Detectability thresholds in networks with dynamic link and community structure

Paolo Barucca,[1,2] Fabrizio Lillo,[3] Piero Mazzarisi,[3] and Daniele Tantari[4]

[1]*University of Zurich, Zurich, Switzerland*
[2]*London Institute for Mathematical Sciences, London, UK*
[3]*Scuola Normale Superiore, Pisa, Italy.*
[4]*Centro Ennio De Giorgi, Scuola Normale Superiore, Pisa, Italy.*
(Dated: January 23, 2017)

We study the inference of a model of temporal networks in which both communities and links keep memory of previous network state. By considering maximum likelihood inference from single snapshot observation of the network, we show that link persistence decreases the detectability threshold, preventing the inference of communities even when they are in principle strong enough, while community persistence tends to restore this possibility. Then we show that the inferred communities share a maximum overlap with those of a specific previous instant of time, corresponding to the maximum of a time-lagged assortativity parameter, and therefore they can be closer to those of the past than of the present. We analytically characterize these transitions as a function of the memory parameters of the model.

Natural and artificial systems change over time, both in their individual elements and in the way they interact. When represented as a temporal network [1], a dynamic system of interacting elements has to include time dependence of links and of community structure. These are the simplest, yet important, mechanisms modeling the tendency of elements (nodes) to remain in the same community and of pair of nodes to keep pre-existent relations (links). Temporal networks can be regarded as a sequence of snapshot networks. As we clarify below, when time correlations are present, the information obtained with inference on individual snapshots might be contaminated by the past history of the system. This is analogous to what happens in multilayer networks [2], for which the analysis cannot be decomposed into the separate analysis over each layer if they are correlated.

In this paper we study the problem of the inference of community structure from single snapshot observation of a temporal network with link and community persistence. Community detection is a long-standing, as well as ill-defined, problem in graph partitioning that has been thoroughly studied in the static network case with various approaches: modularity maximization [3], spectral methods [4, 5], belief-propagation [6], and other heuristic algorithms [7]. For analytical tractability, we focus on stochastic block models (SBMs) with dynamic community structure and link persistence. SBMs have been shown to display a detectability transition [6, 8] in the modular assortativity strength, the ratio between the average degree within a block of nodes and the average degree towards different blocks. Recently the problem was investigated in temporal networks [9–12] and in a specific case of Markovian community structures [13]. In this dynamic network model, it was shown that persistence in communities can help detection, by decreasing the detectability threshold, a weaker assortativity is required to infer communities with respect to the static case. On the contrary, we show that, for static algorithms on individual snapshots, persistence in relations, i.e. links, can hinder detection, eventually causing the detection of old communities instead of the ones present at the time the detection is performed.

The model. We consider a Dynamic Stochastic Block Model (DSBM) with link persistence, i.e. at each time step the presence of a link between two nodes is copied from the previous time with probability ξ, while with probability $1 - \xi$ the link is generated according to a SBM where the community structure changes over time Several models of DSBM were previously introduced for community detection in dynamic networks [13–16]. Our variant mixes together link and community persistence. We remind that a SBM is a classical generative model for static networks with community structure. A network $(V, \boldsymbol{A})$ with $|V| = N$ nodes and adjacency matrix $\boldsymbol{A}$ is generated by a SBM as follow. According to a prior $\{q_r\}_{r=1}^{k}$ over k possible choices, each node $i \in V$ is assigned to a community g_i with probability q_{g_i}. Edges are then generated according to a $k \times k$ affinity matrix $\boldsymbol{p}$ and the community structure $\boldsymbol{g}$: each couple of nodes $i, j \in V$ are linked independently with probability $p_{g_i g_j}$.

In the DSBM the community structure changes over time. It consists of a sequence of networks $(V, \boldsymbol{A}^t)_{t=1}^{T}$, each with its own community structure $\boldsymbol{g}^t$. As in [13], the dynamic of each node's assignment g_i^t is an independent Markov process with transition probability $P(g_i^t | g_i^{t-1}) = \eta\, \delta_{g_i^t, g_i^{(t-1)}} + (1 - \eta)q_{g_i^t}$, meaning that with probability η a node remains in the same community, otherwise it changes randomly to a group r with probability q_r. Since at $t = 0$ labels are assigned according to the prior, it is

$$P(\boldsymbol{g}) = \prod_{i=1}^{N} \left[\prod_{t=1}^{T} \eta\, \delta_{g_i^t, g_i^{(t-1)}} + (1 - \eta)q_{g_i^t} \right] q_{g_i^0} \quad (1)$$

Adding link persistence to the DSBM we obtain

$$P(\boldsymbol{A} \mid \boldsymbol{g}) = \prod_{(ij)}^{N} p_{g_i^0 g_j^0}^{A_{ij}^0} (1 - p_{g_i^0 g_j^0})^{1-A_{ij}^0} \times \qquad (2)$$

$$\times \prod_{t=1}^{T} \xi \, \delta_{A_{ij}^t, A_{ij}^{(t-1)}} + (1-\xi) p_{g_i^t g_j^t}^{A_{ij}^t} (1 - p_{g_i^t g_j^t})^{1-A_{ij}^t}.$$

Thus the two parameters η and ξ can be interpreted as, respectively, the persistence of communities and the persistence of links. Community persistence models the tendency of nodes to remain in the same group over time. Link persistence models the preference of nodes in keeping pre-existent relations over time, for example because of the cost of adding or removing links in socio-economic networks [17].

Here we focus on the common choices of a uniform prior, i.e. $q_r = 1/k \; \forall r = 1, \ldots, k$, and affinity matrix with a constant p_{in} on the diagonal and another constant $p_{out} \leq p_{in}$ off diagonal, the so called assortative *planted partition model*. Moreover we measures the level of assortativity with a parameter $a \in [0,1]$ such that

$$\boldsymbol{p} = a \, k\bar{p}\mathbb{I} + (1-a)\bar{p}\mathbf{1} \qquad (3)$$

interpolating between a fully assortative $k\bar{p}\mathbb{I}$ (proportional to the unity matrix) and a fully random $\bar{p}\mathbf{1}$ (proportional to a matrix of ones) affinity matrix, with fixed mean degree $N/k^2 \sum_{ab} p_{ab} = N\bar{p}$. We are interested in the sparse regime $\bar{p} = \bar{c}/N$, that is the most challenging from the inference perspective, since most of real networks of interest are sparse and because sparsity allows to carry out asymptotically optimal analysis.

The fundamental issue is to study under which conditions we can detect, better than chance, the correct labeling of the latent communities $\boldsymbol{g}$ from the observation of $\boldsymbol{A}$. For the static SBM, it was shown (and proved at least for $k = 2$ [18]) that there exists a sharp threshold below which no algorithm can perform better than chance in recovering the planted community structure. This threshold occurs, in terms of the parametrization (3) at $a = a^c := \bar{c}^{-1/2}$ meaning that, at low assortativity, a community structure may still exists but is undetectable. The Bayesian inference approach consists to consider the posterior distribution over the model parameters

$$P(a|\boldsymbol{A}) = \frac{P(a)}{P(\boldsymbol{A})} \sum_{\boldsymbol{\gamma}} P(\boldsymbol{A}, \boldsymbol{\gamma}|a) \qquad (4)$$

to learn the most likely assortativity $\hat{a}$ given the data, and the posterior distribution of the latent assignments

$$P(\boldsymbol{g}|\boldsymbol{A}, a) = \frac{P(\boldsymbol{A}, \boldsymbol{g}|a)}{\sum_{\boldsymbol{\gamma}} P(\boldsymbol{A}, \boldsymbol{\gamma}|a)} \qquad (5)$$

for inferring a set of statistically significant communities $\hat{\boldsymbol{g}}$. Since maximizing the likelihood (4) needs computing expectations w.r.t the posterior (5), it is called
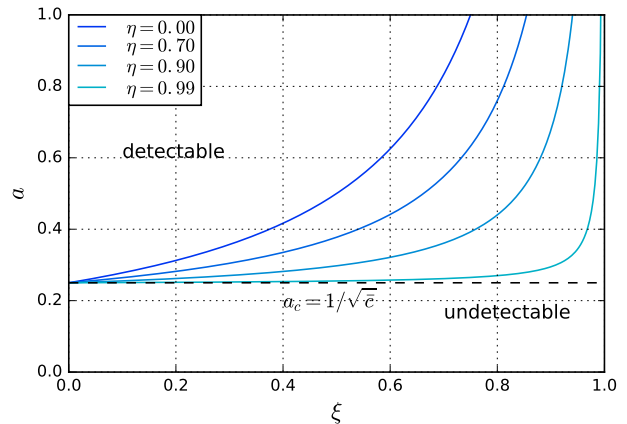

FIG. 1: Asymptotic phase space for single snapshot detectability as function of assortativity a and community/ link persistences, η and ξ, compared to the static threshold a_c.

Expectation-Maximization (EM) procedure [19]. The criticality of this approach stays in the summation over all possible assignments whose number grows exponentially with N. Overtaking this problem is usually done by Monte Carlo (MC) sampling [20] or recently by using belief propagation (BP) algorithms [6, 21]. In this paper we used both methods obtaining similar results. Both provide the optimal $\hat{a}$ together with an estimate of the posterior (5). From that, a partition is obtained by marginalisation, which assigns each node to its most likely group

$$\hat{g}_i = \operatorname{argmax}_r P(g_i = r|\boldsymbol{A}, \hat{a}).$$

This is known [22] to be an optimal estimator, maximising the overlap with the planted assignment

$$q(\boldsymbol{g}, \hat{\boldsymbol{g}}) = \frac{N^{-1} \sum_{i=1}^{N} \delta_{\hat{g}_i g_i} - \max_r q_r}{1 - \max_r q_r}, \qquad (6)$$

where the normalization is chosen to ensure $q = 0$ if labels are assigned by chance.

Single snapshot inference. The inference for the full dynamical model is very complicated due to the double link and community persistence. Here we ask which community structure is inferred from a single snapshot of the dynamic network at a time t. This might occur, for example, if one is unaware that $\boldsymbol{A}^t$ is the result of a dynamic process. Thus we need to compute the posterior $P(\boldsymbol{g}^t|\boldsymbol{A}^t)$ giving the probability of community structure when only the information on the network at time t is used. We prove the following:

Proposition 1. *Given the model above, the posterior $P(\boldsymbol{g}^t|\boldsymbol{A}^t)$ is that of a static SBM with an effective assortativity*

$$a_{\xi,\eta}^t = a \, \epsilon_{\xi,\eta}^t = a \left(\frac{1-\xi}{1-\xi\eta^2} + (\xi\eta^2)^t \frac{\xi(1-\eta^2)}{1-\xi\eta^2} \right). \qquad (7)$$
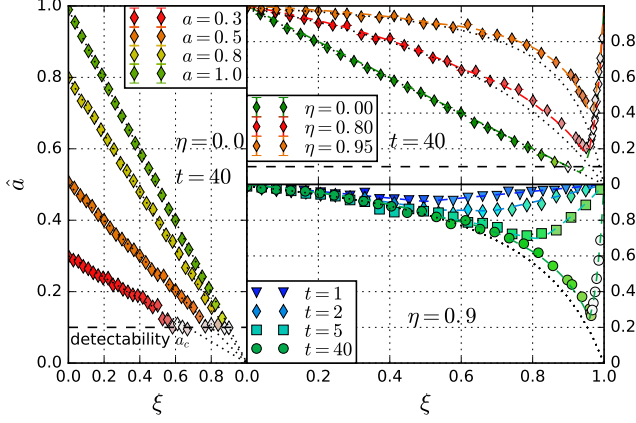
FIG. 2: Effective assortativity $\hat{a}$ inferred using EM from single snapshot observations of a DSBM with link persistence, for different values of ξ, η and t. Black dotted lines represent $a^c_{\xi,\eta}$ while dashed lines are the theoretical optimal $a^\star_t(\xi,\eta)$. Vividness of colours is proportional to the overlap $q(\boldsymbol{g}^t,\hat{\boldsymbol{g}}^t)$ between the planted and the inferred communities.

FIG. 3: Overlap $q(\hat{\boldsymbol{g}}^t, \boldsymbol{g}^{t-\tau})$ between the communities inferred at time t and the planted assignment at time $t - \tau$. Dashed lines are computed by solving the problem in Eq. 12.

Proof. It is sufficient to note that, from Bayes' rule, $P(\boldsymbol{g}^t|\boldsymbol{A}^t) \propto P(\boldsymbol{A}^t|\boldsymbol{g}^t)$, that can be always be written as

$$P(\boldsymbol{A}^t|\boldsymbol{g}^t) = \prod_{(i,j)} (p^t_{g^t_i g^t_j})^{A^t_{ij}} (1 - p^t_{g^t_i g^t_j})^{1-A^t_{ij}}, \qquad (8)$$

with $p^t_{ab} := P(A^t_{ij} = 1|g^t_i = a, g^t_j = b)$. Marginalising over previous network instances we get the recursive equation

$$\begin{aligned} p^t_{ab} &= \xi \ P(A^{t-1}_{ij} = 1|g^t_i = a, g^t_j = b) + (1 - \xi)p_{ab}. \\ &= \xi \ (\eta^2 p^{t-1}_{ab} + (1 - \eta^2)\bar{p}) + (1 - \xi)p_{ab}, \qquad (9) \end{aligned}$$

where in the first equality we have conditioned and summed over A^{t-1}_{ij}, while in the second over g^{t-1}_i, g^{t-1}_j. Since p^0_{ab} is simply p_{ab} we get

$$p^t_{ab} = \left(\xi(1 - \eta^2)\bar{p} + (1 - \xi)p_{ab}\right) \sum_{\ell=0}^{t-1} \left(\xi\eta^2\right)^\ell + \left(\xi\eta^2\right)^t p_{ab},$$

that gives (7) once used the representation (3). $\qquad \square$

The proposition shows that inference of a DSBM with link persistence from a single snapshot reduces to the inference of a static SBM where the detectability threshold is increased by the effect of the dynamics as $a^c \to a^c/\epsilon^t_{\xi,\eta}$ and reduces to the static threshold in absence of link persistence ($\xi = 0$). Note that the detectability threshold from single snapshot is however higher than the threshold of the dynamic problem, i.e. the inference of all the assignments given the observation of the entire network series. For example [13] considers a DSBM without link persistence and shows that the detectability threshold a^c is in general lowered by the communities persistence.
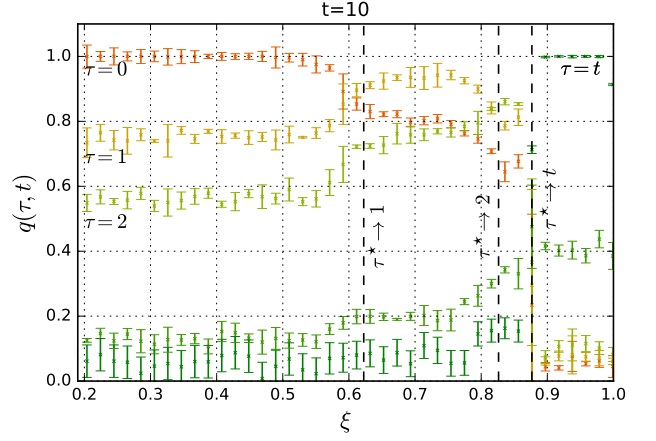
Persistence of links ξ decreases the effective assortative structure of the network, increasing the number of links assigned randomly with respect to those assigned on the base of their group labels, see left Panel of Fig. 2. This effect is partially mitigated by the persistence of communities η since it increases the probability that a link copied from a previous time is not actually random but was in turn assigned through the same community structure. For $t \to \infty$, it is $a^\infty_{\xi,\eta} = a(1-\xi)/(1 - \xi\eta^2)$ (see Fig. 1 for the asymptotic phase space).

Time lagged inference. In the detectable phase, inference on $\boldsymbol{A}^t$ should provide a set of assignment $\hat{\boldsymbol{g}}^t$ correlated with the planted $\boldsymbol{g}^t$ and an inferred assortativity $\hat{a}$ close to $a^t_{\xi,\eta}$. Surprisingly, in some regions $\hat{a}$ is higher than the theoretical one and in particular two transitions appear. The right panels of Fig. 2 show that increasing ξ, $\hat{a}$ becomes larger than $a^\infty_{\xi,\eta}$ (first transition), and later start to increase (second transition).

To understand this phenomenon, given a network sequence of length T generated with parameters (ξ,η,a), we call *time lagged inference* the problem of inferring communities at time $t - \tau$ given the observation of the network at time t. Since the posterior $P(\boldsymbol{g}^{t-\tau}|\boldsymbol{A}^t) \propto \prod_{ij} P(A^t_{ij}|\boldsymbol{g}^{t-\tau})$ we prove:

Proposition 2. *In the above model, the posterior $P(\boldsymbol{g}^{t-\tau}|\boldsymbol{A}^t)$ is that of a static SBM with an effective assortativity*

$$a^{(t,\tau)}_{\xi,\eta} = \xi^\tau a^{t-\tau}_{\xi,\eta} + (1 - \xi)\eta^2 \frac{\eta^{2\tau} - \xi^\tau}{\eta^2 - \xi} a,$$

where $a^t_{\xi,\eta}$ is given by (7).

Proof. As in Proposition 1 it sufficient to compute the quantity $\mathcal{L}^n = P(A^n_{ij} = 1|\boldsymbol{g}^{t-\tau} = \boldsymbol{g})$, evaluated at $n = t$.

For $n \geqslant t - \tau$, keepinig fixed i, j and t , it is

$$\mathcal{L}^n = \sum_{A_{ij}^{n-1}, \boldsymbol{g}^n} P(A_{ij}^n = 1|\boldsymbol{g}^n, A_{ij}^{n-1})P(\boldsymbol{g}^n, A_{ij}^{n-1}|\boldsymbol{g}^{t-\tau} = \boldsymbol{g})$$

$$= \xi \mathcal{L}^{n-1} + (1 - \xi) \sum_{\boldsymbol{g}^n} p_{g_i^n g_j^n} P(\boldsymbol{g}^n|\boldsymbol{g}^{t-\tau} = \boldsymbol{g}). \quad (10)$$

Moreover, defining $\mathcal{T}^n = \sum_{\boldsymbol{g}^n} p_{g_i^n g_j^n} P(\boldsymbol{g}^n|\boldsymbol{g}^{t-\tau} = \boldsymbol{g})$, for $n \geqslant t - \tau$ it is

$$\mathcal{T}^n = \sum_{\boldsymbol{g}^n, \boldsymbol{g}^{n-1}} p_{g_i^n g_j^n} P(\boldsymbol{g}^n|\boldsymbol{g}^{n-1})P(\boldsymbol{g}^{n-1}|\boldsymbol{g}^{t-\tau} = \boldsymbol{g})$$

$$= \eta^2 \mathcal{T}^{n-1} + (1 - \eta^2)\bar{p} \quad (11)$$

Solving (11) and then (10), i.e. the recursive equation $\mathcal{L}^n = \xi \mathcal{L}^{n-1} + (1 - \xi)\mathcal{T}^n$ we get

$$\mathcal{L}^t = \xi^\tau \mathcal{L}^{t-\tau} + (1 - \xi) \sum_{\ell=0}^{\tau-1} \xi^\ell \mathcal{T}^{t-l}$$

$$= \xi^\tau \mathcal{L}^{t-\tau} + (1 - \xi) \sum_{\ell=0}^{\tau-1} \xi^\ell \left(\eta^{2(\tau-\ell)} p_{g_i g_j} + (1 - \eta^{2(\tau-\ell)})\bar{p} \right).$$

Since $\mathcal{L}^{t-\tau}$ corresponds to the non lagged $p_{g_i g_j}^{t-\tau}$ in Proposition 1, we get the result simply using the representation (3) $\square$

The meaning of proposition 2 is that every lagged inference problem has the posterior of a static SBM with effective assortativity $a_{\xi,\eta}^{(t,\tau)}$. Thus fixing t and varying τ we have a sequence of inference problems with the same posterior, same input data $\boldsymbol{A}^t$, and only different effective assortativity, thus detectability threshold. Fig. 3 shows the overlap between the inferred communities $\hat{\boldsymbol{g}}^t$ and the planted ones at $t - \tau$. For small ξ the maximum overlap is with $\boldsymbol{g}^t$, while for larger ξ we observe a series of transitions where the largest overlap is with a $\boldsymbol{g}^{t-\tau}$ with $\tau > 0$. We now show that the τ that maximizes the overlap $q(\hat{\boldsymbol{g}}^t, \boldsymbol{g}^{t-\tau})$ is the one for which the effective assortativity $a_{\xi,\eta}^{(t,\tau)}$ is maximal. To this end we define

$$a_t^\star(\xi, \eta) = \max_{\tau \leqslant t} a_{\xi,\eta}^{(t,\tau)}; \quad \tau_t^\star(\xi, \eta) = \underset{\tau \leqslant t}{\mathrm{argmax}} \, a_{\xi,\eta}^{(t,\tau)} \quad (12)$$

Top left panels of Fig. 4 show that for small link persistence ξ, $\tau_t^\star(\xi, \eta) = 0$, i.e. the EM algorithm solves the problem at the time of the observed snapshot t. At a critical ξ, depending on η and t, it is $\tau_t^\star(\xi, \eta) > 0$, suggesting that the algorithm converges to the solution at time $t - \tau_t^\star$. In fact the dashed lines in Fig. 3 are computed by solving the problem in Eq. 12 and it is clear that they correspond to the transitions in the overlap. Moreover the theoretical $a_t^\star(\xi, \eta)$ is shown in Fig. 2 to be in perfect agreement with the inferred assortativity $\hat{a}$.

To get more intuition, we note that for large t

$$a_{\xi,\eta}^{(\tau)} := \lim_{t \to \infty} a_{\xi,\eta}^{(t,\tau)} = a \left(\xi^\tau \frac{1 - \xi}{1 - \xi\eta^2} + (1 - \xi)\eta^2 \frac{\eta^{2\tau} - \xi^\tau}{\eta^2 - \xi} \right).$$
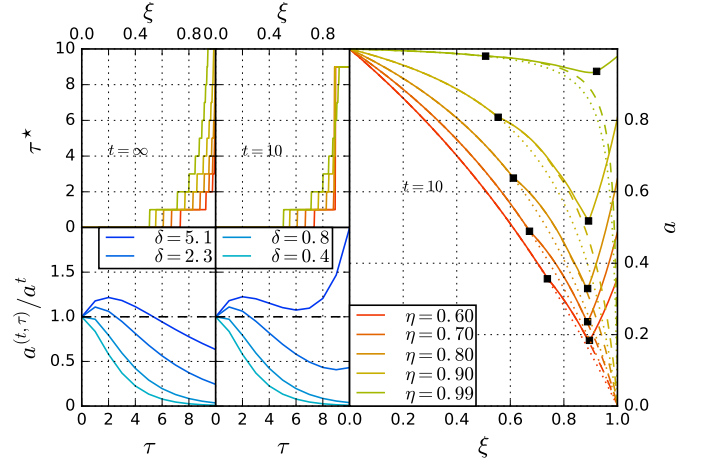

FIG. 4: *Left bottom panels*: $a_{\xi,\eta}^{(t,\tau)}/a_{\xi,\eta}^t$ as function of the lag τ for different δ. *Left top panels*: optimal lag $\tau_t^*(\xi, \eta)$ a function of ξ and η. *Right panel*: Optimal effective assortativity $a_t^\star(\xi, \eta)$ (full lines) compared with the non lagged $a_{\xi,\eta}^t$ (dotted line) and the asymptotic $a_\infty^\star(\xi, \eta)$ (dashed lines). Black squares indicate the two transitions (see text).

Since $a_{\xi,\eta}^{(\tau)} \to 0$ as $\tau \to \infty$, when $a_{\xi,\eta}^{(1)} > a_{\xi,\eta}^{(0)}$, i.e.

$$\delta := \eta^2 \frac{\xi}{1 - \xi} > 1 \quad (13)$$

the maximum of $a_{\xi,\eta}^{(\tau)}$ is not anymore at $\tau = 0$ (see the bottom left panel of Fig. 4).

For finite t, there is a finite size effect since the range of τ is bounded by t. In this situation for large ξ and η the maximum of $a_{\xi,\eta}^{(t,\tau)}$ is achieved at the extremum $\tau = t$ (bottom central panel of Fig. 4). Finally, the right panel of Fig. 4 compares $a_t^\star(\xi, \eta)$, $a_{\xi,\eta}^t$, and $a_\infty^\star(\xi, \eta)$. The black squares indicate the two transitions, the first one from zero to positive τ^* (computed with Eq. 13) and the second when $\tau^* = t$ due to the finite size effect. These correspond to the transitions observed in the empirical analysis of the right panels of Fig. 2.

Conclusions. We studied the detectability transition in a network model where both communities and links are time varying. We focused on static algorithms for temporal networks, where inference is performed on each snapshot network. We found that link persistence is the driver of a new kind of detectability transition, time lagged inference, i.e. the wrong detection of a past community rather than a present community. The framework of persistent dynamic network model we introduced allows to test whether an algorithm of dynamic community detection presents such kind of failure and to design dynamic algorithms. Finding new optimal algorithms will be the object of future research.

————————

[1] P. Holme, and J. Saramaki, Physics reports 519, no. 3: 97-125 (2012).

[2] S. Boccaletti *et al.*, Physics Reports, 544, 1-122 (2014).

[3] M. E. J. Newman, Phys. Rev. E 94, no. 5: 052315 (2016).

[4] B. Hendrickson, and R. Leland. SIAM Journal on Scientific Computing 16.2: 452-469 (1995).

[5] F. Krzakala, C. Moore, E. Mossel, J. Neeman, A. Sly, L. Zdeborová, P. Zhang, Proc. Natl. Acad. Sci. USA 110 20935-20940 (2013).

[6] A. Decelle, F. Krzakala, C. Moore, L. Zdeborová, Phys. Rev. E 84 (6), 066106 (2011).

[7] V. D. Blondel *et al.*, Journal of statistical mechanics: theory and experiment 2008.10: P10008 (2008).

[8] E. Mossel, J. Neeman, and A. Sly, arXiv preprint arXiv:1311.4115 (2013).

[9] P. J. Mucha, T. Richardson, K. Macon, M. A. Porter, and J. Onnela. Science 328, no. 5980: 876-878 (2010).

[10] T. Yang, Y. Chi, S. Zhu, Y. Gong, and R. Jin, Machine learning 82, no. 2: 157-189 (2011).

[11] D. S. Bassett, M. A. Porter, N. F. Wymbs, S. T. Grafton, J. M. Carlson, and P. J. Mucha, Chaos: An Interdisciplinary Journal of Nonlinear Science 23, no. 1: 013142 (2013).

[12] M. Bazzi, M. A. Porter, S. Williams, M. McDonald, D. J. Fenn, and S. D. Howison, Multiscale Modeling & Simulation 14, no. 1: 1-41 (2016).

[13] A. Ghasemian, P. Zhang, A. Clauset, C. Moore, L. Peel, Phys. Rev. X 6, 031005 (2016).

[14] X. Zhang, C. Moore, and M.E.J. Newman, arXiv preprint arXiv:1607.07570, (2016).

[15] K. S. Xu, and A. O. Hero, IEEE Journal of Selected Topics in Signal Processing 8, no. 4: 552-562 (2014).

[16] K. S. Xu, Stochastic Block Transition Models for Dynamic Networks, AISTATS (2015).

[17] L. A. N. Amaral, A. Scala, M. Barthelemy and H.E. Stanley, Proc. Natl. Acad. Sci. USA 97.21: 11149-11152 (2000).

[18] E. Mossel, J. Neeman, and A. Sly, Probability Theory and Related Fields , 1 (2012).

[19] J. Friedman, T. Hastie, and R. Tibshirani, *The elements of statistical learning* (Springer, Berlin, 2001). Vol. 1.

[20] P.T. Peixoto, Physical review letters 110, 14: 148701 (2013).

[21] A. Decelle, F. Krzakala, C. Moore, L. Zdeborová, Phys. Rev. Lett. 107 (6), 065701 (2011).

[22] Y. Iba, Journal of Physics A: Mathematical and General 32, 3875 (1999)