequency distribution of the connectivity degree of nodes varying the value of the parameter p_2 for a network composed by $N = 5{,}000$ nodes and $n_c = 5$ communities.

Real-world networks from social networks to computer networks are generally regarded to be scale-free networks, whose degree distribution follows a power law, at least asymptotically. In this network, the probability distribution of contacts often exhibits a power-law behavior:

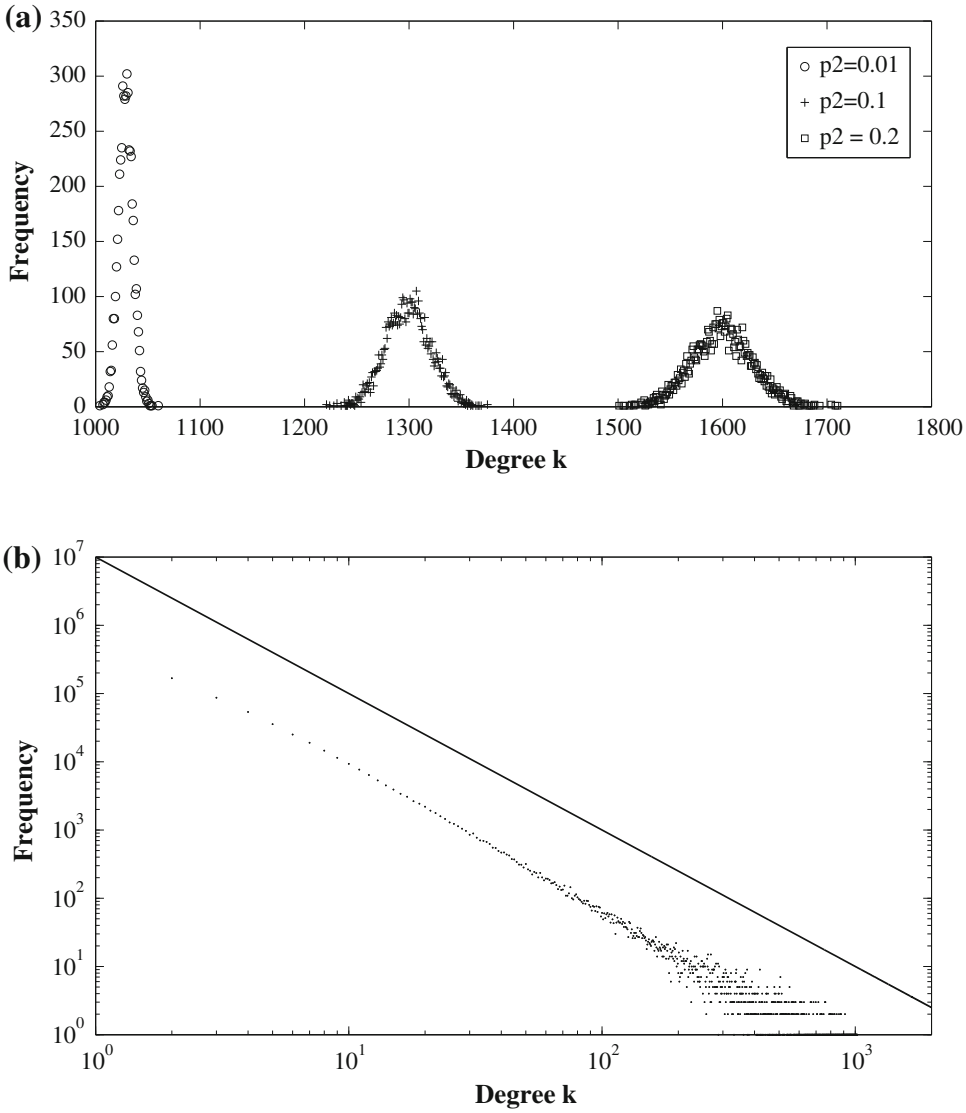


Fig. 7 **a** Random networks: in this figure, we show the frequency distribution of the connectivity degree changing the value of the parameter p_2. The *circles* represent the values for $p_2 = 0.01$, *crosses* for $p_2 = 0.1$ and eventually squares for $p_2 = 0.2$. Here, $s_1 = 1{,}000$ and $n_c = 5$, thus we have generated networks with 5 communities of 1,000 nodes for each. **b** Distribution of connectivity degree for the scale-free network generated with the mechanism described above (*dots*). The *straight line* is a power law curve with exponent $\gamma = 2.5 \pm 1$

$$P(k) \propto ck^{-\gamma}, \tag{4}$$

with an exponent γ between 2 and 3 (Barabási and Albert 1999). For generating networks with this kind of characteristics, we adopt the following mechanism:

1. Start with a fully connected network of m nodes;
2. Add $N - m$ nodes;
3. For each new node add m links;
4. For each of these links choose a node at random from the ones already belonging to the network and attach the link to one of the neighbors of that node, if not already attached.

Through this mechanism we are able to generated scale-free networks with an exponent $\gamma = 2.5$ as shown in Fig. 7b. There, we show the frequency distribution of the connectivity degree for a network of 10^6 nodes. To generate a community structure with a realistic distribution, we first generate n_c scale-free networks as explained above. Then, for all nodes and all outgoing links, we replace the link pointing inside the community with that connecting a neighbor of a random node in a random community with a probability of $p_2 = 1 - p_1$. Thus, the algorithm can be summarized as:

1. Generate n_c communities as scale-free networks with s_1 vertices;
2. For all the vertices, with a probability $p_2 = 1 - p_1$;
   - Delete a random link;
   - Select a random node of another community and create a link with one of its adjacent vertex;
3. End.

In this way, we are able to generate scale-free networks with a well defined community structure. A good measure for the estimation of the strength of the community

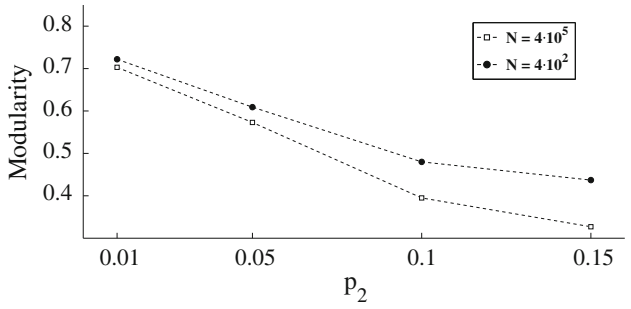


Fig. 8 Different values of modularity (Q) after increasing the mixing parameter p_2 for two networks with $N = 4 \cdot 10^5$ nodes and $N = 4 \times 10^2$ nodes

structure is the so-called modularity (Girvan and Newman 2002). The modularity Q is defined to be:

$$Q = \frac{1}{2}\sum_{vw}\left[A(i,j) - \frac{k_i k_j}{2l}\right]\delta(c_i, c_j), \tag{5}$$

where $l = \frac{1}{2}\sum_{ij} A(i,j)$ is the number of edges in the graph, k is the nodes' connectivity degree and $(k_i k_j)/(2l)$ represents the probability of an edge existing between vertices i and j if connections are made at random but respecting the vertex degrees. The quantity $\delta(c_i, c_j)$ is defined as

$$\delta(c_i, c_j) = \sum_{r}^{n_c} \hat{c}_{ir}\hat{c}_{jr}, \tag{6}$$

where $\hat{c}_{ir}$ is 1 if vertex i belongs to group r, and 0 otherwise.

In Fig. 8 we show the values of modularity for two networks that were generated with the same algorithm, but with different sizes. Here, we consider a network with 4 communities: in the first case $s_1 = 10^5$, while in the second case $s_1 = 10^2$. What one can observe in Fig. 8 is that the modularity's behaviour does not change significantly for different network sizes with the same number of communities.

In the case of scale-free networks, the mean connectivity degree $\langle k \rangle$ is fixed a priori when we choose the number of links the new nodes create. In the case of random networks the mean connectivity is given by equation 3.

2.5 Real world networks

Our mechanism for generating real-world networks is based on a mechanism that emulates people's strategies for acquiring information in social networks, emphasising the local subjective view of an individual and what kind of information the individual can receive when arriving in a new social context (Fig. 9b).

We start from a simple assumption: we suppose that a new individual, or node, arrives in an already structured

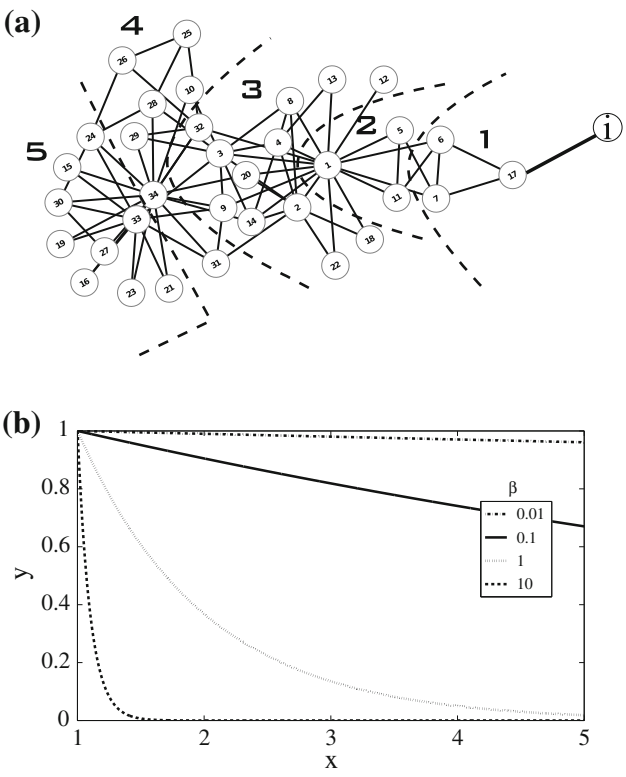
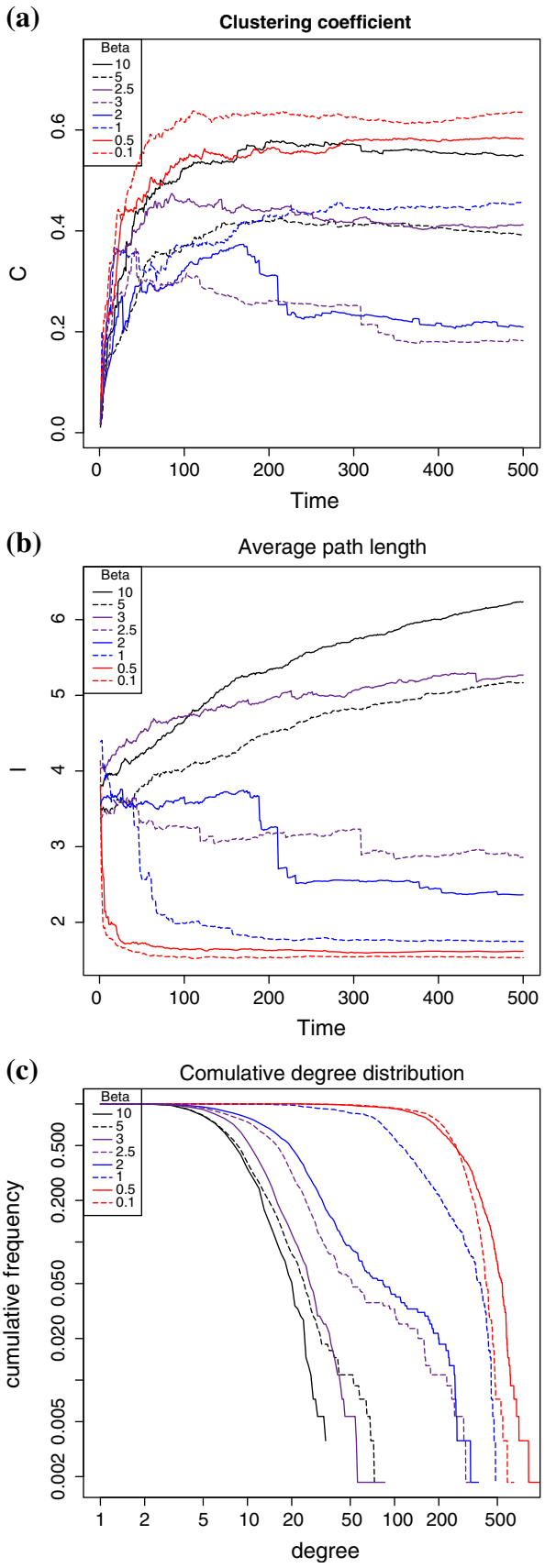


Fig. 9 **a** Starting configuration of the *Zachary's Karate Club* network (Zachary 1977): the new node i links to node 17. The new node has the local hierarchical representation of the network, labelled by levels $1, 2, 3, 4$ and 5. With probability p_1 it could link to nodes of level 1, with probability p_2 to nodes of level 2, and so on. **b** Different probability functions derived from Eq. 7 for different values of parameter β and $a = 1$. The y-axis represents the probability to join with some nodes in the corresponding level shown on the x-axis

network and explores it. According to this idea we begin, at time $t = 0$, with a fully-connected network of N_0 vertexes and then, at each time step, a new node is added to the network. For explaining our idea we describe a simple case, illustrated in Fig. 9a, representing a new node i that joins the *Zachary's Karate Club* network (Zachary 1977) (assuming that the node i is invited by node 17). It discovers n ordered levels: the nodes in level 1 are those adjacent to the connected node (node 17 in Fig. 9a), nodes in level 2 are the friends of one's friends, and so on. With different probabilities $p_1 > p_2 > \ldots > p_n$ it links to nodes in the different levels. This mechanism implements the assumption that the probability of getting new friends in a social context is strictly correlated with the local structure of the network; it is easier that two persons become friends if they a have a friend in common. In this paper we use a simple exponential function, that allows to us to define the probability for the incoming to join with the different levels of network as shown in Fig. 9a, b. Then:

$$y = ae^{-\beta(x-1)}, \tag{7}$$



◀**Fig. 10** Results of **a** clustering coefficient, **b** average path length and **c** final cumulative degree distribution time for different values of the parameter β with $a = 0.95$. (Color figure online)

where x is the considered level. The β represent the "temperature": the probability of joining farther nodes, and a is a normalization constant. In Fig. 9b we show the function (7) for some values of β with $a = 1$. Assuming $-\beta(x-1) = z$ and $e^z = b^z$, we can express the probability distribution of Eq. 7 as $\sum_{z=0}^{\infty} ab^z = a\frac{1}{b-1}$. Setting the previous equation equal to 1 we obtain $a = 1 - b$.

Then, $y = (1-b)b^z = \left(1 - e^{-\beta}\right)e^{-\beta(x-1)} = P(x)$. Because of it is a negative exponential function, by increasing the parameter β the probability to join other levels decreases as shown in Fig. 9 where we reported some results changing the value of β with $a = 0.95$.

Wee can observe very interesting patterns in the Fig. 10a, b corresponding respectively to the mean clustering coefficient and to the average path length during time. In Fig. 10a we observe that the clustering coefficient decreases with increasing values of β from 10 (black continuous line) to 5 (black dashed line) to 2.5 (purple continuous line) to 2 (violet dashed line). For $\beta < 2.5$ the clustering coefficient tends to increase with the maximum values for $\beta = 0.1$ (dashed red line). We observe an analogous trend for the average path length (Fig. 10b) in which for $\beta \geq 3$ it tends to increase during time while we observe a decreasing of this value for $\beta < 3$.

Let us now to introduce an information elaboration mechanism in which each individual i is characterized by a knowledge vector $S_{(i)}$, representing his knowledge of the world. The knowledge vector $S_{(i)}$ is a probability distribution, assuming that $S(i,j)$ is the probability that individual i knows about the community j. It can also be seen as the probability that i belongs to the community "leaded" by j, and therefore, $S(i,j)$ is normalized over the index j. In order to use a compact notation, we arrange the knowledge vectors for all individuals column by column forming a knowledge matrix $S = S^{(t)}$ of the whole network at time t. We initialize the system by setting $S(i,j)(0) = \delta_{ij}$, where δ is the Kronecker delta, $\delta_{ij} = 1$ if $i = j$ and zero otherwise. In other words, at time 0 each node knows only about itself.

The dynamics of the network is given by an alternation of communication and elaboration phases. The communication is implemented as a simple diffusion process, with memory ζ. The memory parameter ζ allows us to introduce some important features of the human cognitive system, for example that recently acquired information have more relevance than information gained in the past (Tulving et al. 1982; Forster and Davis 1984).

In the communication phase, the state of the system evolves as

$$S(i,j)^{(t+1/2)} = \zeta S(i,j)^{(t)} + (1-\zeta)\sum_k A(i,k)S(k,j)^{(t)}, \quad (8)$$

where A is the adjacency matrix. We assume that nodes talk with each other and that nodes with high connectivity degree have greater influence in the process of information's diffusion. This is due to the fact that during a conversation it is more likely to know a vertex with high degree instead of one which has a few links. For this reason, the information dynamics is a function of the adjacency matrix a. The elaboration phase implements elements of fast and frugal heuristics (Gigerenzer and Gaissmaier 2011). When people are asked to take a decision, very rarely do they weight all available pieces of information. If there is some aspect that has a higher importance than others, and one item exhibits it, than the decision is taken, otherwise, the second most important factor is considered, etc. In order to implement an adaptive scheme, we exploit a similarity with the competition dynamics among species.

If two populations x and y are in competition for a given resource, their total abundance is limited. After normalization, we can assume $x + y = 1$, i.e., x and y are the frequency of the two species, and $y = 1 - x$. The reproduction phase is given by $x' = f(x)$, which we assume to be represented by a power $x' = x^\alpha$. For instance, $\alpha = 2$ models birth of individuals of a new generation after binary encounters of individuals belonging to the old generation, with non-overlapping generations (Nicosia et al. 2011).

After normalization $x' = \frac{x^\alpha}{x^\alpha + y^\alpha} = \frac{x^\alpha}{x^\alpha + (1-x)^\alpha}$ and introducing $z = (1/x) - 1$ $(0 \le z < \infty)$, we get the map $z(t+1) = z^\alpha(t)$, whose fixed points (for $\alpha > 1$) are 0 and ∞ (stable attractors) and 1 (unstable), which separates the basins of the two attractors. Thus, the initial value of x, x_0, determines the asymptotic value, for $0 \le x < 1/2$ $x(t \to \infty) = 0$, and for $1/2 < x < 1$ $x(t \to \infty) = 1$. By extending to a larger number of components for a probability distribution $S^{(i)}$, the competition dynamics becomes

$$S_{ij}(t+1) = \frac{S_{ij}(t+1/2)^\alpha}{\sum_k S_{ik}^\alpha (t+1/2)}, \quad (9)$$

and the iteration of this mapping, for $\alpha > 1$, leads to a Kronecker delta, corresponding to the largest component. The parameter α allows us to model a "pruning effect" of the information, which eliminates unnecessary clutter and a clears the way for more information to enter the field of view of the individuals. The convergence time depends on the relative differences among the components and therefore, when coupled with the information propagation phase, it can produce interesting behaviours. The model has two free parameters, the memory m and the exponent α.

Finally, the probability of making a new link (P_n) depends on the joint probabilities of two functions $f(y)$ and $g(S)$: $P_n = f(y) \cdot g(S)$.

In summary: (1) we start with m_0 nodes ($m_0 \ge 1$); (2) at time t a new node, labelled by t, appears in the network; (3) the new node connects with a random node in the network, discovering n levels; with probability p_1 given by Eq. 7 it joins the selected level; (4) the new node links with probability p_2 given by the Eq. 9 to the level's nodes. In this way we take into account the social closeness because of the probability to link to nodes in the network depends on the social distance from the *closest* friend *and* the popularity of nodes given by the information dynamics procedure. Results obtained with the information dynamics algorithm, applied to the network represented in Fig. 9a, are shown in Table 1. It is more likely that a new node will be connected to a node with a high degree, as is also predicted by preferential attachment (Fig. 11).

In order to validate our model we compare predictions from the model with two real networks. The model predictions are averages over ten simulation runs. The first social network is the *Political Blogs Network*. It is a directed network of hyperlinks, with $N = 1{,}490$ nodes, among weblogs on US politics, recorded in 2005 by Adamic and Glance (2005). The degree distribution of the resulting network from a simulation of the model with the same number of nodes is comparable with the real network as shown in Fig. 12a. The mean clustering coefficient of the network generated with our simulations is $C = 0.24$, the average path length is $l = 3.23$, and we obtained a network with the same diameter, $d = 9$.

The second social network contains friendships between users of the website hamsterster.com (The Koblenz Network Collection 2012). The degree distribution of the network is shown in Fig. 12b. The cumulative frequency of the degree distribution is very similar to the network generated with the model (black points in Fig. 12b). The mean clustering coefficient $C = 0.09$, the average path length is

Table 1 Results from the information dynamics algorithm for the different levels in the network shown in Fig. 9a

L	n	p
1	6, 7	0.5–0.5
2	**1**, 5, 11	0.97
3	2, **3**, 4, 8, 9, 12, 13, 14, 18, 20, 22, 32	0.35
4	10, 25, 26, 28, 29, 31, 33 **34**	0.67
5	15, 16, 19, 21, 23, **24** , 27, 30	0.47

L is the number of the level, n is the id of the node and p is the probability to join with the most connected node (bold node)

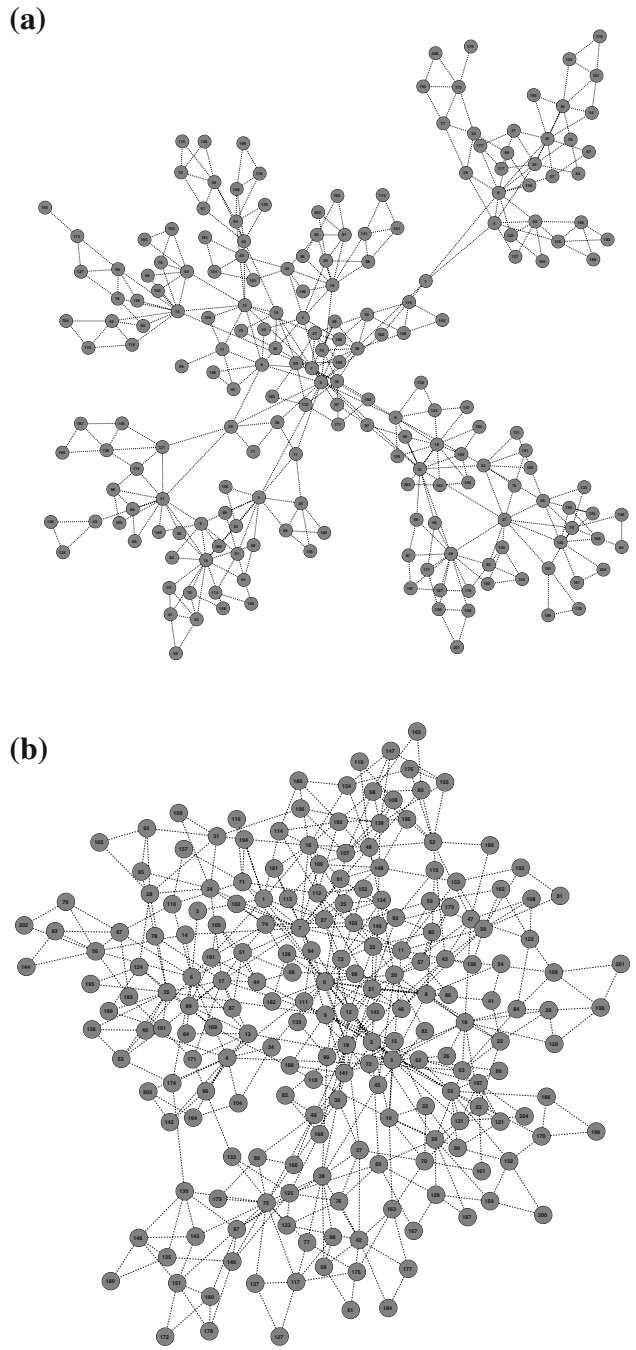


Fig. 11 Two visualizations of networks with $N = 200$ nodes starting from a networks of $N = 5$ nodes. In these two cases the new node links to a random initial node than to one random node of each level with a given probability depending on the parameters β and a. **a** Here $a = 0.95$ and $\beta = 3$: the snapshot indicates strong community structure with communities of various sizes clearly visible. **b** In this case, using $a = 0.95$ and $\beta = 1$, we can not distinguish a communities structure in the network: the decreasing of the parameter β allows to the new node to link with high probability to several levels. Visualization was done using Cytoscape

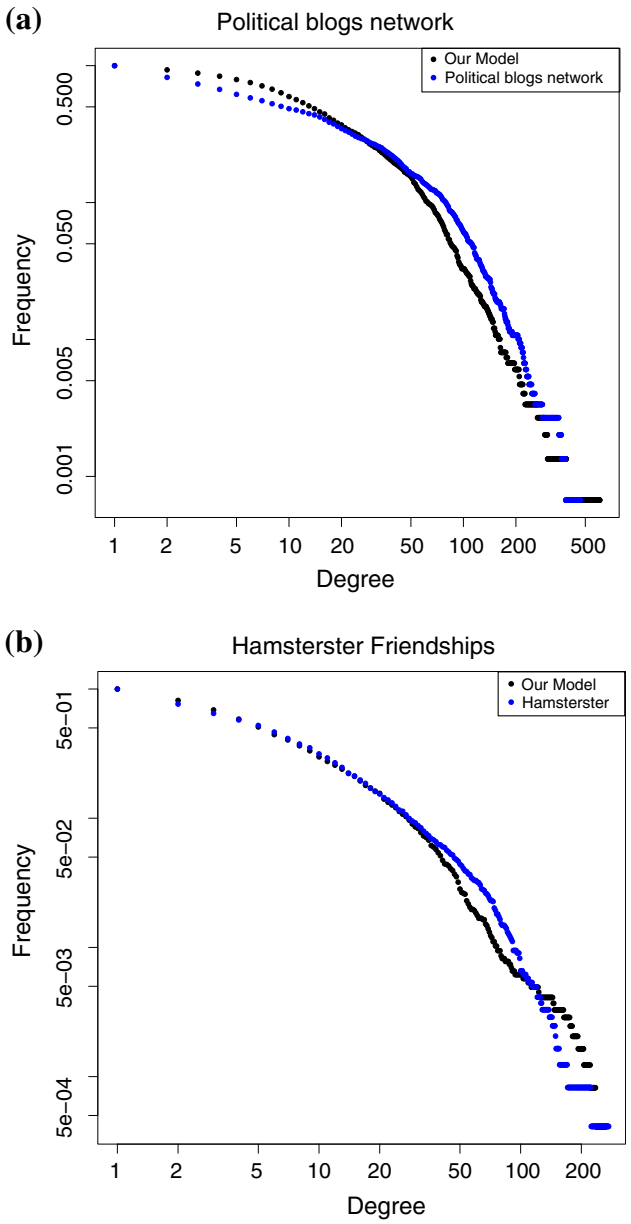


Fig. 12 **a** Cumulative frequency of node degree distributions (log–log scale) of Political Blogs Network (*blue points*) and model generated predictions (*black points*). **b** Cumulative frequency of node degree distributions (log–log scale) of the network of friendships between users of the website hamsterster.com and model generated predictions (*black points*). Model predictions are averaged over 10 simulation runs. (Color figure online)

$l = 3.84$, and the diameter $d = 11$. Another important feature of the model is that the final structure of the network spontaneously arises without any constraint on the degree of the new node. In fact in our simulations we don't assume any constraints on the number of links that the new node can establish: in principle the new node can establish links with all the nodes in the network.

In Fig. 13 we show the results of a simulation of the temporal evolution of network density for the *Political Blogs Network*. The network density is here define as number of links created by the incoming nodes during the

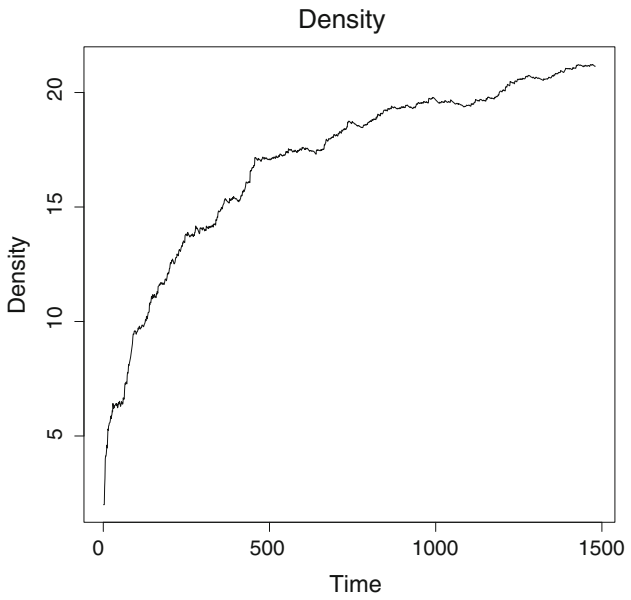


Fig. 13 Simulation of the temporal density evolution of the *Political Blogs Network*

Table 2 Statistics of the social networks: C (mean clustering coefficient), l (average path length) and d (diameter of the network)

	Vertices	Edges	C	l	d
Political Blogs	1,490	19,090	0.24	3.39	9
Simulation	1,490	15,240	0.24	3.23	9
hamsterster.com	2,426	12,534	0.09	3.54	10
Simulation	2,426	12,785	0.09	3.84	11

simulation. From this result, we show that the number of links of the incoming nodes spontaneously increase over the time without any constraint: in real scenarios the number of links that the new node can do depends on the number of nodes already listed in the network at a certain time (for instance in a network of 10 nodes the probability to make a certain number of link is less than in a network of 1,000 nodes). It is a well known result that the density increases in networks that grow over time. It is a very interesting result of our model, because in most methods for emulating growing networks the number of links of the incoming nodes is a fixed parameters that does not change over the time (Papadopoulos et al. 2012)

3 Conclusions

In this paper we presented some simple models for generating different classes of networks by adopting simple local strategies. We also presented a cognitive-inspired growing algorithm that reproduces well the observed dynamics of real networks. The model is based on an information-processing procedure for which new nodes upon their first contacts get information about the local network structure, process them in a competitive way (modeling bounded cognitive resources and the oblivion effect) and choose new nodes with which links are established. Although the model does not put any limit on the number of links, a growing structure very similar to real existing networks emerges.

We varied the parameters of our method to match the characteristic observed in real on-line social networks. We observed interesting correlation between to the mean clustering coefficient and the average path length during time.

For political blog networks, the degree distribution of the network from our simulations showed to be comparable with the real network as shown in Fig. 12a. Also for the friendship Hamster social networks we found that the experimental cumulative frequency is very similar to the network generated with our model as reported in Fig. 12b. Moreover, in Table 2, we reported a quantitative comparison of size of networks, number of generated links, clustering coefficient, average path length and diameter between our generated networks and real ones.

Another important feature of our model is that the final structure of the network spontaneously arise without any constraint on the network. Finally, we were also able to reproduce a well known result that the density increases in networks that grow over time.

References

Adamic LA, Glance N (2005) The political blogosphere and the 2004 U.S. Election: divided they blog. In: Proceedings of the 3rd international workshop on Link discovery (LinkKDD '05). ACM, New York, pp 36–43

Bagnoli F, Massaro E, Guazzini A (2012) Community-detection cellular automata with local and long-range connectivity. In: Sirakoulis GC, Bandini S (eds) Cellular automata. Lecture notes in computer science, vol 7495. Springer, Berlin, pp 204–213

Barabási AL, Albert R (1999) Emergence of scaling in random networks. Science 286(5439):509–512

Barrat A, Barthlemy M, Pastor-Satorras R, Vespignani A (2004) The architecture of complex weighted networks. PNAS 101(11): 3747–3752

Bollobas B (2001) Random graphs. Cambridge University Press, New York

Dorogovtesev SN, Mendes JFF (2003) Evolution of networks. Oxford University Press, Oxford

Euler L (1741) Solutio problematis ad geometriam situs pertinentis. Commentarii academiae scientiarum Petropolitanae 8:128–140

Forster KI, Davis C (1984) Repetition priming and frequency attenuation. J Exp Psyc Learn Memory Cogn 10(4):680–698

Gigerenzer G, Gaissmaier W (2011) Heuristic decision making. Annu Rev Psychol 62:451–482

Girvan M, Newman MEJ (2002) Community structure in social and biological networks. PNAS 99:7821–7826

Internet World Stats (2013) www.internetworldstats.com/facebook.htm. Accessed 1 Feb 2013
König D (1936) Theorie der endlichen und unendlichen Graphen, Liepzig
Massaro E, Bagnoli F (2014) Hierarchical community structure in complex social networks. Acta Physica Polonica 7(2):379–393
Massaro E, Bagnoli F, Guazzini A, Lió P (2012) Information dynamics algorithm for detecting communities in networks. CNSNS 17(11):4294–4303
Newman MEJ, Girvan M (2004) Finding and evaluating community structure in networks. Phys Rev E 69:026113
Nicosia V, Bagnoli F, Latora V (2011) Impact of network structure on a model of diffusion and competitive interaction. EPL 94:68009
Papadopoulos F, Kitsak M, Serrano M, Bogu M, Krioukov D (2012) Popularity versus similarity in growing networks. Nature 489:537–540
Reka A, Barabási AL (2002) Statistical mechanics of complex networks. Rev Mod Phys 74:47–97
Scott J (2000) Social networks analysis: a handbook, 2nd edn. Sage, London
Strogatz SH (2001) Exploring complex networks. Nature 410(6825): 268–276
The Koblenz Network Collection (2012) http://konect.uni-koblenz.de. Accessed 05 May 2012
Tulving E, Schacter DL, Stark HA (1982) Priming effects in word fragment completion are independent of recognition memory. J Exp Psyc Learn Memory Cogn 8(4):336–342
Wasserman S, Faust K (1994) Social networks analysis. University Press, Cambridge
Zachary WW (1977) An information flow model for conflict and fission in small groups. J Anthropol Res 33:452–473