The Value of Temporal Data for Learning of Influence Networks

Munther Dahleh
Laboratory for Information and Decision Systems
MIT

John N. Tsitsiklis
Laboratory for Information and Decision Systems
MIT

Spyros I. Zoumpoulis
Laboratory for Information and Decision Systems
MIT
szoumpou@mit.edu

Working Paper
May 19, 2013

Abstract

We infer local influence relations between networked entities from data on outcomes and assess the value of temporal data by characterizing the speed of learning 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. We propose a parametric model of influence which captures directed pairwise interactions, formulate different variations of the learning problem, and provide theoretical guarantees for correct learning based on sets, sequences, and times.

Our results characterize the sample complexity of learning for the three cases. The asymptotic gain of having access to richer temporal data for the speed of learning is thus quantified in terms of the gap between the derived asymptotic requirements under different data modes. We also evaluate the practical value of learning with richer temporal data, by comparing learning with sets, sequences, and times given actual experimental data. Experiments on both synthetic and real data quantify the improvement due to the availability of richer temporal data, and show that our maximum likelihood methodology recovers the underlying network well.

1 Introduction

Can we infer who influences whom in a network of interacting agents or entities based on timed data of their actions/decisions? How does our inference potential change if, instead of times of actions, we have only access to sequences of agent actions, or just the set of agents who undertake an action?

Consumers adopting a new product; an epidemic spreading across a population; a post trending across Internet blogs; a sovereign debt crisis hitting several countries; a cellular process during which the expression of a gene affects the expression of other genes; all of these are temporal processes governed by local interactions of networked entities, which influence one another. Oftentimes, the outcomes of such processes are observable, possibly with time stamps, yet the underlying network of local interactions is hidden.

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 [1, 2, 3]. 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., [4, 5, 6, 7].

Main questions and contributions. How valuable is the knowledge of time stamps for the inference of the underlying network structure? What is the gain due to having access to increasingly informative temporal data, in terms of the speed of learning? The overarching theme of our work is to quantify the gain in speed of learning, due to having access to richer temporal information. We seek to compare the speed of 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. This is, to the best of our knowledge, a comparison that has not been studied before.

We propose a parametric model of influence which captures directed pairwise interactions, formulate different variants of the learning problem, and provide theoretical guarantees for correct learning with sets, sequences, and times. Our results characterize the sufficient and necessary scaling of i.i.d. samples required for correct learning. The asymptotic gain of having access to richer temporal data à propos of the speed of learning is thus quantified in terms of the gap between the derived asymptotic requirements under different data modes. We also evaluate learning with sets, sequences, and times *with experiments*. Given data on outcomes, we learn the parametric influence model by maximum likelihood estimation. Learning with data of richer temporal detail is shown to be faster and more accurate on synthetic data. On both synthetic and real data, our methodology is shown to recover the underlying network structure well.

Other related literature. 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., [8, 9]) or not (e.g., [10]). Instead of proposing and learning a general graphical model, we instead 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.

Of relevance is also [11], in which knowledge of the graph and of the set of infected nodes are both used to infer the original source of an infection. In contrast, and somewhat conversely, we use knowledge of the set, order, or times of infections to infer the graph.

Last, economists have addressed the problem of identification in social interactions (e.g., [12, 13, 14, 15, 16, 17, 18]) focusing on determining aggregate effects of influence in a group; they classify social interactions into an endogenous effect, which is the effect of group members' behaviors on individual behavior; an exogenous (contextual) effect, which is the effect of group members' observable characteristics on individual behavior; and a correlated effect, which is the effect of group members' unobservable characteristics on individual behavior. In sharp contrast, our approach identifies influence at the individual, rather than the aggregate, level.

2 The Influence Model

A product becomes available at time $t = 0$ and each of $n + 1$ 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, disease, 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 of the adoption and infection 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.

3 Theoretical guarantees for learning networks of influence

Problem formulation. A directed graph $G = (\mathcal{V}, \mathcal{E})$ is a priori given and $\lambda_{ij} = 0$ if edge (i, j) is not in $\mathcal{E}$. We provide theoretical guarantees for learning for the case where each edge in $\mathcal{E}$ carries an influence rate of either zero or infinity, casting the decision problem as a hypothesis testing problem. Given a graph G, lower and upper bounds for the number of i.i.d. products required to learn the correct hypothesis can be sought for different variations of the problem, according to the following axes:

- **Learning one edge versus learning all edges:** We pose two decision problems: learning the influence rate λ_{ij} between two specified agents i, j; and learning *all* the influence rates $\lambda_{ij}, i \neq j$.
- **Different prior knowledge over the hypotheses:** We study this question in the Bayesian setting of assuming a prior on the hypotheses, in the worst case over the hypotheses, as well as in the setting in which we know how many edges carry infinite influence rate.
- **Data of different temporal detail:** We characterize the growth of the minimum number of i.i.d. products required for learning with respect to the number of agents n, when the available data provides information on sets, sequences, or times of adoptions.
- **Different scaling of the horizon rate with respect to the idiosyncratic rates:** We consider different scalings of λ_{hor} with respect to the idiosyncratic rates $\lambda_1, \ldots, \lambda_n$.

We employ the proposed program for the star topology, shown in Figure 1. The star topology is one of the simplest non-trivial topologies, and is illustrative of the difference in the sample complexity between learning scenarios with information of different temporal detail.

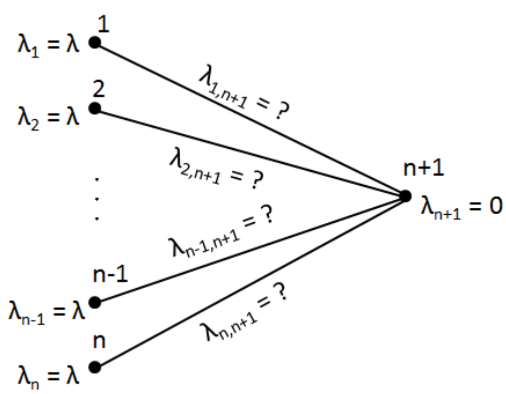

Figure 1: The hypothesis testing problem: what influence does each link carry to the star agent $(n + 1)$: infinite or zero?

In order to simplify the analysis and derive crisp and insightful results, we consider the hypothesis testing problem in which each of n agents influence agent $n + 1$ either with rate zero or infinity. Each of agents $1, \ldots, n$ adopts with rate $\lambda > 0$, while agent $n + 1$ does not adopt unless she is triggered to. There is no influence from agent $n + 1$ to any of the agents $1, \ldots, n$, or from any of the agents $1, \ldots, n$ to any of the agents $1, \ldots, n$. We compare the case when the available data provides information on sets of adopters, to the case when the available data provides information on the sequence of adopters[1].

[1] For this particular formulation, knowledge of times of adoptions would not induce a gain over knowledge of sequences because of our assumption that the influence rates are either zero or infinity, and $\lambda_{n+1} = 0$.

3.1 The Bayesian setting

In the Bayesian setting, we assume that the influence rate on each link is infinite, with probability p, and zero, with probability $1-p$, and that the selection of the rate for each link is independent of the selection for other links.

3.1.1 $p = \frac{1}{2}$

We assume that the influence rate on each link can be zero or infinity with equal probability. Table 1 summarizes the results on the necessary and sufficient number of i.i.d. products for learning.

Table 1: Matching lower and upper bounds for the minimum number of i.i.d. products required to learn the influence model in terms of n, in the Bayesian setting when $p = \frac{1}{2}$, for the two cases of learning the influence between one agent and the star agent and of learning the influence between all agents and the star agent, and for the two cases of learning based on sets of adoptions or sequences of adoptions.

	$\lambda_{hor} = \lambda$		$\lambda_{hor} = n\lambda$	
	Sets	Sequences	Sets	Sequences
Learn one	$\Theta(n^2)$	$\Theta(n)$	$\Theta(n)$	$\Theta(n)$
Learn all	$\Theta(n^2 \log n)$	$\Theta(n \log n)$	$\Theta(n \log n)$	$\Theta(n \log n)$

For example, for the case when $\lambda_{hor} = \lambda$, we have the following sample complexity results[2].

Proposition 1. *To ensure correct learning of $\lambda_{1,n+1}, \ldots, \lambda_{n,n+1}$ with probability $1-\delta$ based on sets of adopting agents, it is sufficient for the number of i.i.d. products to be $O(n^2 \log n)$, and necessary for the number of i.i.d. products to be $\Omega(n^2 \log n)$. To ensure correct learning of $\lambda_{1,n+1}, \ldots, \lambda_{n,n+1}$ with probability $1-\delta$ based on sequences of adoptions, it is sufficient for the number of i.i.d. products to be $O(n \log n)$, and necessary for the number of i.i.d. products to be $\Omega(n \log n)$.*

Proof. For brevity, we only show the second half of Proposition 1, which is illustrative of the reasoning used to prove the rest of our results.

To show the upper bound, consider the following estimator: after k products, decide $\hat{\lambda}_{i,n+1} = \infty$ if and only if there exists a product such that agent i adopts and agent $n+1$ adopts immediately after (and decide $\hat{\lambda}_{i,n+1} = 0$ otherwise). Using the union bound, we relate the probability of error in learning all of $\lambda_{1,n+1}, \ldots, \lambda_{n,n+1}$ to the probability of error in learning $\lambda_{1,n+1}$:

$$\begin{aligned}
\mathbb{P}(\text{error}) &\leq n \cdot \mathbb{P}\left(\hat{\lambda}_{1,n+1} = 0 \mid \lambda_{1,n+1} = \infty\right) \mathbb{P}(\lambda_{1,n+1} = \infty) \\
&= n \cdot \frac{1}{2} \sum_{m=0}^{n-1} \left(1 - \frac{\lambda}{m\lambda + \lambda + \lambda}\right)^k \binom{n-1}{m} \left(\frac{1}{2}\right)^{n-1} \\
&\leq n \cdot \frac{1}{2} \left(\frac{n}{n+1}\right)^k .
\end{aligned}$$

To ensure accurate estimates with probability at least $1-\delta$, for given $\delta \in (0,1)$, it suffices that $k \geq \frac{\log \frac{n}{2\delta}}{\log \frac{n+1}{n}} = O(n \log n)$.

To prove the lower bound, we show that if k is small, there is a high probability event, where it is not clear what a good estimator should decide. No matter how this event is split between the competing hypotheses, the probability of error will be large.

We use X_i to denote the outcome of product i. Having fixed agent j, we say the outcome X_i of product i is in BAD_i^j if one of the following happens: agent j adopts, but agent $n+1$ adopts before

[2]For the sake of brevity, we refrain from providing propositions or proofs for the rest of our results, which are compactly stated in the tables.

her; agent j does not adopt. We say that the outcome $X_1, \ldots, X_k$ is in BAD^j if $X_i \in \text{BAD}_i^j$ for all products $i = 1, \ldots, k$.

We are interested in the probability that for some agent j, it is the case that $X_1, \ldots, X_k \in \text{BAD}^j$. We define A to be the event that each of the agents $1, \ldots, n$ adopts some product before all (other) agents $1, \ldots, n$ with links of rate infinity to agent $n+1$ adopt that product. We define B to be the event that all agents with links of rate infinity to agent $n+1$ adopt some product first among other agents with links of rate infinity to agent $n+1$. Then, we can write

$$\begin{aligned} \mathbb{P}\left(\exists j : (X_1, \ldots, X_k) \in \text{BAD}^j\right) &= 1 - \mathbb{P}(A) \\ &\geq 1 - \mathbb{P}(B). \end{aligned}$$

Let random variable S be the number of i.i.d. products until event A occurs. Let random variable T be the number of i.i.d. products to obtain event B. Then $S \geq T$. The calculation of the expectation of T is similar to the calculation for the coupon collector's problem, after conditioning on the subset of agents $\mathcal{L} \subseteq \{1, \ldots, n\}$ whose influence rate on agent $n+1$ is infinite:

$$\begin{aligned} \mathbb{E}[T] &= \sum_{\mathcal{L} \subseteq \{1,\ldots,n\}} \mathbb{P}(\mathcal{L}) \mathbb{E}[T \mid \mathcal{L}] \\ &= \sum_{m=0}^{n} \mathbb{E}[T \mid \mathcal{L}] \binom{n}{m} \left(\frac{1}{2}\right)^n \\ &= \frac{1}{2^n} \sum_{m=0}^{n} \left(\left(\frac{m\lambda}{m\lambda + \lambda}\right)^{-1} + \left(\frac{(m-1)\lambda}{m\lambda + \lambda}\right)^{-1} + \ldots + \left(\frac{\lambda}{m\lambda + \lambda}\right)^{-1} \right) \binom{n}{m} \\ &= \frac{1}{2^n} \sum_{m=0}^{n} \left(\frac{m+1}{m} + \frac{m+1}{m-1} + \ldots + \frac{m+1}{1} \right) \binom{n}{m} \\ &= \frac{1}{2^n} \sum_{m=0}^{n} (m+1) H_m \binom{n}{m} \\ &= \Omega(n \log n), \end{aligned}$$

where H_m is the mth harmonic number, i.e., $H_m = \sum_{k=1}^{m} \frac{1}{k}$ (and we define $H_0 = 0$), and where the last step follows by Jensen's inequality.

A similar calculation for the variance yields $\text{var}(T) \leq 2(n+1)^2$. By Chebyshev's inequality,

$$\mathbb{P}\left(|T - \mathbb{E}[T]| \geq c(n+1)\right) \leq \frac{2}{c^2},$$

which establishes the $\Omega(n \log n)$ lower bound for the number of products k. □

3.1.2 $p = \frac{1}{n}$

We assume that the influence rate on each link can be infinite with probability $p = \frac{1}{n}$. (In this case, the expected number of agents who can influence agent $n+1$ is $\Theta(1)$.) Table 2 summarizes the results on the necessary and sufficient number of i.i.d. products for learning.

Table 2: Matching lower and upper bounds for the minimum number of i.i.d. products required to learn the influence model, in terms of n, in the Bayesian setting when $p = \frac{1}{n}$, when learning the influence between all agents and the star agent, for the two cases of learning based on sets of adoptions or sequences of adoptions. Notice that no products are needed to learn just one influence rate; an estimator can just guess that $\lambda_{i,n+1} = 0$.

	$\lambda_{hor} = \lambda$		$\lambda_{hor} = n\lambda$	
	Sets	Sequences	Sets	Sequences
Learn all	$\Theta(\log n)$	$\Theta(1)$	$\Theta(n)$	$\Theta(n)$

3.2 The worst-case setting

In the worst-case setting, we assume that each of the influence rates $\lambda_{1,n+1}, \ldots, \lambda_{n,n+1}$ can be either zero or infinity, but we assume no prior over the hypotheses. We provide upper and lower bounds for the minimum number of i.i.d. products required to learn the correct hypothesis assuming that the influence rates on the links are such that the minimum number of i.i.d. products required for learning is maximized (the worst possible). Table 3 summarizes the results on the necessary and sufficient number of i.i.d. products for learning.

Table 3: Matching lower and upper bounds for the minimum number of i.i.d. products required to learn the influence model in terms of n in the worst-case setting, for the two cases of learning the influence between one agent and the star agent and of learning the influence between all agents and the star agent, and for the two cases of learning based on sets of adoptions or sequences of adoptions.

	$\lambda_{hor} = \lambda$		$\lambda_{hor} = n\lambda$	
	Sets	Sequences	Sets	Sequences
Learn one	$\Theta(n^2)$	$\Theta(n)$	$\Theta(n)$	$\Theta(n)$
Learn all	$\Theta(n^2 \log n)$	$\Theta(n \log n)$	$\Theta(n \log n)$	$\Theta(n \log n)$

3.3 The worst-case setting with known scaling of agents with influence rate infinity to $n + 1$

We denote the number of agents with influence rate infinity to agent $n+1$ by ℓ. Table 4 summarizes the results on the necessary and sufficient number of i.i.d. products for learning.

Table 4: Matching lower and upper bounds for the minimum number of i.i.d. products required to learn the influence model in terms of n in the worst-case setting when the scaling of agents ℓ with influence rate infinity to agent $n + 1$ is known, for the two cases of learning based on sets of adoptions or sequences of adoptions.

	$\lambda_{hor} = \lambda$		$\lambda_{hor} = n\lambda$	
	Sets	Sequences	Sets	Sequences
$\ell = 1$	$\Theta(\log n)$	$\Theta(1)$	$\Theta(n)$	$\Theta(n)$
$\ell = \alpha n, \alpha \in (0, 1)$	$\Theta(n^2 \log n)$	$\Theta(n \log n)$	$\Theta(n \log n)$	$\Theta(n \log n)$
$\ell = n - 1$	$\Theta(n^2)$	$\Theta(n)$	$\Theta(n)$	$\Theta(n)$

3.4 Discussion

We characterized the scaling of the number of samples required for learning with sets and sequences, thus theoretically quantifying the gain of learning with sequences over learning with sets in regard to the speed of learning: depending on the setting, learning with sets can take a multiplicative factor of $\Theta(n)$ more samples than learning with sequences, when the horizon rate is moderate (i.e., as large as the idiosyncratic rates). With much smaller horizon, learning with sequences has no gain asymptotically over learning with mere sets, across all the settings we study; the sets of adoptions provide asymptotically all the information pertinent to learning that sequences provide.

4 Experiments

Given data on adoptions, we estimate the idiosyncratic rates of adoption as well as the influence rates (i.e., the network structure) using maximum likelihood estimation. We consider three cases of different temporal detail in the available data (sets, sequences, times), which correspond to different calculations of the log likelihood of the data. Unlike the previous section, we do not restrict the λ_{ij} to be zero or infinite.

We evaluate our methodology on both synthetic and real data, showing that it recovers the network structure well.

4.1 Synthetic data

We generate a directed network according to the Erdős-Rényi model and assign some arbitrary influence rates λ_{ij}, idiosyncratic rates λ_i, and the horizon rate λ_{hor}. We generate data on adoptions of a number of products using the generative model of Section 2, and learn the parameters using maximum likelihood estimation for an increasing sequence of sets of products.

Sets vs. sequences vs. times. Figure 2 showcases the recovery of two different sparse network structures on a network of five agents. We plot the ℓ_1 estimation error, calculated on both the influence rates and the idiosyncratic rates, for different number of products, for the cases of learning with sets, sequences, and times of adoptions. The sparsity pattern is in general recovered more

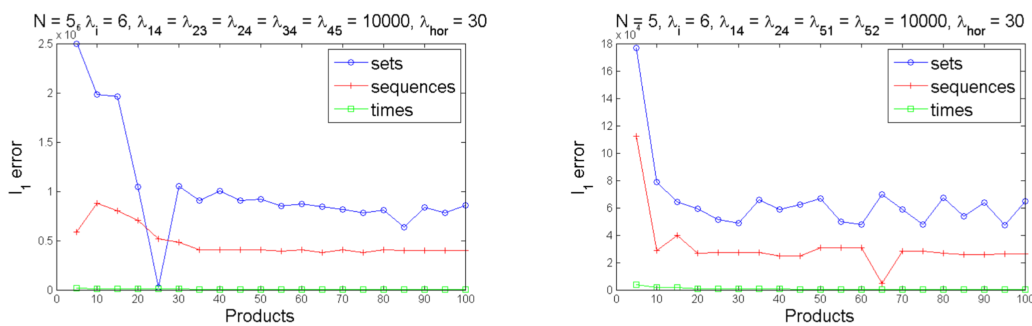

Figure 2: Learning a network of five agents with five high influence rates and all other influence rates zero (left) and four high influence rates and all other influence rates zero (right). We plot the ℓ_1 estimation error calculated on both the influence rates and the idiosyncratic rates. Learning is more accurate and faster with times than with sequences, and with sequences than with sets.

accurately with times than with sequences, and with sequences than with sets; learning with times actually gives accurate estimates of the network influence parameters themselves.

Sequences vs. times. Optimizing the log likelihood function when learning with sets is computationally heavy, as the likelihood term for a specific set is the sum over the likelihood terms for all possible permutations (i.e., orders) associated with the set. We therefore present experiments on networks of larger scale only for learning with sequences and times. Figure 3 showcases the recovery of sparse network structures on networks of 50 and 100 agents. We plot the ℓ_1 estimation error, calculated on both the influence rates and the idiosyncratic rates, for an increasing sequence of products, for the cases of learning with sequences and times of adoptions. Timing information yields significantly better learning.

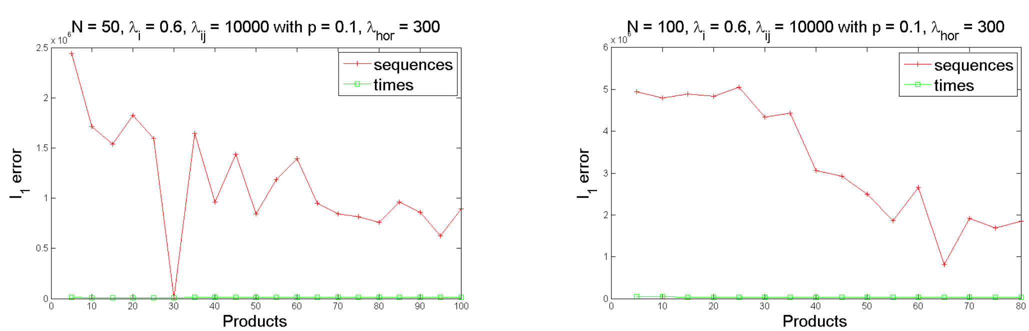

Figure 3: Learning a network of 50 (left) and 100 (right) agents, generated from an Erdős-Rényi model with probability $p = 0.1$ that each directed edge is present. We plot the ℓ_1 estimation error calculated on both the influence rates and the idiosyncratic rates. Learning is more accurate and faster with times than with sequences.

Learning under model mismatch. We have generated data on adoptions using the generative model of Section 2 but adding noise to the realizations of the times of adoptions; this affects also the sets and the sequences of adoptions. Again, we were able to recover the underlying sparsity pattern well, with timing information yielding significantly better learning.

4.2 Real data

We employ our methodology on a mobile applications dataset [19]. The available data captures installations of mobile apps by 55 users during the experimental period of four months, as well as call logs, bluetooth hits, and networks of declared affiliation and friendship among the participants.

The inferred influence rates are highly correlated with the realized communication networks, providing evidence for the soundness of the proposed inference methodology. Table 5 shows the percentage of the edges with top joint influence (i.e., highest sum of inferred $\lambda_{ij} + \lambda_{ji}$, as detected using information on sequences[3] of adoptions) for which either calls, or bluetooth hits, or affiliation, or friendship were reported. We note that some communication edge (calls, bluetooth hits, affiliation, or friendship) exists between two randomly selected nodes with probability 0.8148, which is less than the percentage corresponding to the $20, 50, 100, 200, 400$ edges that carry the highest inferred influence.

Table 5: There is higher probability of communication (calls, bluetooth hits, affiliation, or friendship) in the edges where we detect influence using sequences. A communication edge exists between two randomly selected nodes in the dataset with probability 0.8148.

Out of top	20	joint influence edges, communication happened in	100%
	50		92%
	100		84%
	200		84.50%
	400		81.50%
	800		80.75%

The correlation coefficient between the observations of calls and inferred (joint) influence (using information on sequences of adoptions) per edge is 0.3381. The positive correlation between calls and (joint) influence inferred from sequences is visualized in Figure 4.

[3] Our model is not appropriate for using with information on times of adoptions on this dataset, as the start and end of horizon are not provided. Indeed, when inferred using times, the influence along edges is less correlated with the realized communication data than is the influence inferred using sequences.

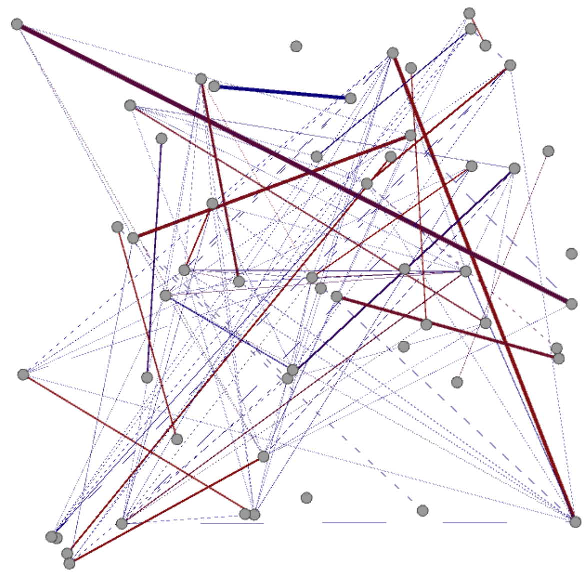

Figure 4: Concurrent visualization of the realized network of calls (the color of each edge denotes the number of calls between the users: closer to blue for lower number of calls, closer to red for higher number of calls) and the inferred network of influence using information on sequences of adoptions (the thickness of edge $i - j$ is proportional to the sum of the inferred influence rates $\lambda_{ij} + \lambda_{ji}$). We observe that edges with higher number of calls (red) are more likely to carry higher influence (thick).

References

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

[2] D. Materassi and M. Salapaka, "On the problem of reconstructing an unknown topology via locality properties of the Wiener filter," *IEEE Transactions on Automatic Control*, vol. 57, no. 7, pp. 1765–1777, 2012.

[3] M. Kolar, L. Song, A. Ahmed, and E. Xing, "Estimating time-varying networks," *Annals of Applied Statistics*, vol. 4, no. 1, pp. 94–123, 2010.

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

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

[6] M. Gomez-Rodriguez, J. Leskovec, and A. Krause, "Inferring networks of diffusion and influence," *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 1019–1028, 2010.

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

[8] V. Chandrasekaran, P. A. Parrilo, and A. S. Willsky, "Latent variable graphical model selection via convex optimization," *Annals of Statistics*, vol. 40, no. 4, pp. 1935–1967, 2012.

[9] M. Choi, V. Tan, A. Anandkumar, and A. Willsky, "Learning latent tree graphical models," *Journal of Machine Learning Research*, vol. 12, pp. 1771–1812, 2011.

[10] A. Anandkumar, V. Tan, F. Huang, and A. Willsky, "High-dimensional structure learning of ising models: Local separation criterion," *Annals of Statistics*, vol. 40, no. 3, pp. 1771–1812, 2012.

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

[12] C. Manski, "Identification of binary choice models," *Journal of the American Statistical Association*, vol. 83, no. 403, pp. 729–738, 1988.

[13] ——, "Identification of endogenous social effects: the reflection problem," *Review of Economic Studies*, vol. 60, pp. 531–542, 1993.

[14] W. A. Brock and S. N. Durlauf, *Interaction-based models*, ser. Handbook of Econometrics, Amsterdam, North-Holland, 2001, vol. 5, pp. 3297–3380.

[15] ——, *Multinomial choice with social interactions*, ser. The Economy as an Evolving Complex System. Oxford and New York: Oxford University Press, 2006, vol. 3, pp. 175–206.

[16] Y. Bramoullé, H. Djebbari, and B. Fortin, "Identification of peer effects through social networks," *Journal of Econometrics*, vol. 150, no. 1, pp. 41–55, 2009.

[17] S. N. Durlauf and Y. M. Ioannides, "Social interactions," *Annual Review of Economics*, vol. 2, pp. 451–478, Sep. 2010.

[18] L. E. Blume, W. A. Brock, S. N. Durlauf, and Y. M. Ioannides, *Identification of Social Interactions*, ser. Handbook of Social Economics. The Netherlands: North-Holland: Elsevier B.V., 2011, vol. 1B, chapter 18, pp. 853–964.

[19] W. Pan, N. Aharony, and A. S. Pentland, "Composite social network for predicting mobile apps installation," in *Association for the Advancement of Artificial Intelligence - 25th Conference on Artificial Intelligence*, 2011.