Knowledge acquisition by networks of interacting agents in the presence of observation errors

J. B. Batista and L. da F. Costa
Institute of Physics at São Carlos, University of São Paulo, P.O. Box 369, São Carlos, São Paulo 13560-970, SP, Brazil
(Received 8 May 2009; revised manuscript received 31 March 2010; published 2 July 2010)

In this work we investigate knowledge acquisition as performed by multiple agents interacting as they infer, under the presence of observation errors, respective models of a complex system. We focus the specific case in which, at each time step, each agent takes into account its current observation as well as the average of the models of its neighbors. The agents are connected by a network of interaction of Erdős-Rényi or Barabási-Albert type. First, we investigate situations in which one of the agents has a different probability of observation error (higher or lower). It is shown that the influence of this special agent over the quality of the models inferred by the rest of the network can be substantial, varying linearly with the respective degree of the agent with different estimation error. In case the degree of this agent is taken as a respective fitness parameter, the effect of the different estimation error is even more pronounced, becoming superlinear. To complement our analysis, we provide the analytical solution of the overall performance of the system. We also investigate the knowledge acquisition dynamic when the agents are grouped into communities. We verify that the inclusion of edges between agents (within a community) having higher probability of observation error promotes the loss of quality in the estimation of the agents in the other communities.

DOI: 10.1103/PhysRevE.82.016103 PACS number(s): 89.75.Fb, 07.05.Mh, 64.60.aq, 01.40.Ha

I. INTRODUCTION

Several important systems in nature, from the brain to society, are characterized by intricate organization. Being naturally related to such systems, humans have been trying to understand them through the construction of models which can reasonably reproduce and predict the respectively observed properties. Model building is the key component in the scientific method. The development of a model involves the observation and measurement of the phenomenon of interest, its representation in mathematical terms, followed by simulations and respective confrontation with further experimental evidences. Because of the challenging complexity of the remaining problems in science, model building has become intrinsically dependent on collaboration between scientists or *agents*. The problem of multiple-agent knowledge acquisition and processing has been treated in the literature (e.g., [1,2]) but often under assumption of simple schemes of interactions between the agents (e.g., lattice or pool). Introduced recently, complex networks [3–7] have quickly become a key research area mainly because of the generality of this approach to represent virtually any discrete system, allied to the possibilities of relating network topology and dynamics. As such, complex networks stand out as being a fundamental resource for complementing and enhancing the scientific method.

The present study addresses the issue of modeling how one or more agents (e.g., scientists) progress while modeling a complex system. We start by considering a single agent and then proceed to more general situations involving several agents interacting through networks of relationships (see Fig. 1). The agents investigating the system (one or more) are allowed to make observations and take measurements of the system as they develop and complement their respective individual models. Errors, incompleteness, noise, and forgetting are typically involved during a such model estimation. The main features of interest include the quality of the obtained models and the respective amount of time required for their estimation. The plural in "models" stands for the fact that the models obtained respectively by each agent are not necessarily identical and will often imply in substantial diversity. Though corresponding to a largely simplified version of real scientific investigation, our approach captures some of the main elements characterizing the involvement of a large number of interacting scientists who continuously exchange information and modify their respective models and modeling approaches. As a matter of fact, in some cases the development of models may even affect the system being modeled (e.g., the perturbation implied by the measurements on the analyzed systems).

Because interactions between scientists can be effectively represented in terms of complex networks (e.g., [8–14]), it is natural to resource to such an approach in our investigation. It is interesting to observe that the agents may not be limited

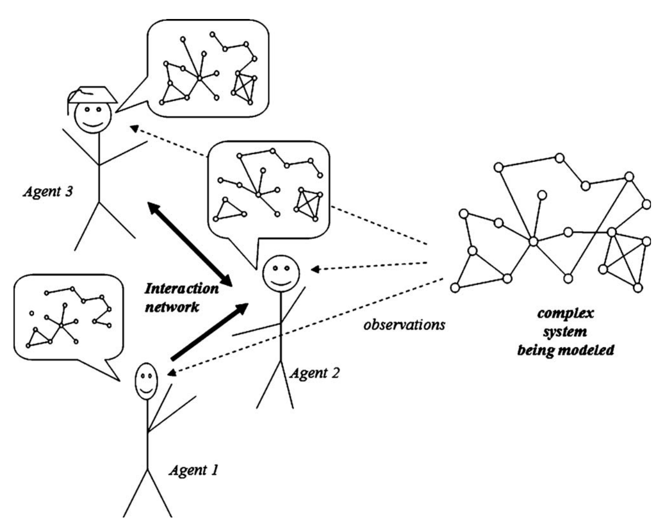

FIG. 1. Agents (scientists) develop their models of complex systems through observation and interactions.

to scientists but can also include intelligent machines and even reference databases and libraries. Though most of the previous approaches to modeling scientific interaction in terms of complex networks have focused on the topology of the collaborations, fewer strategies (e.g., [15]) have addressed the important issue of how the dynamics of learning or knowledge acquisition evolves in such systems, especially with respect to distinct patterns of connectivity between the agents. This is the main motivation of the current work.

In the last years, great interest has been devoted to the study of cooperative behavior, resulting in several models of opinion dynamics. In these models a population of interacting agents holds a set of numerical variables whose states represent opinions about different topics. Inspired by statistical mechanics and social mechanisms, these states evolve governed by mathematical rules that control the dynamics of interaction between agents and the influence of external factors [16]. An overall behavior typically investigated is the condition to the emergence of consensus or polarization among the agents [17]. Models of opinions are commonly divided into two groups. In the first, the variables held by each agent assume a finite set of discrete value. Examples include the Sznajd [18] and Voter [19,20] models, majority rule [21], and social impact theory [22]. On the other side, in the second group, the opinions can vary continuously, and the values can be expressed by real numbers. The most widely known examples are the models of Deffuant [23] and Hegselmann-Krause [24]. Both models assume the *bounded confidence* hypothesis, according to which a pair of agents can influence each other only if their opinions are sufficiently close, specified by a threshold. Useful reviews can be found in [17,25].

In the knowledge acquisition problem studied here, the agents hold not only a single variable, but they also consider a set of continuous states representing their knowledge about the components of the system being modeled. Although agents continuously share their opinions, it is assumed that the estimation of different components of the system proceeds in a completely independent fashion. Furthermore, we consider that a specific agent interacts with all of its neighbors at each time step without regard to the bounded confidence property. In order to enrich our modeling approach, we incorporate the ability of making observations. This important ingredient continuously reinforces the position of the estimations of each agent, while counter-balancing their tendency to become similar after each interaction.

The use of an array of opinions was seminally studied in the model proposed by Axelrod [26] in the context of social dynamics. In this model, single opinions represent possible cultural features and evolve in a coupled way. Also, the Deffuant model introduces the use of a set of opinions limited to binary variables. Extensions were recently proposed in order to address more complex topologies [27] and considering continuous states [28]. In [16], the conditions governing disagreement are investigated by using a model of interacting agents, modeled as Boolean perceptrons and holding multiple issues. Seaver *et al.* [29] studied the influence of conservatism and partisanship in the performance of agents solving coordination tasks.

This paper starts by presenting the general assumptions and specifying and discussing successively more sophisticate levels of modeling. We then focus on the development of a model of a complex system by a team of agents interacting through networks of Erdős-Rényi (ER) and Barabási-Albert (BA) types. During estimation, each agent takes into account not only its observation (subject to a probability of error) but also the average of the models of its neighbors. In order to quantify the performance of an agent, we suggest an individual error based on the difference between the model developed and the system under analysis. Because the influence of an agent is directly related to its degree, the overall error of the system is expressed in terms of the mean of the individual errors weighted by the degree of each respective agent. The obtained results imply a series of interesting and important insights, such as the identification of the substantial role of hubs in affecting the estimation of models by all agents: while a linear relationship is identified between the overall estimation errors and the degree of single agents with different error rates, this relationship becomes superlinear when the degree of each node is considered as the respective fitness. In addition, in networks characterized by presence of communities, intensifying the interactions between agents having higher estimation errors in one of the communities undermines the performance of the agents in the other communities.

II. GENERAL ASSUMPTIONS

The first important assumption is that the complex systems under study be *representable as a discrete structure* (e.g., a graph or complex network) and that information about the several parts of this system can be observed, often with a given probability of error. By *agent*, it is meant any entity capable of making observations or measurements of the system under analysis and storing this information for some period of time. Therefore, any scientist can be naturally represented as an agent. However, automated systems, from measurement stations to more sophisticated reasoning systems, can also be represented as agents. Actually, even less dynamic systems such as books or libraries can be thought of as a sort of passive agents, in the sense that they evolve (new editions and versions) as a consequence of incorporation of the evolution of knowledge.

Each agent is capable of making observations or measurements of the system under investigation. The process of measurement typically involves errors, which can be of several types such as observing a connection where there is none. Such errors can be a direct consequence of the limited accuracy of measurements devices as well as of the priorities and eventual biases (not to mention wishful thinking) of each agent. Several other possibilities, such as the existence of nonobservable portions of the system, can also be considered and incorporated into the model. In addition, the model kept by each agent may undergo degradation as a consequence of noise and/or vanishing memory.

A particularly important element in our approach is the incorporation of different types of interactions between the agents, which can be represented in terms of a graph or network (see Fig. 1). In this case, each agent gives rise to a node, while interactions (e.g., collaborations) among them are represented by links. The single-agent configuration can be immediately thought of as a special case where the graph involves only one node (agent). Several types of interactions are possible including, but being not limited to conversations, attendance to talks or courses (the speaker becomes a temporary hub), article or book, reading, and Internet exchanges (e.g., e-mailing and surfing). Such interactions may also involve errors (e.g., misunderstanding words during a talk), which can be eventually incorporated into our model. It is interesting to observe that the network of interaction between the agents may present dynamic topology, varying with time as a consequence of new scientific partnerships, addition or removal of agents, and so on.

Therefore, the framework considered in our investigation includes the following four basic components: (i) the complex system under analysis S; (ii) one or more agents capable of taking observations or measurements of the system S (subject to error) and of interacting one another; (iii) a network describing the complex system S; and (iv) a network of interaction between the agents. In the following we address, in progressive levels of sophistication, the modeling of knowledge acquisition or model building in terms of complex networks concepts and tools.

III. SINGLE-AGENT MODELING

A very simple situation would be the case where a single agent is allowed to observe, in uniformly random fashion, the elements of the complex system being modeled. Since this system is described by a complex network, an observation lies in detecting the presence or not of links between all pairs of nodes of such a system. The prediction of each edge involves a probability γ of error, i.e., observing a connection where there is none and vice versa. A possible procedure of model building adopted by the agent involves taking the average of all individual observations up to the current time step T, i.e.,

$$\langle x \rangle^T = \frac{1}{T}\sum_{t=1}^{T} x^t, \tag{1}$$

where x is the value of a specific edge (0 if nonexistent and 1 otherwise) and x^t is the observation of x at time step t. Observe that we are considering the observation error to be independent along the whole system under analysis. Let us quantify the error for estimation of any edge, after T time steps as follows:

$$\epsilon(x)^T = |x_S - \langle x \rangle^T|, \tag{2}$$

where x_S is the original value of that edge. It can be easily shown that

$$\lim_{T\to\infty} \epsilon(x)^T = \gamma. \tag{3}$$

Because the observation error is independent among the pairs of nodes, the average of the errors along the whole network is identical to the limit above.

Thus, in this configuration the best situation which can be aimed at by the agent is to reach a representation of the connectivity of the original complex system up to an overall limiting error γ, reached after a considerable period of time. Though the speed of convergence is an interesting additional aspect to be investigated while modeling multiagent knowledge acquisition, we leave this development for a forthcoming investigation.

IV. MULTIPLE-AGENT MODELING

We now turn our attention to a more interesting configuration in which a total of N_a agents interact while making observations of the complex system under analysis, along a sequence of time steps. As before, each agent observes the whole system at each time step, with error probability γ. In case no communication is available between the agents, each one will evolve exactly as discussed in the previous section. However, our main interest in this work is to investigate how *interactions* and *influences* between the multiple agents can affect the quality and speed at which the system of interest is learned. The original system under study is henceforth modeled as a complex network and represented in terms of its respective adjacency matrix A_S. We also assume that the agents exchange information through a complex network of contacts.

One of the simplest and still interesting modeling strategies to be adopted by the agents of such a system involves the following dynamics: at each time step t, each agent observes the connectivity of the complex system with error γ_i, yielding the adjacency matrix O_i^t, and also receives the current matrices K_j^t from each of its immediate neighbors j (i.e., the agents to which it is directly connected). The agent i then calculates the mean matrix $\langle K \rangle_i^t$ of the matrices K_j^t, i.e.,

$$\langle K \rangle_i^t = \frac{1}{k_i}\sum_{j\in\Gamma_i} K_j^t, \tag{4}$$

where Γ_i is the set of the neighbors of the agent i and k_i is its degree. The agent i then makes a weighted average between its current matrix K_i^t and the immediate neighbors mean matrix, i.e.,

$$V_i^t = [a\langle K \rangle_i^t + (1-a)K_i^t], \tag{5}$$

where $0 < a \leq 1$ is a relative weight. We henceforth assume $a=0.5$ so that V_i^t becomes equal to the average between its current matrix and of the neighbors at that time step. Each agent i subsequently adds this value to its current observation, i.e.,

$$K_i^{t+1} = V_i^t + O_i^t, \tag{6}$$

so after T time steps the estimation of the adjacency matrix by the agent i can be given as

$$A_i^T = K_i^T/T. \tag{7}$$

With this formulation, the current matrix K_i^t is a continuous unbounded variable driven by an average process with the neighbors and by accumulating the binary observations. However, the adjacency matrix A_i^t learned by any of the agents has entries with values in the continuous interval [0,1]. Different averaging schemes have been studied in [30] applied to continuous opinion dynamics with bounded confidence.

This simple discrete-type dynamics has some immediate important consequences. By averaging between its current observation and the mean estimation from the neighbors, even the limit error may be actually modified with respect to the single agent situation. It is important to note, also, that the variables were naturally defined as matrices since the agents learn about networked systems, which are commonly represented by the adjacency matrix. Equivalently, this definition can be changed and the mathematical description can be done by using vectors of observations, instead of matrices. The entries of the vectors would represent the estimation of an edge in a network, or in a more general context, any binary information to be observed.

Several variations and elaborations of this model are possible, including the consideration of noise while transmitting the observations, forgetting dynamics, the adoption of other values of a, as well as the evolution in time of the network of agents including different levels of affinity between them. Furthermore, the parameter γ could depend not only on the agent performing an observation but also on the component of the system being modeled, in our case, each element of the adjacency matrix A_S. As a consequence, the probability of error during an estimation, γ, would also be represented as a matrix. Another alternative to increase the level of sophistication of the knowledge acquisition dynamics is the possibility of predicting possible missing links in the studied system. This problem was recently investigated by Clauset *et al.* [31] taking into account the hierarchical structure presented in real networks.

In this paper, however, we focus attention on the multiagent model described in this section with error being present only during the observation by each agent. In the remainder of our work, we report results of numerical simulation considering three particularly important situations: (i) multiple agents with equal observation errors; (ii) as in (i) but with one of the agents having different observation error; and (iii) multiple-agent scheme applied in networks with community structure and one of the communities having higher error rate. Interesting results are obtained for all these three configurations including the analytical solution of the overall behavior of the system.

V. CASE EXAMPLES: RANDOM AND SCALE-FREE NETWORKS OF AGENTS

In this section we investigate further the dynamics of multiagent learning by considering theoretical simulations performed in a fixed collaboration network. We assume that the agents collaborate through a uniformly random model (ER) as well as a scale-free model (BA). For simplicity's sake, we consider only a single realization of such each of these models in our subsequent investigation. Each matrix K_i is initiated with random values uniformly distributed between 0 and 1. The original network to be learnt by the agents is a Barabási-Albert network containing $N=50$ nodes and average degree equal to 6, represented by the respective adjacency matrix A_S. The performance of each agent is quantified in terms of the error between the original network and the models obtained for that agent after a sufficiently large number of time steps (henceforth assumed to be equal to $T=300$ time steps). This error is calculated as

$$\varepsilon_i^T = \frac{1}{N^2}\sum_{m=1}^{N}\sum_{n=1}^{N}|A_i^{T,m,n} - A_S^{m,n}|, \tag{8}$$

where A_i is the adjacency matrix representation of the model of agent i at time step T. The overall performance of the set of agents is henceforth expressed in terms of the average of the above error weighted by the degree considering all agents, which is here called the *overall error* of the system:

$$E^T = \frac{1}{\sum_{i=1}^{N_a} k_i}\sum_{i=1}^{N_a} k_i \varepsilon_i^T. \tag{9}$$

Because the influence of an agent in the network is correlated with the number of its neighbors, such a definition of the overall error takes into account the respective importance of the models (and their intrinsic errors) of each agent.

The Appendix provides the analytical description of the behavior of the overall error at large times, E:

$$E = \lim_{T\to\infty} E^T = \frac{1}{\langle k\rangle N_a}\sum_{i=1}^{N_a} k_i\gamma_i, \tag{10}$$

where $\langle k\rangle N_a = \sum_{i=1}^{Na} k_i$. Thus, it is straightforward to realize that in case of all agents with the same probabilities of observation error, equal to γ, also the overall error tends to γ. In addition, we can apply the result of the overall error to the two following situations, assuming a network of agents with average degree $\langle k\rangle$ and maximum and minimum degree equal to $k_{\max}$ and $k_{\min}$, respectively:

(a) One agent j with degree k_j having a probability of observation error $\gamma_j = p\gamma (0 \le p \le 1/\gamma)$ and all other agents with an error rate γ:

$$E = \frac{\left(\sum_{i=1}^{N_a} k_i\gamma\right) - \gamma k_j + p\gamma k_j}{\langle k\rangle N_a},$$

$$E = \frac{k_j\gamma(p-1)}{\langle k\rangle N_a} + \gamma, \tag{11}$$

which is a linear relation with respect to the degree of the agent with different error rate. Note that there is no difference between any two networks of agents with the same average degree and number of vertices.

(b) One agent j with degree k_j having an observation error proportional to its degree $\gamma_j (\gamma \le \gamma_j \le 1)$ and all other agents with an error rate γ:

$$E = \frac{k_j(1-\gamma)(k_j - k_{\min})}{\langle k\rangle N_a(k_{\max} - k_{\min})} + \gamma, \tag{12}$$

which becomes a quadratic relation with respect to the degree of the agent with different error rate.

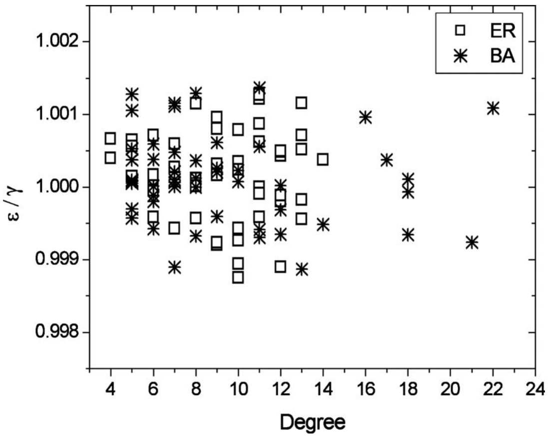

FIG. 2. The individual estimation errors in terms of the degree of the agents obtained for the Erdős-Rényi (ER) and Barabási-Albert (BA) models with N_a=50 and $\langle k\rangle$=9.4. Results obtained by averaging over 100 realizations of the dynamics.

A. Fixed observation errors

In this first configuration, all agents have the same estimation error γ=0.2, and therefore the same influence, over the averages obtained by each agent. However, agents with higher degree, especially hubs, are still expected to influence more strongly the overall dynamics as a consequence of the fact that their estimated models are taken into account by a larger number of neighbors. We assume a single realization of the ER and BA models, both containing 50 agents and average degree equal to 9.4. Figure 2 shows the distribution of the individual errors ε_i^T of each agent obtained for this case in terms of the respective degrees. As could be expected, the errors obtained among the agents are very similar one another. Thus, there is no major difference in the learning quality among the agents. A rather different situation arises in the next sections, where we consider different error rates.

B. Varying observation errors

We now repeat the previous configuration in two situations: (i) one of the agents having half or twice the error rate of the others and (ii) having an error proportional to its degree. Simulations are performed independently while placing the higher error rate at all of the possible agents, while the respective overall errors are calculated. In addition, we compared the behavior for single realizations of ER and BA networks of agents.

Figure 3(a) shows the results obtained while considering error probability equal to 0.2 for one of the agents and γ =0.4 for all other agents. It is possible to identify a substantial difference of the mean quality of the models which depends linearly on the degree of the agent with different error rate. More specifically, it is clear that the overall errors are much smaller for agents with higher degrees. In other words, the best models will be obtained when the hubs have smaller observation errors, influencing strongly the rest of the agents through the diffusion, along time, of their respective estimated models.

Figure 3(b) shows the results obtained when the estimation error of one of the agents is 0.4 while the rest of the agents have error rate γ=0.2. The opposite effect is verified, with the overall error increasing linearly with the degree of the agent with higher estimation error. Moreover, as expected, no difference was found between the ER and BA models in both cases.

Figure 4 shows the results obtained when one of the agents has a higher error rate than the others and proportional to its degree. Similar results to the previous case were found, but with a nonlinear behavior. In addition, accordingly with the analytical predictions, a separation was verified between the ER and BA networks (ER presents higher values than BA). This means that when we consider two agents with the same degree, each in one network, the overall error is lower for the BA model since their hubs influence more strongly to increase the quality of the estimated models.

A better picture of the influence of the degree over the model development by other agents is provided in Fig. 5, respectively, to the situation having twice the error probability (γ_j=0.4) for the agent j with the highest degree. Figures 5(a) and 5(c) consider the ε_i^T of individual agents in terms of their respective degrees for the BA and ER models, respectively, both containing 300 agents and average degree equal

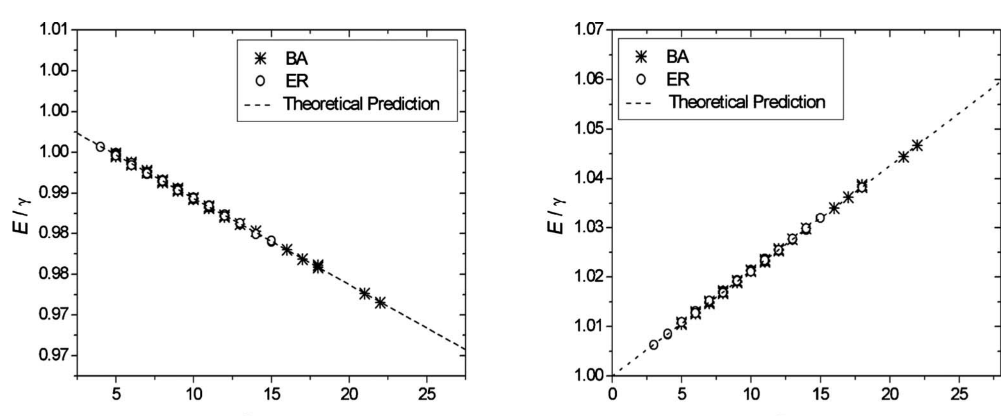

FIG. 3. The overall estimation errors in terms of the degree of the agents with different error rates for ER and BA models with N_a =50 and $\langle k\rangle$=9.4: (a) one agent with an error rate equal to 0.2 and all other agents with γ=0.4 and (b) one agent with 0.4 and all other nodes with γ=0.2. The dashed line represents the analytical solution. Results obtained by averaging over 50 realization of the dynamics.

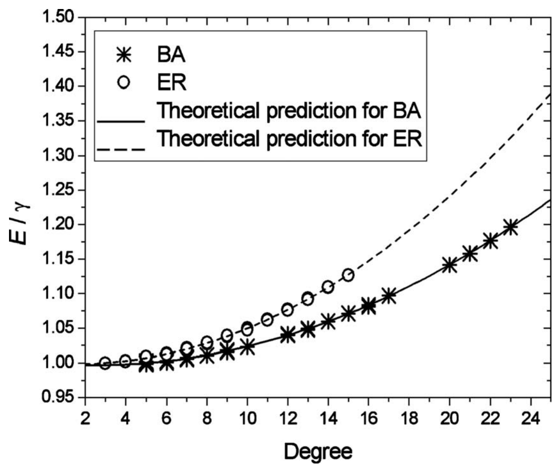

FIG. 4. The overall estimation errors in terms of the degree of the agents with higher error rates: one agent with an error proportional to its degree and all other agents with $\gamma=0.2$. The dashed and straight lines represent the analytical solution for the ER and BA models, respectively. Simulations for a single realization of each model with $N_a=50$ and $\langle k \rangle=9.4$. Results obtained by averaging over 50 realizations of the dynamics.

to 4.7. Dark points represent the neighborhood of the less accurate agent. It is clear from these results that the degree of the differentiated agent affects the estimation of the whole set of agents, especially for its neighborhood, shifting substantially the average individual errors. Observe also that, in both cases, the largest individual error results precisely at the less accurate agent. Furthermore, nodes with high degrees

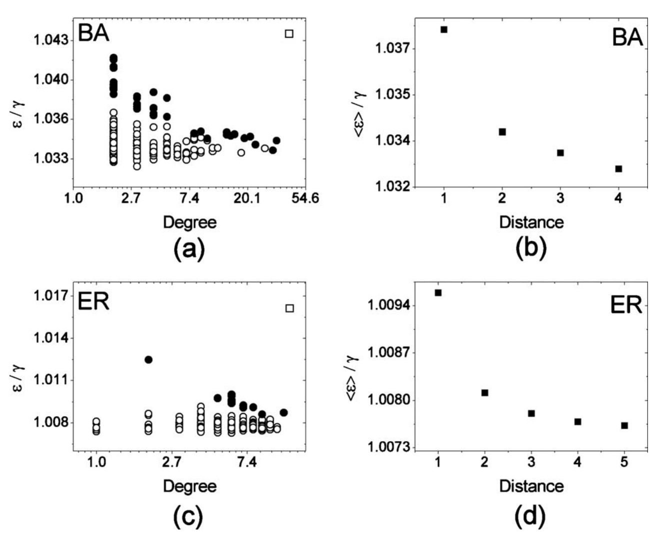

FIG. 5. The individual estimation errors in terms of the degree of the agents in the case of the agent with the highest degree having an error rate equal to 0.4 and all other agents with $\gamma=0.2$ [(a) and (c)]: □, the agent with the highest error rate; ●, the neighborhood of the agent with different error rate. ○, the other agents. Figures (b) and (d) show the mean value of the individual errors of the set of nodes with the same topological distance from the agent with the highest error rate. Simulations for the ER and BA models with $N_a=300$ and $\langle k \rangle=4.7$. Results obtained by averaging over 50 realizations of the dynamics.

tend to be less influenced by the agent with different error rate. Figures 5(b) and 5(d) complement the analysis considering the mean value of the individual errors of agents with the same topological distance from the agent with different error probability. The results show that the closer agents tend to be more affected than the peripherals.

From the results above, we immediately verify that the presence of the less accurate agent produces a nonuniform error distribution. Since the individual error of an agent is produced by averaging between its model and the mean estimation from the neighbors, the direct contact (and proximity) with agents with higher error rates alters considerably its final model. Moreover, this effect is more strongly felt by agents with few connections.

A complementary way to verify the influence of the connectivity on the knowledge acquisition is to consider the temporal evolution of the individual estimation errors in a single realization of the dynamics. Figures 6(a) and 6(b) show this analysis for a BA network when the agent with smallest and highest degree, respectively, has twice the error rate of the other agents. Considering all agents, initially there is a significant difference between the estimated models, which tends to be minimized as the agents interact and information diffuses through the network. A more detailed inspection of the agents with the same error rate (lower panel) reveals that their models are more affected when the less accurate agent has a higher degree.

Figure 7 shows the standard deviation of individual errors in terms of the degrees of the agents with different error rate for the three situations above. Differently from the first two situations [Figs. 7(a) and 7(b)], we detected a positive correlation when one agent has an error rate proportional to its degree [Fig. 7(c)].

C. Varying observation errors in communities

After studying how a single agent can affect the overall performance of the entire network, we can extend our analysis to quantify the influence of a set of nodes when the knowledge acquisition dynamics is applied to networks of agents with community structure. The presence of modules, or clusters, with nodes densely connected compared with the number of the intercommunity edges is an important property found in many real networked systems and is acknowledged to be related with the similarity between the nodes and the hierarchical organization of networks [32].

In our modeling, the synthetic networks of agents with community organization were generated by using the algorithm proposed by Lancichinetti *et al.* [33]. This generator is capable of producing networks with heterogeneous distributions both in the degree of the vertices and in the community size. Furthermore, the fraction of edges that each agent shares with agents in other communities can be also specified by an input parameter, μ, the *mixing parameter*.

During the simulations, all agents of one of the communities had twice the error rate of the others (it was equal to 0.4). We analyzed how the inclusion of more links between the agents with higher observation errors influences the individual errors of the agents in the other communities. The

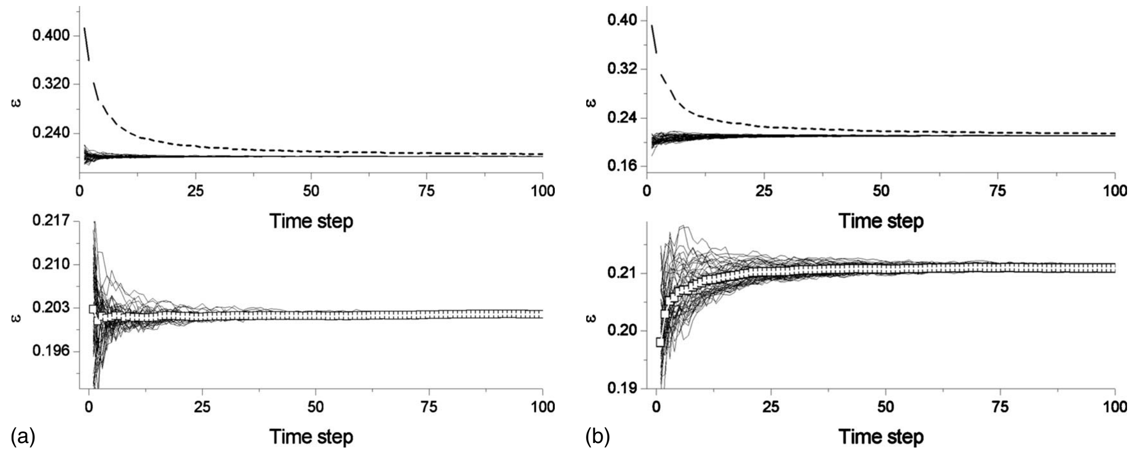

FIG. 6. Temporal evolution of the individual errors considering the agent with (a) smallest and (b) highest degree having an observation error rate equal to 0.4 and all other agents with $\gamma=0.2$. Upper panel: all agents are considered (dashed line represents the less accurate agent). Lower panel: only the agents with the same error rate (dark lines) and the mean value of their individual errors. Simulations for BA model assume $N_a=50$ and $\langle k \rangle=9.4$.

dynamics was performed in networks with $N_a=256$ agents considering variations of the mixing parameter, the degree distribution and the size of the community with higher error rate.

Figure 8(a) shows the results of the dynamics applied to realizations of networks of agents with four communities of same size and all agents with the same degree (equal to 16) but varying the mixing parameter. For each network we considered one of the equal-sized communities having a higher observation error. As shown in this figure, the addition of edges between the agents of this community increases the individual errors of the agents in the other communities and increases with the number of intercommunity edges (the value of μ). Since the dynamics is based on an averaging process, in the sense that the individual models are updated considering the models of the neighborhood, the agents with higher observation errors tend to estimate less accurate models when the intracommunity communication is intensified. Indeed, the addition of new edges implies in more agents with higher errors to influence the development of their estimations. Finally, the intercommunity links is through which the loss of performance is propagated to the other communities.

Similar results are shown in Figs. 8(b) and 8(c). For the former, we considered a single realization of the network of agents in which all of them have the same degree, equal to 16, and the mixing parameter was set equal to 0.2, while the agents were grouped into five communities of different sizes. Independent simulations were performed considering all the agents of each community having a higher error rate. Again, a positive correlation was found between the number of added links between the agents in the community of higher observation error and the mean value of the individual errors of the other agents. As expected, this effect increased with the size of the community with higher error rate.

Figure 8(c) shows the results for agents grouped into 4 communities of same size with μ equal to 0.2. The degrees of the agents exhibited a power law form (with exponent $\lambda=-2$ and $\langle k \rangle=16$) so that each community had a different mean degree. As before, the influence of each community was considered independently, while we measured the mean value of the individual errors of the agents in the accurate communities. Since the communities had the same size and the agents shared, very roughly, the same fraction of edges with agents in other communities, the greater the average degree of a community, the greater the number of intercom-

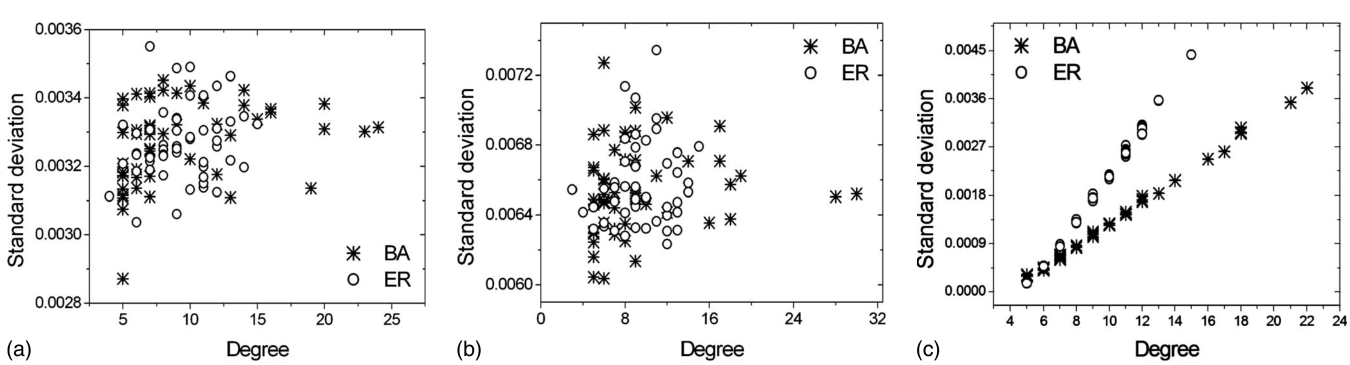

FIG. 7. The standard deviation of individual errors in terms of the degree of the agents with different error rates: (a) one agent with $\gamma=0.2$ and all other agents with 0.4; (b) one agent with $\gamma=0.4$ and all other agents with 0.2; and (c) one agent with an error proportional to its degree and all other agents with 0.2. Results obtained by averaging over 50 realizations of the dynamics.

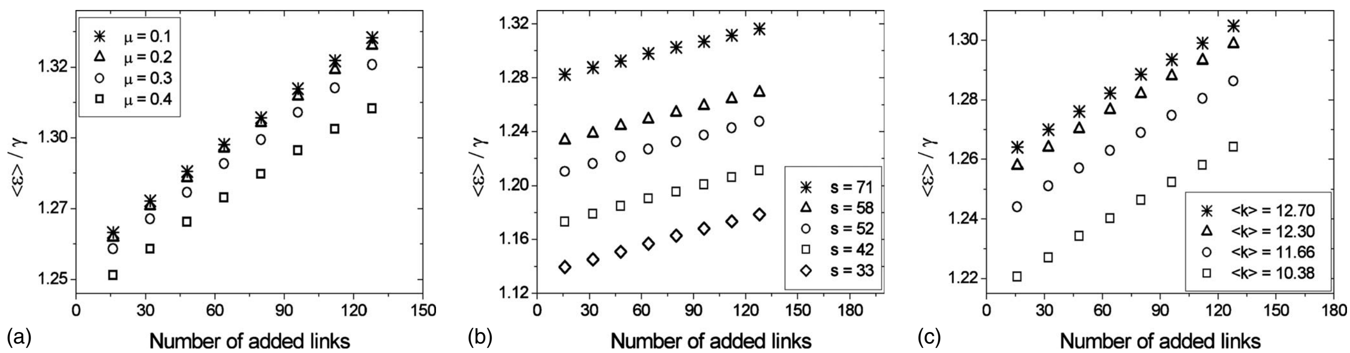

FIG. 8. The dependency between the number of added links in the community of higher observation error and the mean value of the individual errors of the agents in the other communities for networks with N_a=256 agents: (a) all agents with the same degree, equal to 16, grouped into four equal-sized communities and varying the mixing parameter; (b) all agents with the same degree, equal to 16, grouped into five communities with different sizes and setting μ=0.2; (c) agents with degrees distributed as a power-law grouped into four communities of the same size and μ=0.2. Results obtained by averaging over 50 realizations of the dynamics.

munity edges linking to this community. This explains the increase in the individual errors with the average degree of the less accurate community.

VI. CONCLUDING REMARKS

The important problem of scientific interaction has been effectively investigated in terms of concepts and tools of complex networks research (e.g., [8–13]). However, most of the interest has been so far concentrated on characterizing the connectivity between scientists and institutions. The present study reported what is possibly the first approach to model the dynamics of knowledge acquisition (building a model of a complex system) by a system of multiple agents interacting through specific types of complex network. Several configurations were considered at increasing levels of sophistication. Each agent was assumed to make an observation of the system of interest at each time step with error probability γ.

A series of interesting and important results have been identified analytically and through simulations. First, we have that the individual estimation error tends to γ when the agents do not interact one another. However, different individual errors were observed when the agents were allowed to consider the average of models at each of their immediate neighbors.

Special attention was given to the cases in which one of the agents has a different observation error, yielding the important result that the overall error in such a configuration tends to be correlated with the degree of the agent with different observation error. More specifically, we demonstrated that this correlation is linear when one agent has an error half or twice the error rate of the others and nonlinear when it is proportional to its degree. In other words, the connectivity will have a substantial influence over the models developed by the agents, for better or for worse. This behavior still holds when the overall error assumes a simpler form defined as the average of the individual errors, without taking into account the weights of the respective degrees. It is interesting to observe that agents with many connections will imply strong influences over the whole network even in case those agents have no special fitness. Such an effect is a direct consequence of the fact that those agents are heard by more people. In case the hubs have higher observation errors, worse models are obtained throughout the agent's network. In particular, the negative influence of the hubs is more strongly felt by its neighborhood. Finally, we have shown that when the agents are clustered into communities, the addition of edges in the communities with different error rate leads to the change in the individual errors of the agents in the other groups.

This investigation has paved the way to a number of subsequent works, including but not being limited to: consideration of model degradation along time, other learning strategies, other types of networks, observation errors conditional to specific local features (e.g., degree or clustering coefficient) of the network being modeled, as well as other distribution of observation errors among the agents.

ACKNOWLEDGMENTS

L.d.F.C. thanks CNPq (Grant No. 308231/03-1) and FAPESP (Grant No. 05/00587-5) for sponsorship. Part of this work was developed during the author's Visiting Scholarship at St. Catharine's College, University of Cambridge, UK.

APPENDIX: ANALYTICAL DESCRIPTION OF THE OVERALL ERROR

We now present the mathematical formalization of the above concepts. We start with the difference equation that characterizes the dynamics of each agent i:

$$2K_i^{t+1} = \left[\frac{1}{k_i}\sum_{j\in\Gamma_i} K_j^t\right] + K_i^t + 2O_i^t, \tag{A1}$$

where K_i^t is the current matrix and O_i^t is the observation of the agent i at time step t. The term in brackets is the mean of the matrices received from the set Γ_i of the neighbors of i. Now we consider the continuous approximation given by

$$2\frac{d}{dt}K_i(t) = \left[\frac{1}{k_i}\sum_{j\in\Gamma_i} K_j(t)\right] - K_i(t) + 2O_i(t). \tag{A2}$$

Dividing by t and knowing that $\frac{1}{t}\frac{d}{dt}K_i(t) = \frac{d}{dt}\left(\frac{K_i(t)}{t}\right) + \frac{1}{t^2}K_i(t)$, we have

$$2k_i\frac{d}{dt}\left[\frac{K_i(t)}{t}\right] + 2k_i\frac{K_i(t)}{t^2} = \left[\sum_{j\in\Gamma_i}\frac{K_j(t)}{t}\right] - k_i\frac{K_i(t)}{t} + 2k_i\frac{O_i(t)}{t}, \tag{A3}$$

and adding, for every agent i:

$$\frac{d}{dt}\sum_{i=1}^{N_a}\left[\frac{k_iK_i(t)}{t}\right] = \frac{1}{t}\sum_{i=1}^{N_a}k_iO_i(t) - \frac{1}{t}\sum_{i=1}^{N_a}\left[\frac{k_iK_i(t)}{t}\right], \tag{A4}$$

since $\Sigma_i\Sigma_{j\in\Gamma_i}\frac{K_j(t)}{t} = \Sigma_i\frac{k_iK_i(t)}{t}$. Finally, we define $J \equiv \Sigma_i\frac{k_iK_i(t)}{t}$:

$$t\frac{d}{dt}J(t) = \sum_{i=1}^{N_a}k_iO_i(t) - J(t). \tag{A5}$$

So, after time T:

$$J(T) = \sum_{i=1}^{N_a}k_i\left[\frac{1}{T}\int_0^T O_i(t)dt\right]. \tag{A6}$$

For an agent i the probability of observing an edge where there is none and vice versa is γ_i. So, for an entry $A_S^{m,n}$ of the system under observation and for large values of T, the average of the observation is

$$\lim_{T\to\infty}\left[\frac{1}{T}\int_0^T O_i^{m,n}(t)dt\right] = = \begin{cases}(1-\gamma_i) & \text{if } A_S^{m,n}=1\\ (\gamma_i) & \text{if } A_S^{m,n}=0.\end{cases} \tag{A7}$$

Applying this result in Eq. (A5), we have

$$\lim_{T\to\infty} J^{m,n}(T) = \begin{cases}\sum_i k_i(1-\gamma_i) & \text{if } A_S^{m,n}=1\\ \sum_i k_i(\gamma_i) & \text{if } A_S^{m,n}=0.\end{cases} \tag{A8}$$

Subtracting $\Sigma_ik_iA_S^{m,n}$ and adding for all m and n we have

$$\sum_{m=1}^{N}\sum_{n=1}^{N}\left|\lim_{T\to\infty}\left(J^{m,n}(T) - \sum_{i=1}^{N_a}k_iA_S^{m,n}\right)\right| = \sum_{i=1}^{N_a}\left(\sum_{m=1}^{N}\sum_{n=1}^{N}\gamma_ik_i\right),$$

$$\sum_{i=1}^{N_a}k_i\left\{\frac{1}{N^2}\sum_{m=1}^{N}\sum_{n=1}^{N}\left|\lim_{T\to\infty}\left(\frac{K_i^{m,n}(T)}{T} - A_S^{m,n}\right)\right|\right\} = \sum_{i=1}^{N_a}k_i\gamma_i \tag{A9}$$

for both cases. The term in the sum is precisely the individual error of an agent i weighted by its degree. Then, dividing by the sum of the degrees, we have the analytical expression for the overall error of the system (E) for large times;

$$\frac{\sum_{i=1}^{Na}k_i\varepsilon_i}{\sum_i k_i} = \frac{\sum_{i=1}^{Na}k_i\gamma_i}{\sum_i k_i}$$

$$\Rightarrow E = \frac{1}{\langle k\rangle N_a}\sum_{i=1}^{N_a}k_i\gamma_i. \tag{A10}$$

[1] V. Botti, F. Barger, A. Crespo, D. Gallardo, I. Ripoll, E. Onaindia, and L. Hernandez, *Sharing Temporal Knowledge by Multiple Agents* (Springer, New York, 1993).
[2] P. S. Sajja, Interdisc. J. Info., Know., and Manag. **3**, 1 (2008).
[3] R. Albert and A. L. Barabási, Rev. Mod. Phys. **74**, 47 (2002).
[4] S. N. Dorogovtsev and J. F. F. Mendes, Adv. Phys. **51**, 1079 (2002).
[5] M. E. J. Newman, SIAM Rev. **45**, 167 (2003).
[6] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D. Hwang, Phys. Rep. **424**, 175 (2006).
[7] L. da F. Costa, F. A. Rodrigues, G. Travieso, and P. R. Villas Boas, Adv. Phys. **56**, 167 (2007).
[8] D. J. de Solla Price, Science **149**, 510 (1965).
[9] M. E. J. Newman, Proc. Natl. Acad. Sci. U.S.A. **98**, 404 (2001).
[10] M. E. J. Newman, Phys. Rev. E **64**, 016131 (2001).
[11] M. E. J. Newman, Phys. Rev. E **64**, 016132 (2001).
[12] J. Moody, Am. Sociol. Rev. **69**, 213 (2004).
[13] A.-L. Barabási, H. Jeong, Z. Néda, E. Ravasz, A. Schubert, and T. Vicsek, Physica A **311**, 590 (2002).
[14] L. da F. Costa, M. A. Rodrigues, and F. N. Silva, Physica A **387**, 6201 (2008).
[15] L. da F. Costa, Phys. Rev. E **74**, 026103 (2006).
[16] A. C. R. Martins, R. Vicente, and N. Caticha, J. Stat. Mech.: Theory Exp. **2009**, 03015.
[17] C. Castellano, S. Fortunato , and V. Loreto, Rev. Mod. Phys. **81**, 591 (2009).
[18] K. Sznajd-Weron and J. Sznajd, Int. J. Mod. Phys. C **11**, 1157 (2000).
[19] P. Clifford and A. Sudbury, Biometrika **60**, 581 (1973).
[20] R. A. Holley and T. M. Liggett, Ann. Probab. **3**, 643 (1975).
[21] S. Galam, Eur. Phys. J. B **25**, 403 (2002).
[22] B. Latan, Am. Psychol. **36**, 343 (1981).
[23] G. Deffuant, D. Neu, F. Amblard, and G. Weisbuch, Adv. Complex Syst. **3**, 87 (2000).
[24] R. Hegselmann and U. Krause, J. Artif. Soc. Soc. Simul. **5**, 1 (2002).
[25] J. Lorenz, Int. J. Mod. Phys. C **18**, 1819 (2007).
[26] R. Axelrod, J. Conflict Resolut. **41**, 203 (1997).
[27] M. Purvis, A. Alaali, and B. T. R. Savarimuthu, *WI-IAT, 2008* (IEEE Computer Society, Washington, D.C., 2008), pp. 394–397.

[28] J. Lorenz, Eur. J. Econ. Soc. Syst. **19**, 213 (2006).
[29] S. M. D. Seaver, A. A. Moreira, M. Sales-Pardo, R. D. Malmgren, D. Diermeier, and L. A. N. Amaral, Eur. Phys. J. B **67**, 369 (2009).
[30] R. Hegselmann and U. Krause, Comput. Econ. **25**, 381 (2005).
[31] A. Clauset, C. Moore, and M. E. J. Newman, Nature (London) **453**, 98 (2008).
[32] S. Fortunato, Phys. Rep. **486**, 75 (2010).
[33] A. Lancichinetti, S. Fortunato, and F. Radicchi, Phys. Rev. E **78**, 046110 (2008).