Identification of complex network topologies through delayed mutual information

Pierre-Alain Toupance [*,**] **Bastien Chopard** [*]
Laurent Lefevre [**]

[*] *University of Geneva, Geneva, Switzerland*
bastien.chopard@unige.ch
[**] *Univ. Grenoble Alpes, Grenoble INP, LCIS, Valence, France*
{pierre-alain.toupance,laurent.lefevre}@lcis.grenoble-inp.fr

Abstract: The definitions of delayed mutual information and multi-information are recalled. It is shown how the delayed mutual information may be used to reconstruct the interaction topology resulting from some unknown scale-free graph with its associated local dynamics. Delayed mutual information is also used to solve the community detection problem. A probabilistic voter model defined on a scale-free graph is used throughout the paper as an illustrative example.

Keywords: complex systems, information theory, probabilistic models, interaction topology, adjacency matrix identification, community reconstruction, voter model

1. INTRODUCTION

In Conant (1976), the author already pointed out the interest of information theoretic approaches for the analysis of complex dynamical systems. He specifically highlighted their additive complexity which make them attractive for large scale systems. These approaches are based on the computation and analysis of information flows (and losses) between subsystems and components of the complex system. Therefore it seems quite logical to use them for the analysis of control systems topologies, with applications in mind to coarse graining or partitioning (decomposition) problems, for instance. Delayed mutual information was introduced in Schreiber (2000) - with appropriate conditioning of transition probabilities - to distinguish driving and responding elements through the analysis of the correlation between two stochastic signals. Introducing the time delay and detecting asymmetry in the interaction of subsystems allows one to distinguish information that is actually exchanged between two subsystems from shared information due to common history and input signals.

The structural analysis of complex systems dynamics and input-output properties has a long history. Approaches have been developed which make use of interconnection graph or "inference diagram", describing the existing (analytical) relations between a priori given input, state and output variables (Lin (1974); Siljak (2011). Such approaches give structural results on controllability and observability, together with efficient graph algorithms. Some recent results on structural controllability and observability using this approach are presented in Liu et al. (2011, 2013). They require a priori knowledge of the system interconnection topology. When trying to isolate the most influential or measurable nodes in some complex system, this information is often missing or incomplete. Besides, these results on controllability and observability only conclude to some existing causal relation between the considered sets of variables, but not how much the dynamics of a node may be measured or controlled from another node.

Therefore we proposed in Toupance (2019) the use of delayed mutual and multi-informations to analyze the most influential components in a complex system with no a priori knowledge on the interconnection topology. This approach is non-intrusive in the sense that it may be performed by simply sampling the state dynamics, even if the underlying dynamics is unknown. We proved - on the example of the so-called voter model - that the nodes may be ranked according to their influence (the impact on the average opinion of the entire group) by monitoring the time-delayed multi-information and that this ranking closely relates to controllability/observability grammians singular values.

In this paper we investigate how this delayed mutual information approach may be used to reconstruct the interconnection topology (for instance the incidence matrix when the dynamics is defined on a graph). Our goal is to develop an approach which is only based on the sampling of the state dynamics and could be further applied online, for instance with a moving time frame for the computation of the mutual information.

In many practical situations, the topological structure of a complex network is unknown or uncertain and the topological structure of a complex network plays a pivotal role in it dynamic, and control. Therefore, exploring the underlying topology of a complex network is of great importance. This problem has therefore been studied in many papers and some articles propose an information theoretic approach. In neuroscience this approach seems to be preferred. For example, in the papers Becq et al.

(2017) and Amblard and Michel (2013), the authors were interested in constructing the graph of interactions between the areas of the brain applied to encephalogram analysis. They identified the topology of this complex system using the transfer entropy (Schreiber (2000)) and the Granger causality concept (Granger (1988)). Wilmer et al. (2012) used a time-delayed mutual information for detecting nonlinear synchronization in electrophysiological data.

Ideally the approach should be effective when changes occur in the topology, making us able to detect theses changes and reconstruct quickly the new system topology. We will consider again - as an illustration example - a probabilistic voter model where the vote dynamics is defined on a scale-free graph representing somehow the influence between agents in a social network. The quantitative nature of the mutual information and multi-information suggests a new way to detect groups or communities in the complex system, based on the information flux between these communities. We will show how this partitioning algorithm works on the example of the voter model and will suggest that it could be an alternative approach either for coarse graining or partitioning, or for estimation, control or diagnosis purpose (see for instance Ocampo-Martínez et al. (2011)).

The paper is organized as follows : section 2 introduces the voter model which will be used throughout the paper (subsection 2.1) and the metrics from the theory of information that we will use (subsection 2.2), together with a summary of results previously obtained when using these metrics to measure the relative influence of the agents in the voter model. Our main contributions are presented in section 3 where it is shown that time-delayed multi-information and mutual information allow us to reconstruct the interconnection topology of a complex system (subsection 3.1) and partition the graph into communities which are defined from the information flows (subsection 3.2).

2. THE VOTER MODEL

2.1 Description of the model

Simple models that abstracts the process of opinion formation have been proposed by many researchers Castellano et al. (2009); Galam et al. (1998). The version we consider here is an agent-based model defined on a graph of arbitrary topology, whether directed or not.

A binary agent occupies each node of the network. The dynamics is specified by assuming that each agent i looks at every other agent in its neighborhood, and counts the percentage ρ_i of those which are in the state +1 (in case an agent is linked to itself, it obviously belongs to its own neighborhood). A function f is specified such that $0 \leq f(\rho_i) \leq 1$ gives the probability for agent i to be in state +1 at the next iteration. For instance, if f would be chosen as $f(\rho) = \rho$, an agent for which all neighbors are in state +1 will turn into state +1 with certainty. The update is performed synchronously over all n agents.

Formally, the dynamics of the voter model can be express as

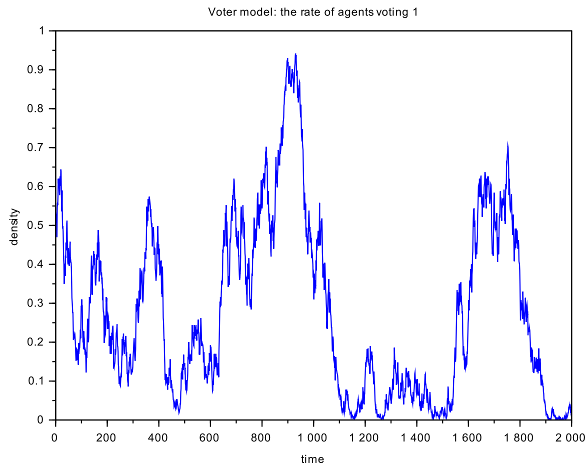

Fig. 1. Time evolution of the density of opinion 1 with noise $\epsilon = 0,001$ and $n = 200$ agents connected through a scale-free network.

$$s_i(t+1) = \begin{cases} 1 & \text{with probability } f(\rho_i(t)) \\ 0 & \text{with probability } 1 - f(\rho_i(t)) \end{cases} \quad (1)$$

where $s_i(t) \in \{0, 1\}$ is the state of agent i at iteration t, and

$$\rho_i(t) = \frac{1}{|N_i|} \sum_{j \in N_i} s_j(t). \quad (2)$$

The set N_i is the set of agents j that are neighbors of agent i, as specified by the network topology.

The global density of all n agents with opinion 1 is obviously obtained as

$$\rho(t) = \frac{1}{n} \sum_{i=1}^{n} s_i(t) \quad (3)$$

In what follows, we will use a particular function f,

$$f(\rho) = (1-\epsilon)\rho + \epsilon(1-\rho) = (1-2\epsilon)\rho + \epsilon \quad (4)$$

The quantity $0 \leq \epsilon \leq 1/2$ is called the noise. It reflects the probability to take a decision different from that of the neighborhood.

To illustrate the behavior of this model, we consider a random scale-free graph G, as simple instance of a social network (Barabási et al. (2000)). We use the algorithm of Béla Bollobás (Bollobás and Riordan (2003)) to generate this graph.

Figure 1 shows the corresponding density of agents with opinion 1, as a function of time. We can see that there is a lot of fluctuations due to the fact that states "all 0's" or "all 1's" are no longer absorbing states when $\epsilon \neq 0$.

2.2 Delayed Mutual and multi-information

Let us consider a set of random variables $X_i(t)$ associated with each agent i, taking their values in a set A. For instance, $X_i(t) = s_i(t)$ would be the opinion of agent i at iteration t.

Since we want to assess the temporal causality, we measure the influence of the vote of agent i at time t on the vote of agent j at time $t+\tau$, we define the τ-delayed mutual information $w_{i,j}$ as

$$\omega_{i,j}(t,\tau) = I(X_i(t), X_j(t+\tau)) \quad (5)$$
$$= \sum_{(x,y)\in A^2} p_{xy} \log\left(\frac{p_{xy}}{p_x p_y}\right) \quad (6)$$

with $p_{xy} = \mathbb{P}(X_i(t) = x, X_j(t+\tau) = y)$ and $p_x = \mathbb{P}(X_i(t) = x)$ and $p_y = \mathbb{P}(X_j(t+\tau) = y)$

We also define the τ-delayed multi-information w_i as a measure of the influence of agent i on all the others

$$\omega_i(t,\tau) = I(X_i(t), Y_i(t+\tau)) \quad (7)$$
$$Y_i(t+\tau) = \sum_{k\neq i} X_k(t+\tau) \quad (8)$$

These information metrics can be computed by sampling. In the sequel we will consider $N = 10^5$ instances of the system in order to perform an ensemble average. At the begining of each simulation, the probabilities at time zeros are reset, and the random seed is changed. For example to compute an approximation of the joint probability of $(X_i(t), X_j(t+\tau))$, we count among these N simulations the number of couple $(k,l) \in \{0,1\}^2$ that we obtained.

According to the central limit theorem, we know that, with this number of instances, we obtain a precision of 3×10^{-2} with a risk of 5% for the approximate values of the probabilities that we compute.

The τ-delayed multi-information can be used as a measure of the influence of opinion of each node i on the vote of the other agents. For instance, Fig. 2 shows $\omega_i(\tau = 2)$ in a steady state, where the origin of time is arbitrary. We observe that some agents i exhibit a more pronounced peak of multi-information towards the rest of the system, suggesting that the opinion of these agents may affect the global opinion of all agents. Note that this results is obtained only by probing the systems, without modifying any of its components. For this reason, we describe this approach as "non-intrusive".

2.3 Controllability and information theoretical

In the paper "Controllability of the Voter Model : an information theoretic approach" (Toupance (2019)), we showed that the delayed multi-information is indeed a good metrics to identify the influential agents of the system. To illustrate this result we compared $w_i(\tau)$ with the impact of forcing the vote of agent i to 1 at all time. As a result of this forcing, the density of vote

$$\rho(t) = \frac{1}{n}\sum_{j=1}^{n} s_j(t) \quad (9)$$

oscillates around a value different from that observed when agent i is free. This variation of the average value of ρ is defined as the *influence* of agent i on the whole system. This measure of influence is "intrusive" as it is the result of an action on the system.

The two metrics (intrusive and non-intrusive) are shown in Fig. 3, with a color representations, for the given scale-free graph. It shows that the multi-information (right panel) detects correctly the influence of the nodes (left panel) since the variation of gray levels are similar in both cases. In this way we can then identify, by non-intrusive observations, which agents are those whose control will be the most influential to the system when their vote is forced.

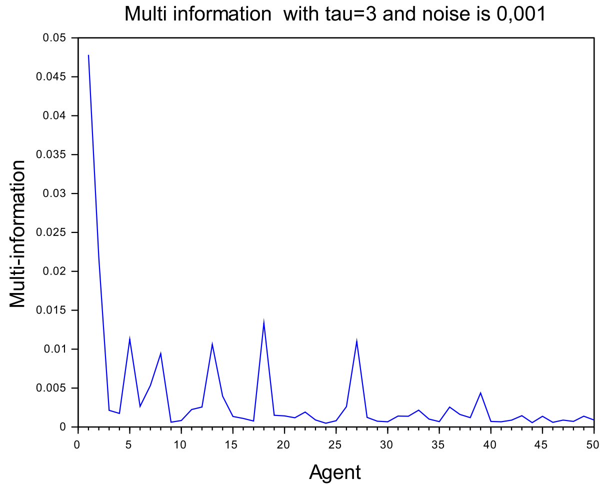

Fig. 2. τ-delayed multi-information $w_i(\tau)$ as a function of i, for a test graph G, similar but not identical to the graph shown in Fig. 3. The case show here has $n = 50$ agents and a noise level $\epsilon = 0.001$ and τ-delayed multi-information is computed at initial state.

The strong correlation that exists between influence and the delayed multi-information can be proven rigorously in the case of a 1D unidirectional voting model. In Toupance (2019) it is shown that the probability π_i that agent i votes 1 when the system is stationary, is

$$\pi_i = \frac{1}{2} + \frac{1}{2}\exp\left[-\frac{i}{\ell_c}\right] \quad (10)$$

where ℓ_c is defined as

$$\ell_c = \frac{1}{\ln\left(\frac{1+2\epsilon}{1-2\epsilon}\right)} \quad (11)$$

We have also shown that delayed mutual information decreases exponentially with the distance between two agents and the noise. By simulation, we obtained that the multi-information between agents i and j (where $j > i$) with a delay $\tau = j - i$ is

$$\omega_{i,j}(j-i) = \alpha_i \exp\left(-\lambda_i(j-i)\right)$$

where λ_i depends on the noise level, ϵ. Figure 2.3 shows that λ_i is proportional to $1/\ell_c$, confirming the strong link between our concept of influence and that of time-delayed multi-information.

3. TOPOLOGY OF THE SYSTEM

3.1 1-delayed mutual information and adjacency matrix

After the results presented in the previous section about the link between controllability and information theory, we are interested here in the identification of topology of a system. The aim is to us our information metrics to construct the interaction topology of the unknown graph underlying the dynamics of a complex system.

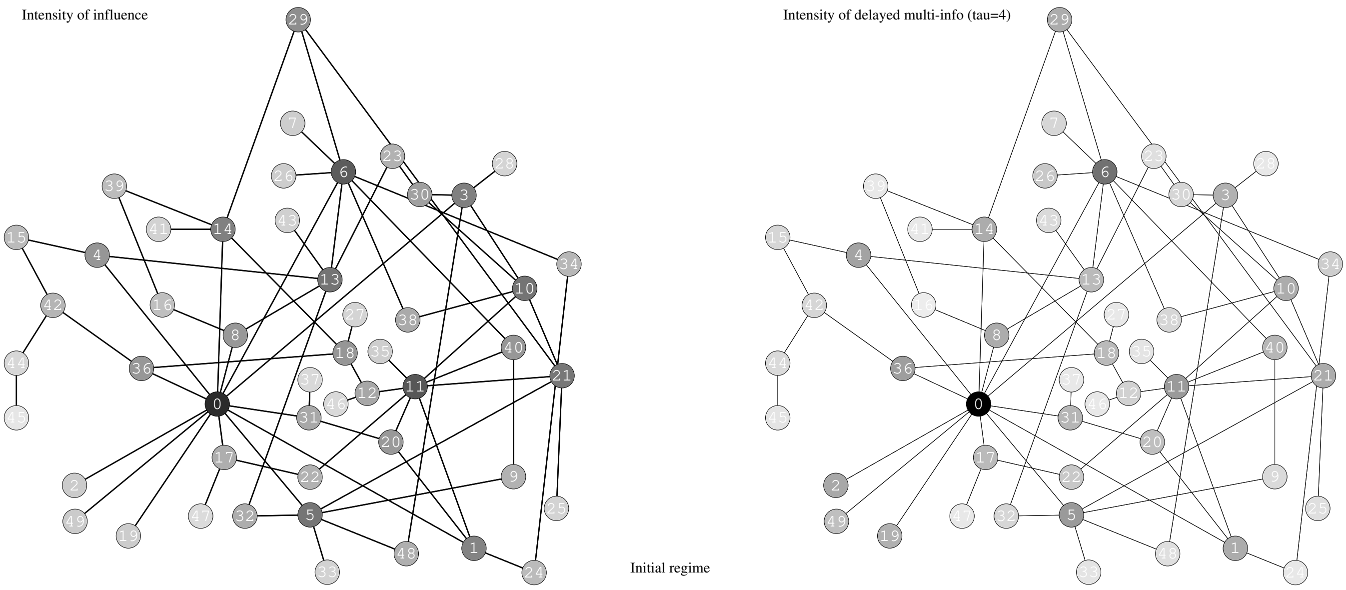

Fig. 3. Scale free graph colored as a function of the values of the influence (left) and the τ-delayed multi-information (right), for $\tau = 4$. The value of τ is chosen so as to match the diameter of the graph. In this case, the multi-information is computed in the transient regime that follows the initial state.

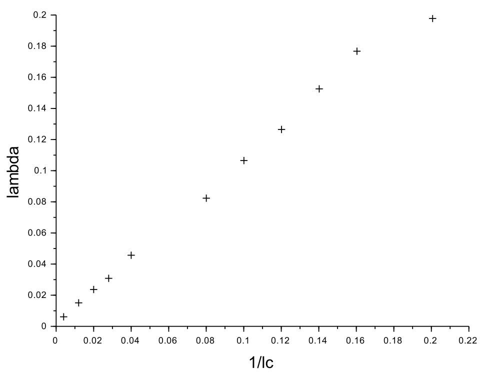

Figure 4 shows the values of 1-delayed mutual information $\omega_{i,j}(1)$ between agent i and all the others in the system. These values were calculated by sampling, when the system has reached its steady state. In this case, the highest values of $\omega_{i,j}(1)$ are obtained for the neighbors of agent i. For $i = 42$ we observed a peak for agents 15, 36, 42, 44 (the agent is neighbor of itself).

Thus, we can use the 1-Delayed Mutual information to get the edges of a graph. For each agent i, we fixed a threshold T_i for $\omega_{i,j}(1)$ that indicates that j is a neighbor of i. This threshold is defined as

$$T_i = \mu_i + a_i\sigma_i$$

where μ_i is the mean of the values of the 1-delayed mutual information between agent i and the others agents, and σ_i is the standard deviation of these values.

There are two possible values for a_i according to the following criteria

$$a_i = \begin{cases} 0.2 & \text{if } w_i(t,\tau) > \alpha + \dfrac{1}{2}\beta \\ 0.7 & \textit{otherwise} \end{cases} \tag{12}$$

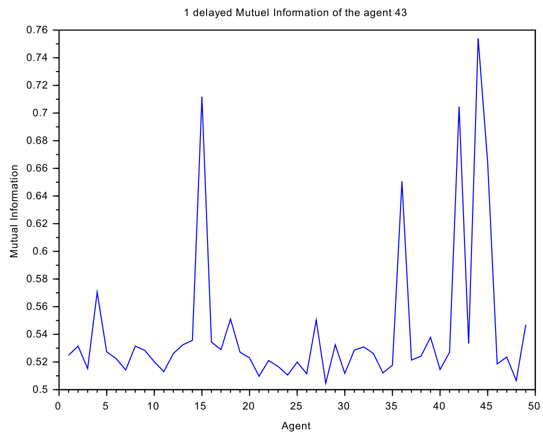

Fig. 4. 1 Delayed mutual information between agent $i = 42$ and the rest of the system. Peaks are visible for the neighbors of i.

where $w_i(t,\tau)$ is agent i's $\tau-$delayed multi-information at time t, as defined in eq. (7). The value of t is chosen to be in the initial regime and τ is taken large enough to capture the influence over the rest of the system. The values α and β are respectively the average and the standard deviation of $w_i(t,\tau)$ over i. In the first case ($a_i = 0.2$), the agent i is considered as very influential.
The estimation of the adjacency matrix, denoted $M = (m_{i,j})_{1\leqslant i,j\leqslant N}$, is defined by :

$$m_{i,j} = \begin{cases} 1 & \text{if } \omega_{i,j}(1) > T_i \text{ or } \omega_{j,i}(1) > T_j \\ 0 & \text{otherwise} \end{cases}$$

In other words, when $\omega_{i,j}(1) > T_i$ or $\omega_{j,i}(1) > T_j$, it is assumed that agents i and j are neighbors.

The values of a_i were determined by an empirical method. We have tested all possible values of a from 0 to 1 with

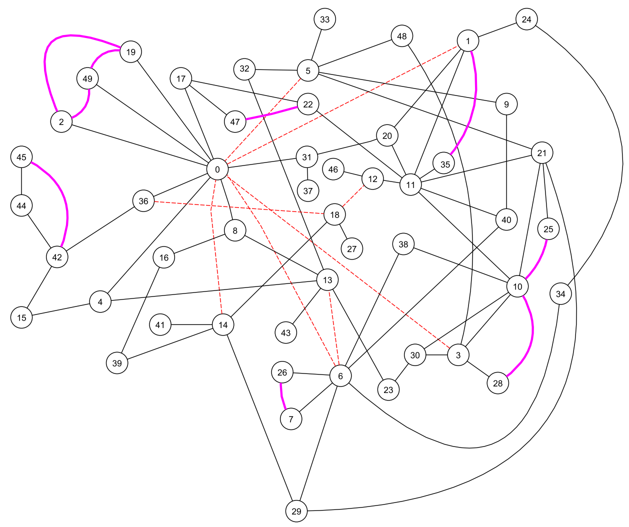

Fig. 5. Graph built with 1-delayed mutual information calculated when the system is in a steady state. The dashed red edges are the ones that have not been found, and the solid pink edges are those that were wrongly added.

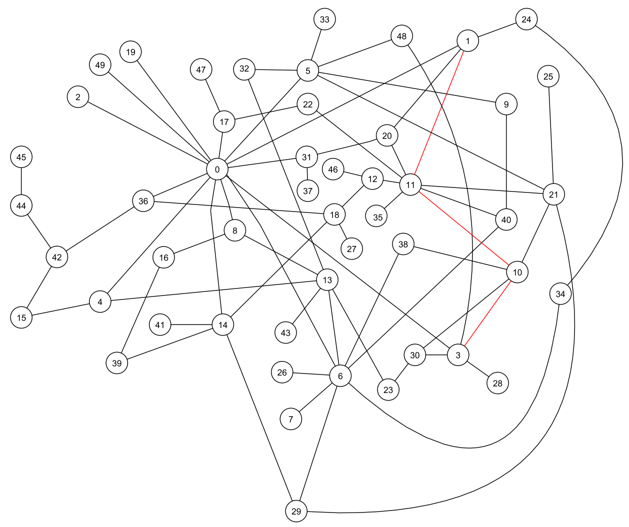

Fig. 6. Graph built with the 1-delayed mutual information computed when the system evolves from its initial state. The red edges are the ones that have not been found.

a step of 0.1 over several scale free graphs, and we have chosen the values in order to minimize the error rate r between this matrix and the actual adjacency matrix A.

The error rate is defined by $r = \frac{\Delta(M,A)}{n^2}$, where $\Delta(M, A)$ is the Hamming distance, namely the number of values that differ between M and A. n is the number of agents in the graph.

For example, Fig. 5, shows the graph that is reconstructed by this procedure, and compares it to the original graph. In this case, the error rate is $r = 1.3\%$. An even better result is obtained when 1-delayed mutual information is computed during the initial transient regime (see Fig. 6). The error rate is now $r = 0.24\%$. In the transient regime the results are probably better because it really probes the direct influences. In order to provoke such a transient regime, one may disrupt the system by temporarily increasing the noise, while calculating the mutual information. This method has been tested by randomly generating 20 scale free graphs. The average error rate obtained is 0.9% and the standard deviation is 0.0026.

3.2 Community detection

The existence and structures of communities in a graph is an important concept in the analysis of social networks. Communities are sub-graph with dense internal connections and sparse connections between these sub-graphs. In our case, partitioning the agents in communities should allow us to better control the network. In (Papadopoulos et al. (2012)) the author presents different method to determine communities. In our case, we will rather use the delayed mutual information. In a scale free graph, we expect a seed-centric approaches for community detection (see Kanawati (2014)). In the voter model these seeds are the most influential nodes, each of them characterizing a different community. In the first part of this paper, we showed that the delayed multi-information can determine these seeds : if we want to partition into k communities, to find the seeds, we determine the k largest values of multi-information.

Then, to build the corresponding community, we define the proximity $p(i, j)$ between two agents i and j as

$$p(i,j) = \frac{1}{r}\sum_{\tau=1}^{r} \omega_{i,j}(\tau) \qquad (13)$$

where r is an approximation of radius of the graph. Bollobás and Riordan (2004) proved that the diameter of a scale-free random graph is asymptotically $\frac{\log(n)}{\log(\log(n))}$ where n is the size of the graph. We can choose

$$r = \lfloor \frac{1}{2}\frac{\log(n)}{\log(\log(n))} \rfloor$$

this value is approximately the radius of the graph.

Each agent j is associated with the community of the seed agent i, where i is obtained as

$$i = \arg\max_{i \in I} p(i,j) \qquad (14)$$

where I is the set of seed agents.

The communities obtained with this algorithm are shown in Figs. 7.

To evaluate the quality of the partitioning, we compute their modularity Q (see Newman (2006)) defined by :

$$Q = \frac{1}{2m}\sum_{i,j}\left(A_{i,j} - \frac{q_i q_j}{2m}\right)\delta(c_i, c_j) \qquad (15)$$

where m is the number of edges of the graph, $A_{i,j}$ is the coefficient (i, j) of the adjacency matrix, q_i is the degree of agent i, c_i is the community of agent i, and δ is the Kronecker symbol, we have

$$\delta(c_i, c_j) = \begin{cases} 1 & \text{if } i \text{ and } j \text{ are in the same community} \\ 0 & \text{otherwise} \end{cases}$$

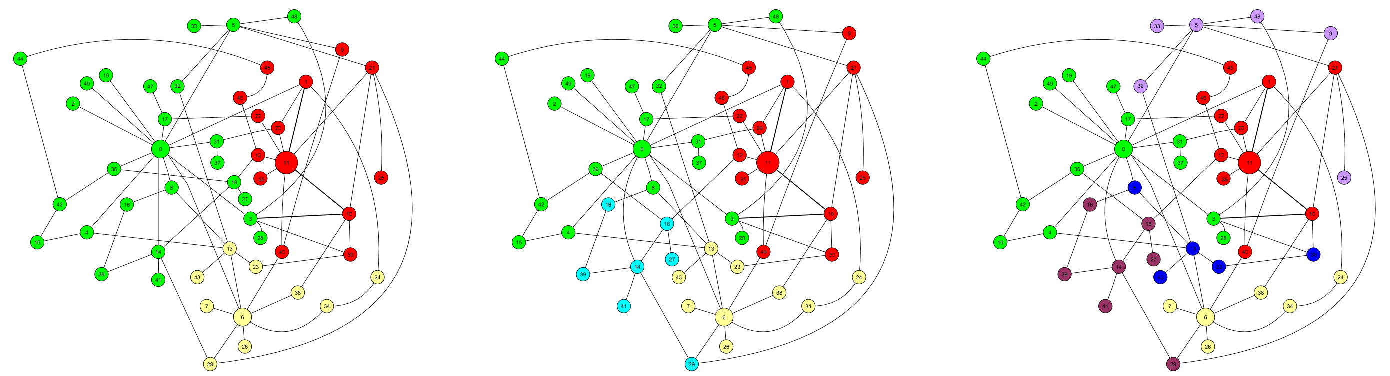

Fig. 7. Partition of the graph in 3, 4 and 6 communities.

Table 1. Modularity for different partitioning.

Number of communities	3	4	6
Q	0.3607	0.4208	0.4382

For our example, the coefficients of modularity are shown in table 3.2. This indicates a good assortativity of the proposed communities.

4. CONCLUSION

In this paper we showed that information theory is related to control theory and further that is offers efficient tools to determine the unknown interaction topology of a complex dynamical system. The results are obtained using a sampling approach on a voter model defined on a scale free graph. The key information-theoretic quantities introduced in this work is the τ-delayed mutual information and multi-information. In addition to reconstructing the graph topology, it can be used to determine communities that partition the graph.

In a future work we plan to extend the present results in three directions: (1) to detect possible changes in the graph topology over time. This will be obtained by computing the information metrics over sliding time windows; (2) To used the detected communities to apply different control strategies; (3) to use community as a way to reduce the complexity of the full system (model reduction).

REFERENCES

Amblard, P.O. and Michel, O.J. (2013). The relation between granger causality and directed information theory: A review. *Entropy*, 15(1), 113–143.

Barabási, A.L., Albert, R., and Jeong, H. (2000). Scale-free characteristics of random networks: the topology of the world-wide web. *Physica A: statistical mechanics and its applications*, 281(1-4), 69–77.

Becq, G., Amblard, P.O., Salin, P., and Comte, J.C. (2017). Détection dynamique de graphes appliquée à l'analyse de signaux électroencéphalographiques issus du sommeil de rat.

Bollobás, B. and Riordan, O. (2004). The diameter of a scale-free random graph. *Combinatorica*, 24(1), 5–34.

Bollobás, B. and Riordan, O.M. (2003). Mathematical results on scale-free random graphs. *Handbook of graphs and networks: from the genome to the internet*, 1–34.

Castellano, C., Fortunato, S., and Loreto, V. (2009). Statistical physics of social dynamics. *Rev. Mod. Phys.*, 81, 591–646.

Conant, R.C. (1976). Laws of information which govern systems. *IEEE transactions on systems, man, and cybernetics*, (4), 240–255.

Galam, S., Chopard, B., Masselot, A., and Droz, M. (1998). Competing species dynamics: Qualitative advantage versus geography. *Eur. Phys. J. B*, 4, 529–531.

Granger, C.W. (1988). Some recent development in a concept of causality. *Journal of econometrics*, 39(1-2), 199–211.

Kanawati, R. (2014). Seed-centric approaches for community detection in complex networks. In *International Conference on Social Computing and Social Media*, 197–208. Springer.

Lin, C.T. (1974). Structural controllability. *Automatic Control, IEEE Transactions on*, 19(3), 201–208.

Liu, Y.Y., Slotine, J.J., and Barabasi, A.L. (2011). Controllability of complex networks. *Nature*, 473, 167.

Liu, Y.Y., Slotine, J.J., and Barabási, A.L. (2013). Observability of complex systems. *Proceedings of the National Academy of Sciences*, 110(7), 2460–2465.

Newman, M.E. (2006). Modularity and community structure in networks. *Proceedings of the national academy of sciences*, 103(23), 8577–8582.

Ocampo-Martínez, C., Bovo, S., and Puig, V. (2011). Partitioning approach oriented to the decentralised predictive control of large-scale systems. *Journal of Process Control*, 21(5), 775–786.

Papadopoulos, S., Kompatsiaris, Y., Vakali, A., and Spyridonos, P. (2012). Community detection in social media. *Data Mining and Knowledge Discovery*, 24(3), 515–554.

Schreiber, T. (2000). Measuring information transfer. *Physical review letters*, 85(2), 461.

Siljak, D.D. (2011). *Decentralized control of complex systems*. Courier Corporation.

Toupance, P-A L Lefèvre, a.B.C. (2019). Controllability of the voter model: an information theoretic approach. *J. Comp. Science, accepted; arXiv:1911.06540*.

Wilmer, A., de Lussanet, M., and Lappe, M. (2012). Time-delayed mutual information of the phase as a measure of functional connectivity. *PloS one*, 7(9).