Learning network structure via Hawkes processes

Michael G. Moore and Mark A. Davenport
School of Electrical and Computer Engineering
Georgia Institute of Technology
Email: {mmoore90,mdav}@gatech.edu

I. THE NETWORK LEARNING PROBLEM

Many of the most challenging problems in modern signal processing and machine learning involve the study of complex networks. Networks arise in a wide variety of applications, ranging from the analysis of traditional communication networks and social networks to problems in genomics and fMRI. In all of these settings, one of the most fundamental questions involves how to discover the relationships between different nodes in a network. In some applications, this information is (seemingly) readily available. For example, in many social networks, the relationships for a particular user can be immediately identified by looking at the user's "friends." However, in many cases the links are far less clear. How can we learn this information when it is not readily available to us? Moreover, even if we have some estimate of the network, how can we determine which relationships are superficial and which ones are truly meaningful?

In many of the applications mentioned above, it can be difficult to directly observe the types of interactions between nodes that would lead to the most direct approach for estimating the network structure. For example, this might be because of technical constraints or privacy concerns in the context of a communication or social network, or because of limited measurement ability in many biological applications. However, in these applications it may still be possible to observe the activity of a particular node, even when the interaction is ambiguous. For example, we may be able to observe a particular user transmitting in a communication network without necessarily being able to determine with whom they are communicating. In this and many other applications, the only information that we can observe regarding the network structure is the timing of events at various nodes in the network. We aim to quantify when it is possible to accurately recover the structure of a network from this kind of simple information about the co-occurrence of events. We will see that, under natural assumptions on the network structure and the number of observations, reliable recovery is indeed possible.

II. HAWKES PROCESSES

Our approach to addressing this problem builds on a probabilistic model for the timing of events. Event times are modeled using a relatively simple point process, the *Hawkes process*, which differs from a Poisson process in that the rate is a function of the process history [1, 2]. This dependence allows it to to capture the action/reaction behavior prevalent in such networks [3–7]. The rate function for node i of N in a multidimensional Hawkes process is given by

$$\lambda_i(t) = \mu_i + \sum_{j=1}^{N} A_{ij} \sum_{\ell=1}^{K^{(j)}} \gamma\left(t - t_\ell^{(j)}\right), \quad (1)$$

where μ_i is a base rate, A_{ij} controls how node i reacts to events at node j, $t_\ell^{(j)}$ is the time of the ℓth of $K^{(j)}$ events at node j, and $\gamma(t)$ is a causal, nonnegative, integrable function that dictates the temporal structure of reactions to events.

An observation of a Hawkes process over the interval $[0, T]$ has a negative log-likelihood given by

$$\mathcal{L}(\boldsymbol{\mu}, \boldsymbol{A}) = \sum_{i=1}^{N} \left(\int_0^T \lambda_i(t) dt - \sum_{k=1}^{K^{(i)}} \log \lambda_i\left(t_k^{(i)}\right) \right), \quad (2)$$

which can be optimized over $\boldsymbol{\mu}$ and $\boldsymbol{A}$ to obtain a maximum-likelihood estimate of the parameters. The matrix $\boldsymbol{A}$ (sometimes called the *infectivity matrix*) can be interpreted as the adjacency matrix of our network, where the magnitude of each entry is proportional to the strength of the corresponding relationships. In practice, we often expect $\boldsymbol{A}$ to be highly structured. For example, in many networks we expect the number of relationships to be relatively small, in which case the matrix $\boldsymbol{A}$ would be sparse. In so-called "small-world networks" $\boldsymbol{A}$ might consist of the combination of sparse and low-rank components [4]. Such structure can be imposed via simple penalized maximum-likelihood estimation procedures.

III. RELIABLE RECOVERY

A key question regarding our approach concerns how many observations are necessary for reliable recovery of the network structure. Here we will consider the case where the network (and $\boldsymbol{A}$) is r-sparse (i.e., $\boldsymbol{A}$ has only r nonzeros). To infer a relationship, we expect that we would need to observe it in action. Assuming that interactions are initiated independently between related nodes, by analogy to the well-known *coupon collector* problem we expect to need at least $O(r \log r)$ events to ensure that all relationships are expressed so that they may be inferred.

To verify this, we provide some simulations relating network recovery to the number of relationships and observed events. We set nonzero values of $\boldsymbol{A}$ to be (suitably scaled) Bernoulli random variables. We examine the error rate in terms of the support recovery of $\boldsymbol{A}$ as a function of the number of events we observe compared to the number of relationships in the network.
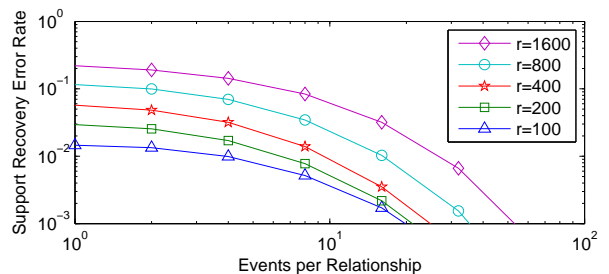

Fig. 1. Support recovery rates as a function of the number of events observed per relationship. Each line denotes a different number of relationships (nonzeros) in the 100×100 infectivity matrix.

Figure 1 suggests that the number of observations needed per relationship to achieve a certain error rate is roughly $O(r \operatorname{polylog} r)$, which coincides with our intuition. In practice, we observe that (for moderate network sizes), it is sufficient to observe an average of roughly 10–30 events per relationship. We also note, however, that matrices $\boldsymbol{A}$ with large dynamic ranges (the example here is binary) require more observations to guarantee support recovery, since weaker relationships are more rarely expressed.

REFERENCES

[1] A. G. Hawkes, "Spectra of some self-exciting and mutually exciting point processes," *Biometrika*, vol. 58, no. 1, pp. 83–90, Apr. 1971.

[2] J. G. Rasmussen, *Temporal point processes: the conditional intensity function*. January, 2011.

[3] C. Blundell, J. Beck, and K. A. Heller, "Modelling reciprocating relationships with hawkes processes," in *Advances in Neural Information Processing Systems 25*, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2012, pp. 2600–2608.

[4] K. Zhou, H. Zha, and L. Song, "Learning social infectivity in sparse low-rank networks using multi-dimensional hawkes processes," in *Proceedings of the Sixteenth International Conference on Artificial Intelligence and Statistics*, 2013, pp. 641–649.

[5] A. Stomakhin, M. B. Short, and A. L. Bertozzi, "Reconstruction of missing data in social networks based on temporal patterns of interactions," *Inverse Problems*, vol. 27, no. 11, p. 115013, Nov. 2011.

[6] M. Farajtabar, N. Du, M. Gomez-Rodriguez, I. Valera, H. Zha, and L. Song, "Shaping social activity by incentivizing users," in *Advances in Neural Information Processing Systems 27*, Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2014, pp. 2474–2482.

[7] P. Reynaud-Bouret, V. Rivoirard, and C. Tuleau-Malot, "Inference of functional connectivity in neurosciences via hawkes processes," in *1st IEEE Global Conference on Signal and Information Processing*, Austin, United States, Dec. 2013.