The Value of Temporal Data for Learning of Influence Networks: a Characterization via Kullback-Leibler Divergence

Munther A. Dahleh, John N. Tsitsiklis, and Spyros I. Zoumpoulis

Abstract— We infer local influence relations between networked entities from data on outcomes and assess the value of temporal data by formulating relevant binary hypothesis testing problems and characterizing the speed of learning of the correct hypothesis via the Kullback-Leibler divergence, under three different types of available data: knowing the set of entities who take a particular action; knowing the order that the entities take an action; and knowing the times of the actions.

I. INTRODUCTION

An article trending in the blogosphere [17], a topic trending on an online social network [24], computer malware spreading across a network [15], customers sequentially adopting a new product on an e-commerce platform [14], are all examples of temporal processes governed by local interactions between networked entities, which influence one another. Due to the increasing capability of data acquisition technologies, rich data on the outcomes of such processes are oftentimes available (possibly with time stamps), yet the underlying network of local interactions is hidden. In this work, we infer who influences whom in a network of interacting entities based on data of their actions/decisions, and quantify the gain of learning based on times of actions, versus sequences of actions, versus sets of actions.

a) **Background and related literature:** Untangling and quantifying local influences in a principled manner, based on observed outcomes, is a challenging task, as there are many different confounding factors that may lead to seemingly similar phenomena. In recent work, inference of causal relationships has been possible from multivariate time-series data [18], [20], [16]. Solutions for the influence discovery problem have been proposed, which, similarly to this work, treat time explicitly as a continuous random variable and infer the network through cascade data, e.g., [10], [21], [12], [11]. The focus of our work is not just to infer the underlying network, but rather to quantify the gain in speed of learning, due to having access to richer temporal information.

Most closely related to this work are [8], [2], [1], [22], [9], [13]. The first five all derive sample/trace complexity results for the network inference problem. [2] and [13] share with us the question of reconstructing a graph from traces defined as sets of unordered nodes. Similarly to [1], we

assume exponentially distributed infection times, yet we are after the influence rate of each edge, rather than learning whether each edge exists or not. Our scope differs from the works mentioned above, as we wish to compare explicitly the speed of learning when having access to datasets with times of actions, versus just sequences of actions, versus just sets of actions; and we expand our discussion in [8] by using Kullback-Leibler divergence as a measure for the speed of learning.[1] Furthermore, the models assumed by the works mentioned above differ from the model we study, mainly in that we allow for self-induced infections (not just in the initial seeding), which makes the inference problem harder.

Recent research has focused on learning graphical models (which subsumes the question of identifying the connectivity in a network), either allowing for latent variables (e.g., [5], [6]) or not (e.g.,[3]). Instead of proposing and learning a general graphical model, we focus on a simple parametric model that can capture the sequence and timing of actions naturally, without the descriptive burden of a standard graphical model.

b) **Overview:** The overarching theme of our work is to quantify the gain in speed of learning of parametric models of influence, due to having access to richer temporal information. Using Kullback-Leibler (KL) divergence as a measure for speed of learning, we compare learning under three different cases of available data: (i) the data provides merely the set of agents/entities who took an action; (ii) the data provides the (ordered) sequence of agents/entities who took an action, but not the times; and (iii) the data provides the times of the actions. It is clear that learning is no slower with times than it is with sequences, and no slower with sequences than with sets; yet, what can we say about *how much* faster learning is with times than with sequences, and with sequences than with sets? This is, to the best of our knowledge, a comparison that has not been studied systematically before.[2]

We propose a parametric model of influence which captures directed pairwise interactions. We focus on learning the influence model in three particular instances, which we cast as respective binary hypothesis testing problems: Which of two agents influences a third agent? Is an agent influenced by another agent, or are her decisions self-induced? And is the influence between two agents large or small? We view these three questions as building blocks for understanding complex interactions in general networks. Given a hypothesis testing problem, the Kullback-Leibler divergence between

---

[1][1] also makes use of the Kullback-Leibler divergence, in order to quantify the sample complexity when timestamps are disregarded.

[2][22] finds such a comparison highly relevant.

the distributions pertaining to the two competing hypotheses yields the best achievable asymptotic exponent for the probability of error in a Neyman-Pearson setting. For each of the proposed hypothesis testing problems, we compare the Kullback-Leibler divergences in the cases of learning based on data of sets of decisions; learning based on data of sequences of decisions; and learning based on data of times of decisions.

We show that, when the data for each independent observation is collected over a small horizon, the sets of decisions provide almost all the necessary information for learning, and there is no value in richer temporal data for moderate values of the influence parameter. When the data for each independent observation is collected over a large horizon, then sequences have a large gain over sets, and times have smaller gain over sequences for moderate values of the influence parameter, for the first and second hypothesis testing problems; for the third problem, in which the two hypotheses are asymmetric with respect to the rate with which the influenced agent adopts, sequences have no gain over sets, while times have a large gain even for small values of the influence parameter. When the data for each independent observation is collected over a moderate horizon, times have some gain over sequences and sets for moderate values of the influence parameter.

The rest of the paper is organized as follows. Section II introduces the influence model and formulates the binary hypothesis testing problems. Section III studies the problem of which of two peers influences a third agent crucially ("peer" and "agent" are used interchangeably). Section IV studies the problem of whether an agent is influenced by another, or her decisions are self-induced. Section V studies the problem of whether the influence between two agents is large or small. Section VI concludes.

II. MODEL AND PROBLEM FORMULATION

A product becomes available at time $t = 0$ and each of a collection of n agents may adopt it or not. (In this paper the word "product" is used throughout, but could be interchanged by any of the following: information, behavior, opinion, virus, etc., depending on the context.) Agent i adopts it at a time that is exponentially distributed with rate $\lambda_i \geq 0$. After agent i adopts, the rate of adoption for all other agents $j \neq i$ increases by $\lambda_{ij} \geq 0$. The overall time horizon, capturing the duration of the adoption process, is modeled as an exponentially distributed random variable with rate λ_{hor}. No adoptions are possible after the end of the horizon. We study the adoption decisions for a collection of products, assuming that the parameters are static across products, and adoptions across products are independent.

As in our work [8], the proposed cascade model suggests a recursive definition for the times of adoption for each agent given a product c, which we denote $\{T_c^i\}_{i=1}^n$. We define $T_c^i = \infty$ if agent i does not adopt product c. Given product c, we consider the following three data modes:

- learning with *sets* of adoptions: the learner observes vector $\left(1_{\{T_c^1 < \infty\}}, \ldots, 1_{\{T_c^n < \infty\}}\right)$, i.e., whether each

agent adopts or not;
- learning with *sequences* of adoptions: the learner observes vector $\left(R_c^1, \ldots, R_c^n\right)$, where R_c^i denotes the relative order of T_c^i in $\{T_c^j\}_{j=1}^n$. If $T_c^i = \infty$, define $R_c^i = \infty$. The learner, that is, observes who adopts and in what order.
- learning with *times* of adoptions: the learner observes vector $\left(T_c^1, \ldots, T_c^n\right)$, i.e., the times at which agents adopt (if at all).

We propose three binary hypothesis testing problems:
1) Which of two peers influences you crucially?
2) Are you influenced by your peer, or is it yourself?
3) Does your peer influence you a lot or a little?

In the context of binary hypothesis testing in the Neyman-Pearson setting, the Chernoff-Stein lemma yields the asymptotically optimal exponent for the probability of error of one type, under the constraint that the probability of error of the other type is less than ϵ. More specifically, given hypotheses H_1 and H_2, and corresponding probability measures P_1, P_2, the best achievable exponent for the probability of error of deciding in favor of H_1 when H_2 is true, given that the probability of deciding in favor of H_2 when H_1 is true is less than ϵ, is given by the negative Kullback-Leibler divergence between the two measures $-D(P_1||P_2) \equiv -\mathbb{E}_{P_1}\left[\log \frac{dP_1}{dP_2}\right]$, where $\frac{dP_1}{dP_2}$ denotes the Radon-Nikodym derivative of the two measures (see, for example, [7]).

For each hypothesis testing problem, we observe i.i.d. observations drawn from the true distribution; the observations can be a collection of sets of adopters, a collection of sequences of adopters, or a collection of times of adoptions, depending on how much information is provided in the available data. We use $KL_{\text{set}}, KL_{\text{sequence}}, KL_{\text{time}}$ to denote the Kullback-Leibler divergence of the two distributions pertaining to the two hypotheses under the cases of learning with data which only provides the sets of adopters, learning with data which provides the sequence of adopters but not exact times, and learning with data which provides exact times of adoptions, respectively. A greater Kullback-Leibler divergence implies a faster decaying probability of error, which in turn means that fewer i.i.d. observations are needed in order for the probability of error to become sufficiently small. We are interested in the relation between the Kullback-Leibler divergences for the cases of sets, sequences, and times of adoptions;[3] this relation reveals how faster learning becomes asymptotically with temporally richer data.

III. WHICH OF TWO PEERS INFLUENCES YOU CRUCIALLY?

We consider the hypothesis testing problem illustrated in Figure 1. In words, according to Hypothesis I, agent 1 influences agent 3 with rate α, while agent 2 influences agent 3 with rate 1; according to Hypothesis II, agent 1 influences agent 3 with rate 1, while agent 2 influences agent 3 with rate α.

---

[3]It is clear that $KL_{\text{set}} \leq KL_{\text{sequence}} \leq KL_{\text{time}}$; we are interested in the relative scaling of the KL divergences and how it changes with the parameters.
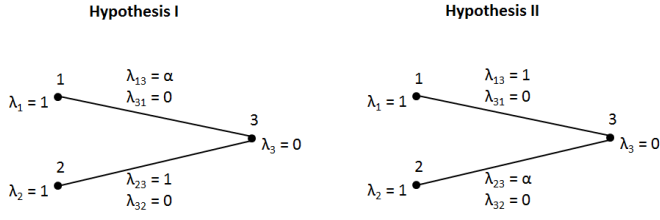
Fig. 1. The hypothesis testing problem: which of agents 1, 2 crucially influences agent 3?

The probability mass functions needed for the calculation of $KL_{\text{set}}, KL_{\text{sequence}}$ are straightforward to compute. For example, the probability of the *sequence* of agents $\{1, 2\}$ occurring is $\lambda_1/(\lambda_1 + \lambda_2 + \lambda_{hor}) \cdot \lambda_2/(\lambda_2 + \lambda_{13} + \lambda_{hor}) \cdot \lambda_{hor}/(\lambda_{13} + \lambda_{23} + \lambda_{hor})$, the probability of the *sequence* of agents $\{2, 1\}$ is $\lambda_2/(\lambda_1 + \lambda_2 + \lambda_{hor}) \cdot \lambda_1/(\lambda_1 + \lambda_{23} + \lambda_{hor}) \cdot \lambda_{hor}/(\lambda_{13} + \lambda_{23} + \lambda_{hor})$, and the probability of the *set* of agents $\{1, 2\}$ is the sum of the two. Denoting with p_A^i the probability of a sequence of adopters given by ordered set A according to Hypothesis i, we can write

$$
\begin{aligned}
KL_{\text{set}} = {} & p_\emptyset^I \log \frac{p_\emptyset^I}{p_\emptyset^{II}} + p_{\{1\}}^I \log \frac{p_{\{1\}}^I}{p_{\{1\}}^{II}} + p_{\{2\}}^I \log \frac{p_{\{2\}}^I}{p_{\{2\}}^{II}} \\
& + \left(p_{\{1,2\}}^I + p_{\{2,1\}}^I \right) \log \frac{p_{\{1,2\}}^I + p_{\{2,1\}}^I}{p_{\{1,2\}}^{II} + p_{\{2,1\}}^{II}} \\
& + p_{\{1,3\}}^I \log \frac{p_{\{1,3\}}^I}{p_{\{1,3\}}^{II}} + p_{\{2,3\}}^I \log \frac{p_{\{2,3\}}^I}{p_{\{2,3\}}^{II}} \\
& + \left(p_{\{1,2,3\}}^I + p_{\{2,1,3\}}^I + p_{\{1,3,2\}}^I + p_{\{2,3,1\}}^I \right) \\
& \cdot \log \frac{p_{\{1,2,3\}}^I + p_{\{2,1,3\}}^I + p_{\{1,3,2\}}^I + p_{\{2,3,1\}}^I}{p_{\{1,2,3\}}^{II} + p_{\{2,1,3\}}^{II} + p_{\{1,3,2\}}^{II} + p_{\{2,3,1\}}^{II}}
\end{aligned}
$$

and

$$
\begin{aligned}
KL_{\text{sequence}} = {} & p_\emptyset^I \log \frac{p_\emptyset^I}{p_\emptyset^{II}} + p_{\{1\}}^I \log \frac{p_{\{1\}}^I}{p_{\{1\}}^{II}} + p_{\{2\}}^I \log \frac{p_{\{2\}}^I}{p_{\{2\}}^{II}} \\
& + p_{\{1,2\}}^I \log \frac{p_{\{1,2\}}^I}{p_{\{1,2\}}^{II}} + p_{\{2,1\}}^I \log \frac{p_{\{2,1\}}^I}{p_{\{2,1\}}^{II}} \\
& + p_{\{1,3\}}^I \log \frac{p_{\{1,3\}}^I}{p_{\{1,3\}}^{II}} + p_{\{2,3\}}^I \log \frac{p_{\{2,3\}}^I}{p_{\{2,3\}}^{II}} \\
& + p_{\{1,2,3\}}^I \log \frac{p_{\{1,2,3\}}^I}{p_{\{1,2,3\}}^{II}} + p_{\{2,1,3\}}^I \log \frac{p_{\{2,1,3\}}^I}{p_{\{2,1,3\}}^{II}} \\
& + p_{\{1,3,2\}}^I \log \frac{p_{\{1,3,2\}}^I}{p_{\{1,3,2\}}^{II}} + p_{\{2,3,1\}}^I \log \frac{p_{\{2,3,1\}}^I}{p_{\{2,3,1\}}^{II}}.
\end{aligned}
$$

From the log-sum inequality $\sum_{i=1}^n a_i \log \frac{a_i}{b_i} \geq \left(\sum_{i=1}^n a_i \right) \log \frac{\sum_{i=1}^n a_i}{\sum_{i=1}^n b_i}$ for nonnegative numbers, $a_1, a_2, \ldots, a_n$ and $b_1, b_2, \ldots, b_n$, it is clear that $KL_{\text{set}} \leq KL_{\text{sequence}}$.

For the probability density functions needed for the calculation of KL_{time}, one can consider $T_1, T_2, T_{13}, T_{23}, T_{hor}$, the exponentially distributed random variables modeling the time of adoption by agent 1, the time of adoption by agent 2, the time of adoption by agent 3 due to agent 1, the time of adoption by agent 3 due to agent 2, and the end of the horizon, respectively. Then $T_3 = \min(T_{13}, T_{23})$ models the time of adoption by agent 3 (due to either agent 1

or agent 2). One can then consider the joint density of $T_1, T_2, T_{13}, T_{23}, T_{hor}$. Nevertheless, the available data only captures realization of random variables that occurred before the realization of T_{hor}. Therefore, the calculation of the joint densities should account only for observed outcomes.

In Figures 2 and 3 we plot the KL divergences and their ratios against influence rate α, for different horizon rates.
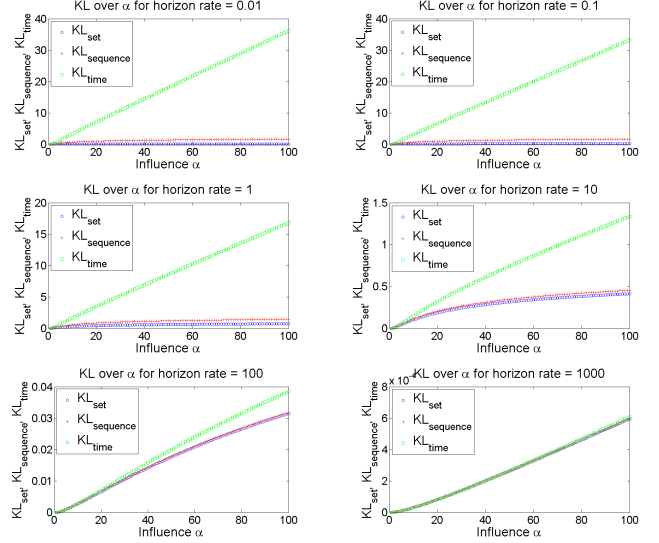

Fig. 2. Which of two peers influences you? Plots of KL_{set} (circles), KL_{sequence} (crosses), KL_{time} (squares) against influence rate α for different horizon rates.
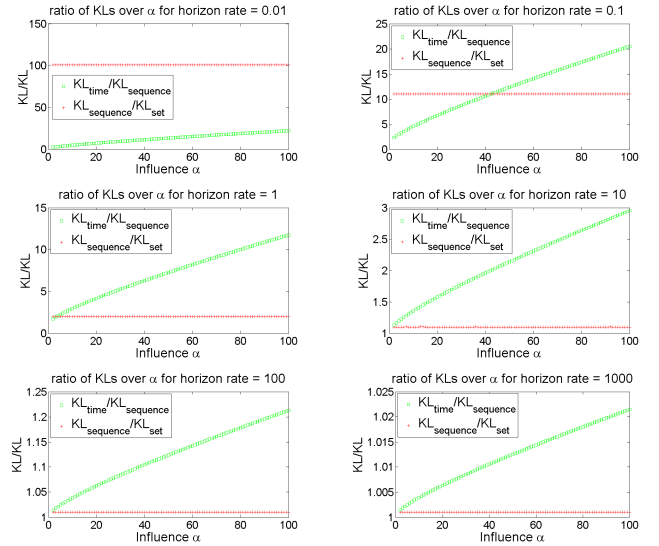

Fig. 3. Which of two peers influences you? Plots of $KL_{\text{sequence}}/KL_{\text{set}}$ (crosses), $KL_{\text{time}}/KL_{\text{sequence}}$ (squares) against influence rate α for different horizon rates.

In the large horizon regime (i.e., when λ_{hor} is small), knowing the sequences of adoptions has a large gain over knowing the sets of adoptions. On the other hand, knowing the times of adoptions has large gain over knowing the sequences of adoptions only for large enough values of the

influence rate α. For small values of α, the gain of times over sequences is small compared to the gain of sequences over sets.

In the small horizon regime (i.e., when λ_{hor} is large), the sets of adoptions give almost all the information for learning, and there is no much further value in richer temporal data. Sequences have no gain (in the limit as $\lambda_{hor} \longrightarrow \infty$) over sets, while for times to have significant gain over sequences, the rate of influence α has to be large.

In the moderate horizon regime, knowing the times of adoptions has some value over knowing merely sequences or sets of adoptions even for small values of the influence rate α.

IV. ARE YOU INFLUENCED BY YOUR PEER OR IS IT YOURSELF?

We consider the hypothesis testing problem illustrated in Figure 4. In words, according to Hypothesis I, agent 1 influences agent 2 with rate α, while agent 2 adopts herself with rate 1; according to Hypothesis II, agent 1 influences agent 2 with rate 1, while agent 2 adopts herself with rate α.
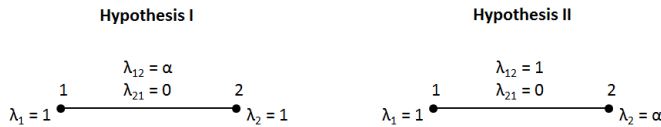

Fig. 4. The hypothesis testing problem: is agent 2 influenced by agent 1, or does she have a high individual rate?

The plots of the KL divergences and their ratios are omitted, as the qualitative analysis is similar to the analysis of the hypothesis testing problem of which of two peers influences crucially another agent. In the large horizon regime (i.e., when λ_{hor} is small), knowing the sequences of adoptions has a large gain over knowing the sets of adoptions, and knowing the times of adoptions yields much smaller gain over knowing just sequences for small values of the rate α. In the small horizon regime (i.e., when λ_{hor} is large), the sets of adoptions give almost all the information for learning, and there is no much further value in richer temporal data. In the moderate horizon regime, knowing the times of adoptions has some value over knowing merely sequences or sets of adoptions even for small values of the rate α.

The value of learning with sequences of adoptions over learning with sets of adoptions can be readily decided analytically by looking at the relevant limits. First we write

$$
\begin{aligned}
KL_{\text{set}}(\alpha, \lambda_{hor}) = & \frac{\lambda_{hor}}{2 + \lambda_{hor}} \log \frac{1 + \alpha + \lambda_{hor}}{2 + \lambda_{hor}} \\
& + \frac{1}{2 + \lambda_{hor}} \frac{\lambda_{hor}}{1 + \alpha + \lambda_{hor}} \log \frac{1 + \alpha + \lambda_{hor}}{2 + \lambda_{hor}} \\
& + \frac{1}{2 + \lambda_{hor}} \frac{\lambda_{hor}}{1 + \lambda_{hor}} \log \frac{1 + \alpha + \lambda_{hor}}{\alpha(2 + \lambda_{hor})} \\
& + \left(\frac{1}{2 + \lambda_{hor}} \frac{1 + \alpha}{1 + \alpha + \lambda_{hor}} + \frac{1}{2 + \lambda_{hor}} \frac{1}{1 + \lambda_{hor}} \right) \\
& \cdot \log \frac{\frac{1}{2+\lambda_{hor}} \frac{1+\alpha}{1+\alpha+\lambda_{hor}} + \frac{1}{2+\lambda_{hor}} \frac{1}{1+\lambda_{hor}}}{\frac{1}{1+\alpha+\lambda_{hor}} \frac{1+\alpha}{1+\alpha+\lambda_{hor}} + \frac{\alpha}{1+\alpha+\lambda_{hor}} \frac{1}{1+\lambda_{hor}}}
\end{aligned}
$$

and

$$
\begin{aligned}
KL_{\text{sequence}}(\alpha, \lambda_{hor}) = & \frac{\lambda_{hor}}{2 + \lambda_{hor}} \log \frac{1 + \alpha + \lambda_{hor}}{2 + \lambda_{hor}} \\
& + \frac{1}{2 + \lambda_{hor}} \frac{\lambda_{hor}}{1 + \alpha + \lambda_{hor}} \log \frac{1 + \alpha + \lambda_{hor}}{2 + \lambda_{hor}} \\
& + \frac{1}{2 + \lambda_{hor}} \frac{\lambda_{hor}}{1 + \lambda_{hor}} \log \frac{1 + \alpha + \lambda_{hor}}{\alpha(2 + \lambda_{hor})} \\
& + \frac{1}{2 + \lambda_{hor}} \frac{1 + \alpha}{1 + \alpha + \lambda_{hor}} \log \frac{1 + \alpha + \lambda_{hor}}{2 + \lambda_{hor}} \\
& + \frac{1}{2 + \lambda_{hor}} \frac{1}{1 + \lambda_{hor}} \log \frac{1 + \alpha + \lambda_{hor}}{\alpha(2 + \lambda_{hor})}.
\end{aligned}
$$

For fixed α, we have

$$
\lim_{\lambda_{hor} \longrightarrow 0} KL_{\text{set}}(\alpha, \lambda_{hor}) = 0
$$

$$
\lim_{\lambda_{hor} \longrightarrow 0} KL_{\text{sequence}}(\alpha, \lambda_{hor}) = \frac{1}{2} \log \frac{1 + \alpha}{2} + \frac{1}{2} \log \frac{1 + \alpha}{2\alpha},
$$

which implies

$$
\lim_{\lambda_{hor} \longrightarrow 0} \frac{KL_{\text{sequence}}(\alpha, \lambda_{hor})}{KL_{\text{set}}(\alpha, \lambda_{hor})} = \infty,
$$

for $\alpha \neq 1$, establishing that in the large horizon regime, learning with sequences yields significant gain over learning with sets.

For fixed α, and adopting the definition of the Kullback-Leibler divergence which uses the convention $0 \log \frac{0}{0} = 0$, we have

$$
\begin{aligned}
\lim_{\lambda_{hor} \longrightarrow \infty} \frac{KL_{\text{sequence}}(\alpha, \lambda_{hor})}{KL_{\text{set}}(\alpha, \lambda_{hor})} & = \frac{\frac{\lambda_{hor}}{2+\lambda_{hor}} \log \frac{1+\alpha+\lambda_{hor}}{2+\lambda_{hor}}}{\frac{\lambda_{hor}}{2+\lambda_{hor}} \log \frac{1+\alpha+\lambda_{hor}}{2+\lambda_{hor}}} \\
& = 1,
\end{aligned}
$$

which establishes that in the small horizon regime, learning with sequences has insignificant gain over learning with sets.

We can reach the same conclusions by looking at the limit as $\alpha \longrightarrow \infty$, for fixed λ_{hor}. Indeed, for fixed λ_{hor}, we have

$$
\lim_{\alpha \longrightarrow \infty} \frac{KL_{\text{sequence}}(\alpha, \lambda_{hor})}{KL_{\text{set}}(\alpha, \lambda_{hor})} = \frac{\lambda_{hor} + 1}{\lambda_{hor}},
$$

which in turn becomes arbitrarily large for $\lambda_{hor} \longrightarrow 0$, and converges to 1 for $\lambda_{hor} \longrightarrow \infty$.

V. DOES YOUR PEER INFLUENCE YOU A LOT OR A LITTLE?

We consider the hypothesis testing problem illustrated in Figure 5. In words, according to Hypothesis I, agent 1 influences agent 2 with rate α; according to Hypothesis II, agent 1 influences agent 2 with rate 1.
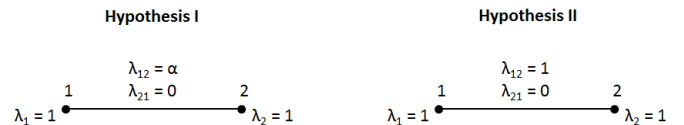

Fig. 5. The hypothesis testing problem: is agent 2 influenced by agent 1 a lot or a little?

In Figures 6 and 7 we plot the KL divergences and their ratios against influence rate α, for different horizon rates.

In the large horizon regime (i.e., when λ_{hor} is small), knowing the times of adoptions has a large gain over
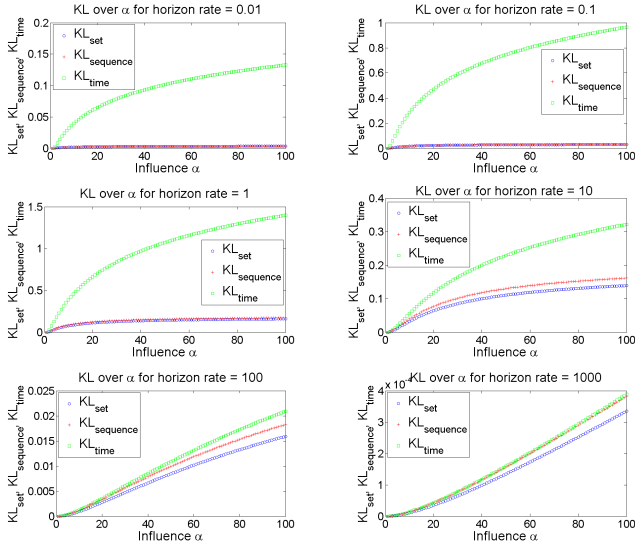
Fig. 6. Does your peer influence you a lot or a little? Plots of KL_{set} (circles), KL_{sequence} (crosses), KL_{time} (squares) against influence rate α for different horizon rates.
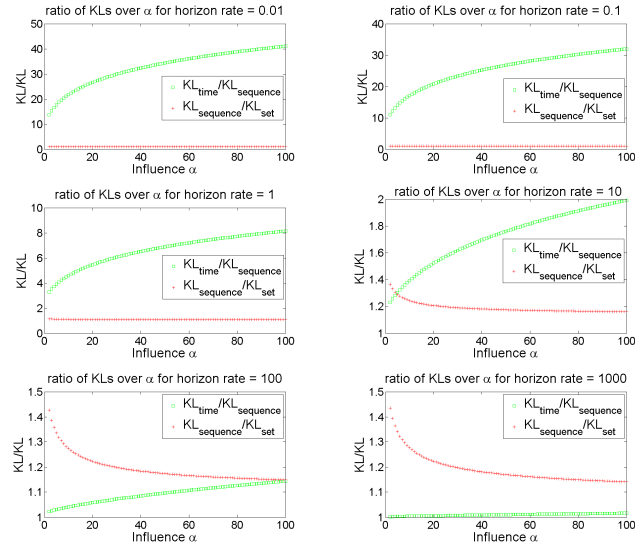

Fig. 7. Does your peer influence you a lot or a little? Plots of $KL_{\text{sequence}}/KL_{\text{set}}$ (crosses), $KL_{\text{time}}/KL_{\text{sequence}}$ (squares) against influence rate α for different horizon rates.

knowing the sequences or sets of adoptions. On the other hand, knowing the sequences of adoptions does not have value over knowing just the sets of adoptions.

In the small horizon regime (i.e., when λ_{hor} is large), the sets of adoptions give almost all the information for learning, and there is no much further value in richer temporal data. Sequences have no significant gain over sets, while times have even less gain over sequences for moderate values of the influence rate α.

In the moderate horizon regime, knowing the times of adoptions has some value over knowing merely sequences or sets of adoptions even for small values of the influence rate α. For constant α, the gain becomes larger for larger

horizon.

The reason why time has significant gain in the large horizon regime, even for small values of the influence rate, is the difference between the two hypotheses with respect to the total rate with which agent 2 adopts, after agent 1 has adopted (which is $1 + \alpha$ under Hypothesis I, and 2 under Hypothesis II). When the horizon is long, having time information allows for more accurate learning of the rate with which agent 2 adopts after agent 1 has adopted, and therefore for better asymptotic optimal error exponent in the hypothesis test.

The value of learning with sequences of adoptions over learning with sets of adoptions can be readily decided analytically by looking at the relevant limits. First we write that $KL_{\text{set}}(\alpha, \lambda_{hor})$ is equal to

$$\frac{1}{2 + \lambda_{hor}} \frac{\lambda_{hor}}{1 + \alpha + \lambda_{hor}} \log \frac{2 + \lambda_{hor}}{1 + \alpha + \lambda_{hor}}$$
$$+ \left(\frac{1}{2 + \lambda_{hor}} \frac{1 + \alpha}{1 + \alpha + \lambda_{hor}} + \frac{1}{2 + \lambda_{hor}} \frac{1}{1 + \lambda_{hor}} \right)$$
$$\cdot \log \frac{\frac{1 + \alpha}{1 + \alpha + \lambda_{hor}} + \frac{1}{1 + \lambda_{hor}}}{\frac{2}{2 + \lambda_{hor}} + \frac{1}{1 + \lambda_{hor}}}$$

and $KL_{\text{sequence}}(\alpha, \lambda_{hor})$ is equal to

$$\frac{1}{2 + \lambda_{hor}} \frac{\lambda_{hor}}{1 + \alpha + \lambda_{hor}} \log \frac{2 + \lambda_{hor}}{1 + \alpha + \lambda_{hor}}$$
$$+ \frac{1}{2 + \lambda_{hor}} \frac{1 + \alpha}{1 + \alpha + \lambda_{hor}} \log \frac{(1 + \alpha)(2 + \lambda_{hor})}{2(1 + \alpha + \lambda_{hor})}.$$

For fixed λ_{hor}, we have

$$\lim_{\alpha \longrightarrow \infty} \frac{KL_{\text{sequence}}(\alpha, \lambda_{hor})}{KL_{\text{set}}(\alpha, \lambda_{hor})} = \frac{\frac{1}{2 + \lambda_{hor}} \log \frac{2 + \lambda_{hor}}{2}}{\frac{1}{1 + \lambda_{hor}} \log \frac{1 + \frac{1}{1 + \lambda_{hor}}}{\frac{2}{2 + \lambda_{hor}} + \frac{1}{1 + \lambda_{hor}}}}$$
$$= \frac{1 + \lambda_{hor}}{2 + \lambda_{hor}} \frac{\log \frac{2 + \lambda_{hor}}{2}}{\log \frac{(2 + \lambda_{hor})^2}{3 \lambda_{hor} + 4}},$$

which in turn converges to 1 both for $\lambda_{hor} \longrightarrow 0$ and for $\lambda_{hor} \longrightarrow \infty$, using l'Hôpital's rule. In fact, the gain of learning with sequences over learning with sets is insignificant asymptotically as $\alpha \longrightarrow \infty$, for all horizon rates.

VI. CONCLUSION

Understanding the value of temporal data for learning interactions between networked agents or entities can provide valuable insights for a broad spectrum of disciplines. Theorists and practitioners alike in marketing and sociology are interested in learning models of influence among individuals; biologists want to infer how genes interact in gene regulatory networks; financial analysts and economists want to know how firms or sectors interconnect in the economy. The nodes in the network can be taken to be individuals, genes, or firms/sectors, respectively. Of course, the details of a parametric influence model should be tuned according to the needs of its specific application.

In this work, we studied and compared the speed of learning an influence model when learning based on sets, sequences, and times of adoptions, by means of the Kullback-Leibler divergence, for three hypothesis testing problems:

Which of two agents influences a third agent? Is an agent influenced by another agent, or are her decisions self-induced? And is the influence between two agents large or small? We view these three problems as the building blocks for understanding complex patterns of influence that can arise in general networks.

In [25], we also propose a characterization of the speed of learning using the Fisher information. In addition, in [25], [8], we characterize the learnable networks and the sample complexity of learning which edges truly exist, assuming prior knowledge of a "super graph", and we recover reliably the network of influence from real data using a maximum likelihood estimator. In agreement with our findings in this paper, it is shown that when the horizon is small, the sets of decisions provide almost all the necessary information for learning, and there is no value in richer temporal data (sequences or times of decisions).

The discussion in this work remains silent about more general network topologies, and casts the problem of learning the parameters of the influence model as a hypothesis testing problem between only two competing hypotheses. This limitation can be addressed by studying families of more complex networks and characterizing the growth of the number of i.i.d. observations required for learning with respect to the size of the network, especially under different scenarios for the richness of the temporal information in the available data. In addition, one can study the same question assuming different infection and horizon models.

ACKNOWLEDGMENTS

REFERENCES

[1] B. Abrahao, F. Chierichetti, R. Kleinberg, and A. Panconesi. Trace complexity of network inference. In *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 491–499, 2013.

[2] Kareem Amin, Hoda Heidari, and Michael Kearns. Learning from contagion (without timestamps). In *Proceedings of the 31st International Conference on Machine Learning, JMLR W&CP*, volume 32, pages 1845–1853, 2014.

[3] A. Anandkumar, V.Y.F Tan, F. Huang, and A.S. Willsky. High-dimensional structure learning of Ising models: Local separation criterion. *Annals of Statistics*, 40(3):1771–1812, 2012.

[4] L. E. Blume, W. A. Brock, S. N. Durlauf, and Y. M. Ioannides. *Identification of social interactions*, volume 1B of *Handbook of Social Economics*, Chapter 18, pages 853–964. Elsevier B.V., The Netherlands: North-Holland, 2011.

[5] V. Chandrasekaran, P. A. Parrilo, and A. S. Willsky. Latent variable graphical model selection via convex optimization. *Annals of Statistics*, 40(4):1935–1967, 2012.

[6] M.J. Choi, V. Tan, A. Anandkumar, and A. Willsky. Learning latent tree graphical models. *Journal of Machine Learning Research*, 12:1771–1812, 2011.

[7] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley-Interscience, New Jersey, second edition, 2006.

[8] Munther A. Dahleh, John N. Tsitsiklis, and Spyros I. Zoumpoulis. *The Value of Temporal Data for Learning of Influence Networks*, volume 8877 of *Lecture Notes in Computer Science. Web and Internet Economics: 10th International Conference WINE 2014 Proceedings*, pages 322–323. Springer, 2014.

[9] Hadi Daneshmand, Manuel Gomez-Rodriguez, Le Song, and Bernhard Schoelkopf. Estimating diffusion network structures: Recovery conditions, sample complexity & soft-thresholding algorithm. In *Proceedings of the 31st International Conference on Machine Learning, JMLR W&CP*, volume 32, pages 793–801, 2014.

[10] N. Du, L. Song, A. Smola, and M. Yuan. Learning networks of heterogeneous influence. *Advances in Neural Information Processing Systems*, 25, 2012.

[11] M. Gomez-Rodriguez, D. Balduzzi, and B. Schölkopf. Uncovering the temporal dynamics of diffusion networks. In *Proceedings of the 28th International Conference on Machine Learning*, 2011.

[12] M. Gomez-Rodriguez, J. Leskovec, and A. Krause. Inferring networks of diffusion and influence. In *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1019–1028, 2010.

[13] V. Gripon and M. Rabbat. Reconstructing a graph from path traces. In *Proceedings of 2013 IEEE International Symposium on Information Theory*, 2013.

[14] D. Kempe, J. Kleinberg, and E. Tardos. Maximizing the spread of influence through a social network. In *Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 137–146, 2003.

[15] J. O. Kephart and S. R. White. Directed-graph epidemiological models of computer viruses. In *Proceedings of IEEE Symposium on Security and Privacy*, pages 343–359, 1991.

[16] M. Kolar, L. Song, A. Ahmed, and E.P. Xing. Estimating time-varying networks. *Annals of Applied Statistics*, 4(1):94–123, 2010.

[17] K. Lerman and R. Ghosh. Information contagion: An empirical study of the spread of news on Digg and Twitter social networks. In *Proceedings of the International AAAI Conference on Weblogs and Social Media*, 2010.

[18] A. C. Lozano and V. Sindhwani. Block variable selection in multivariate regression and high-dimensional causal inference. *Advances in Neural Information Processing Systems*, 23, 2010.

[19] C. Manski. Identification of endogenous social effects: The reflection problem. *Review of Economic Studies*, 60:531–542, 1993.

[20] D. Materassi and M.V. Salapaka. On the problem of reconstructing an unknown topology via locality properties of the Wiener filter. *IEEE Transactions on Automatic Control*, 57(7):1765–1777, 2012.

[21] S. Myers and J. Leskovec. On the convexity of latent social network inference. *Advances in Neural Information Processing Systems*, 23, 2010.

[22] P. Netrapalli and S. Sanghavi. Finding the graph of epidemic cascades. In *Proceedings of the 12th ACM SIGMETRICS/PERFORMANCE Joint International Conference on Measurement and Modeling of Computer Systems*, pages 211–222, 2012.

[23] D. Shah and T. Zaman. Rumors in a network: Who's the culprit? *IEEE Transactions on Information Theory*, 57:5163–5181, 2011.

[24] Z. Zhou, R. Bandari, J. Kong, H. Qian, and V. Roychowdhury. Information resonance on Twitter: Watching Iran. In *Proceedings of the First Workshop on Social Media Analytics, ACM*, pages 123–131, 2010.

[25] Spyridon Ilias Zoumpoulis. *Networks, Decisions, and Outcomes: Coordination with Local Information and the Value of Temporal Data for Learning Influence Networks*. Ph.D. dissertation, Massachusetts Institute of Technology, Department of Electrical Engineering and Computer Science, June 2014.