# SPARSE TOPOLOGY IDENTIFICATION FOR POINT PROCESS NETWORKS

Syed Ahmed Pasha

Department of Electrical Engineering Air University Islamabad 44000, Pakistan

#### ABSTRACT

Increasingly in the 'big data' era, point process network data is appearing e.g. in social media but also genomics, high frequency finance and neurosicence. And the most basic problem is that of identifying network structure. As with all network system identification problems this is enabled by the use of sparsity. Here, for the first time, we develop a vector  $l_0$  sparsity approach for identification of a network of interacting Hawkes processes. We apply the new method to genomic data on a transcriptional regulatory network in embryonic stem cells and the results are compared with previous work and experimental findings.

*Index Terms*— genomics, network topology, point process, system identification, sparsity

## 1. INTRODUCTION

Networks are ubiquitous but understanding of their dynamics is not. Particularly for network data generated in areas such as the social sciences (social media) [1],[2], biology (genomics)[3], finance (high frequency or high sampling rate data) [4], a fundamental issue is the identification of the network topology. Analog network data currently dominates but point process data is increasingly becoming available in all the areas above.

Network system identification methods based on sparsity, have been developed in a number of literatures; but only some of these treat models that deal with dynamics.

The emphasis in this sparsity dynamics literature has been on  $l_1$  penalised formulations [5],[6],[7],[8] since these typically lead to convex optimization problems. But convexity is no longer guaranteed for point process models which have a non-quadratic likelihood. Thus the  $l_0$  penalty, which is known to produce greater sparsity [9] becomes attractive even though it comes with non-convexity.

In our previous work on point process network identification [10],[11],[12],[13] we have considered only the  $l_1$  case. Here, for the first time, we develop a vector or group  $l_0$  penalized approach. Because of the nonconvexity of the  $l_0$  penalty, the  $l_0$  algorithm has to be rebuilt from the ground up.

The remainder of the paper is organized as follows. In section 2 we briefly review the Hawkes-Laguerre point process model. In section 3 we develop the new algorithm. In section 4 we provide a comparative analysis of some genomic data showing the superiority of the  $l_0$  approach over the  $l_1$ . Section 5 contains the conclusion.

Victor Solo

School of Electrical Engineering & Telecom. University of New South Wales Sydney NSW 2052 Australia

#### 2. THE HAWKES-LAGUERRE MODEL

Consider a network of d dynamic processes represented by nodes. A causal interaction between two nodes is represented by a directed link. The output at each node is a spike train and furthermore observable. Then, the network can be modelled by a multivariate point process  $N_t$  that is observed on the interval  $0 < t \le T$ . The nodal processes are the scalar counting processes  $N_{k,t}$ ,  $k = 1, \dots, d$  where  $N_{k,t} = \#$  events of the k-th process in (0, t].

We assume that No-Simultaneity [14] (or orderliness [15]) holds, that is, in a very fine time increment  $\delta$  with high probability only zero or one event of any type can occur. Then, we can define the vector stochastic intensity

$$P(N_{k,t}^{\delta} = 1 | \mathcal{H}^t) = \mu_{k,t} \delta + o(\delta), k = 1, ..., d$$

where  $N_{k,t}^{\delta} = N_{k,t+\delta} - N_{k,t}$ ,  $\mathcal{H}^t$  is the history of the vector counting process  $N_s, 0 \le s < t$  and  $\lim_{\delta \to 0} \frac{o(\delta)}{\delta} = 0$ .

The output at node k can be modelled as

$$y_{k,t} = \mu_{k,t} + \nu_{k,t}, \ k = 1, ..., d$$

where  $y_{k,t} = \frac{1}{\delta} N_{k,t}^{\delta}$  and  $\nu_{k,t}$  is a martingale increment noise. The Hawkes vector stochastic intensity model [16, 17] is

$$\mu_{k,t} = c_k + \sum_1^d \int_{-\infty}^t g_{kj}(u) dN_{j,t-u}, \qquad (2.1)$$
$$\equiv \mu_t = c + \int_{-\infty}^t G(u) dN_{t-u}$$

where c is the d-vector of background firing rates and G(.) is a  $d \times d$  matrix whose (k, j)-th entry  $g_{kj}(.)$  is the Hawkes impulse response (HIR) of the directed link to node k from node j.

Positivity of the stochastic intensity can be ensured with  $c_k > 0$ ,  $g_{kj}(.) \ge 0$ . Furthermore, stationarity is ensured by requiring that  $\Gamma = \int_0^\infty G(u) du$  have largest eigenvalue strictly less than 1. Then the Hawkes (stationary) rate [16, 17] is  $\mu = (I - \Gamma)^{-1}c$ .

The HIR  $g_{kj}$  can be expanded in Laguerre polynomials, which are crucial on two counts. Firstly they provide a causal basis [18]; secondly the basis elements are positive and this enables positivity of the HIR as follows

$$g_{kj}(u) = \Sigma_1^{p_{kj}} \beta_{kj,l} \phi_{kj,l}(u)$$

with

$$\phi_{kj,l}(u) = \beta_{kj,o} \frac{(u\beta_{kj,o})}{(l-1)!}^{l-1} e^{-\beta_{kj,o}u}$$

This work was supported partly by the Australian Research Council and also under grant NPRU 5233 of the Higher Education Comission of Pakistan.

where  $\beta_{kj,o}$  is the reciprocal of the time constant. This is similar to [19, 20]. But the addition of the (l-1)! in the denominator and the multiplier  $\beta_{kj,o}$  in the numerator ensure

$$\int_0^\infty \phi_{kj,l}(u) du = 1$$

so that we get the simple expression

$$\Gamma = [\Gamma_{kj}]$$

$$= [\int_0^\infty g_{kj}(u) du] = [\sum_l \beta_{kj,l}]$$

Substituting the Laguerre basis expansion in (2.1), the stochastic component due to node j is

$$\int_{-\infty}^{t} g_{kj}(u) dN_{j,t-u} = \Sigma_{1}^{p_{kj}} \beta_{kj,l} \int_{-\infty}^{t} \phi_{kj,l}(u) dN_{j,t-u}$$
$$= \Sigma_{1}^{p_{kj}} \beta_{kj,l} \psi_{kj,l;t}$$

where crucially,  $\psi_{kj,l;t}$  can be precomputed and so can be assumed known.

We thus have a linear (in the parameters) model for the stochastic intensity of node k

$$\mu_{k,t} = c_k + \sum_{j=1}^d \sum_{l=1}^p \beta_{kj,l} \psi_{kj,l;t}, \ k = 1, ..., d$$
(2.2)

where we have taken  $p = \max\{p_{kj}\}_{k=1,\dots,d,j=1,\dots,d}$ .

Note that the Laguerre expansion provides a valid causal basis expansion for any positive  $\beta_{kj,o}$  and so  $\beta_{kj,o}$  can be specified by the user and need not be estimated. Since  $\frac{1}{\beta_{kj,o}}$  is a time constant the choice of  $\beta_{kj,o}$  controls the extent of memory in the system and this interpretation facilitates its choice.

# 3. NETWORK TOPOLOGY IDENTIFICATION

We partition the observation interval  $0 < t \le T$  into fine time bins of width  $\delta$ . Let  $T = n\delta$  and  $t = m\delta$ , m = 0, ..., n - 1. The output at the k-th node is then  $y_{k,m} = y_{k,m\delta}$ . Let  $\mathbf{y}_k$  denote the output vector at node k after subtracting the background rate, i.e.,  $\mathbf{y}_k = (y_{k,0}, y_{k,1}, y_{k,2}, ..., y_{k,n-1})^T - c_k \mathbf{1}$ . Then, the Laguerre model (2.2) yields the linear regression model

$$\mathbf{y}_k = X_k \beta_k + v_k \tag{3.3}$$

where  $\beta_k = (\beta_{k1,1}, ..., \beta_{k1,p}, \beta_{k2,1}, ..., \beta_{kd,p})^T$  and  $X_k = (X_{k1}, ..., X_{kd})$  with

$$X_{kj} = \begin{pmatrix} \psi_{kj,1;0} & \dots & \psi_{kj,p;0} \\ \vdots & \ddots & \vdots \\ \psi_{kj,1;n-1} & \dots & \psi_{kj,p;n-1} \end{pmatrix}$$
  
The parameter vector  $\beta_k$  can be grouped as  $\beta_k = \begin{bmatrix} \beta_{k1} \\ \vdots \\ \beta_{kd} \end{bmatrix}$  where

 $\beta_{kj} = (\beta_{kj,1}, ..., \beta_{kj,p})^T$ . Then, the column-wise partition  $X_{kj}$  corresponds to the  $\beta_{kj}$  group. Note that the number of parameters in each group is equal to p. This is not a necessary condition. If  $p_{kj}$  is the number of parameters in the *j*-th group, then the parameter vector  $\beta_k$  is of length  $\sum_{j=1}^d p_{kj}$  and the  $X_{kj}$  partition has the corresponding number of  $p_{kj}$  columns.

The network topology identification problem under the sparsity constraint can be formulated as a regularized least squares problem.

### 3.1. Penalized Least Squares Problem

The vector  $l_0$ -regularized least squares ( $l_o$ -LS) criterion for grouped variables is

$$J(\beta) = \sum_{k=1}^{d} \left( \|\mathbf{y}_{k} - X_{k}\beta_{k}\|^{2} + \lambda_{k}\sum_{j=1}^{d} I(\|\beta_{kj}\| \neq 0) \right) \quad (3.4)$$

where  $\lambda_k > 0$  is the regularization parameter and  $I(||\beta_{kj}|| \neq 0) = 1$  if  $||\beta_{kj}|| \neq 0$ . It is crucial to note the vector or group nature of the penalty which enables whole node to node links to be removed simultaneously. This is not the same as the scalar  $l_0$  penalty defined as  $\sum_{l=1}^{p} I(\beta_{kj,l} \neq 0)$  which takes values in the interval [0, p] and can ensure that some but not all entries in  $\beta_{kj}$  are zero. Also note that the background rate (hidden in  $y_k$ ) is not regularised.

Note that the system identification problem is further compounded by the non-negativity constraints on  $\beta_k$  and positivity constraints on  $c_k$  for k = 1, ..., d.

The variance of the regression term in (3.4) is of order  $c_k/\delta$ where  $c_k$  is of order the mean (or median) count of the k-th counting process. For dimensional consistency the penalty <u>parameter</u> must have the units of  $c_k/\delta$ . So we can replace the penalty parameter  $\lambda_k$  in (3.4) by  $\lambda_o c_k/\delta$  where  $\lambda_o$  is nondimensional. We then have only to choose one tuning parameter  $\lambda_o$ . This is a crucial feature to make the least squares criterion workable in the non-standard point process setting.

Equation (3.4) is separable in the dimension d which allows the minimization to be performed as d separate optimizations for the criterion

$$J_{k}(\beta_{k}) = \|\mathbf{y}_{k} - X_{k}\beta_{k}\|^{2} + \lambda_{k}\Sigma_{j=1}^{d}I(\|\beta_{kj}\| \neq 0)$$
(3.5)

#### 3.2. Cyclic Descent Minimization

We propose a cyclic descent approach to minimize (3.5). In step 1, we minimize (3.5) with respect to  $\beta_{k1}$  with coefficient groups  $\beta_{kj}, j = 2, ..., d$  fixed at their current value. In step r, we minimize (3.5) with respect to  $\beta_{kr}$  with coefficient groups  $\beta_{kj}, j =$  $\{1, ..., d\} \setminus \{r\}$  fixed and so on. An additional step is also needed to estimate the background rates  $c_k$ .

For the r-th step, we can rewrite (3.5) in the alternative form

$$J_{k}(\beta_{kr}) = \|v_{k,-r} - X_{kr}\beta_{kr}\|^{2} + \lambda_{k}\Sigma_{j\neq r}I_{kj} + \lambda_{k}I_{kr}$$
(3.6)

where  $v_{k,-r} = \mathbf{y}_k - \sum_{j \neq r} X_{kj} \beta_{kj}$  and  $I_{kj} = I(\|\beta_{kj}\| \neq 0)$ .

We assume without loss of generality that  $X_k$  is group orthonormal (ON), that is  $X_{kj}^T X_{kj} = I, j = 1, ..., d$ . If  $X_{kj}$  is not ON we rewrite

$$X_{kj}\beta_{kj} = [X_{kj}(X_{kj}^T X_{kj})^{-\frac{1}{2}}][(X_{kj}^T X_{kj})^{\frac{1}{2}}\beta_{kj}] = \bar{X}_{kj}\bar{\beta}_{kj}$$

where now  $\bar{X}_{kj}$  is ON. Once the optimization is completed we compute  $\beta_{kj} = (X_{kj}^T X_{kj})^{-\frac{1}{2}} \bar{\beta}_{kj}$ .

Continuing, since a multiplication by  $X_{kr}^T$  in the first term in (3.6) preserves the norm, (3.6) can be rewritten as

$$J_{k}(\beta_{kr}) = \|z_{kr} - \beta_{kr}\|^{2} + \lambda_{k}I_{kr}$$
(3.7)

where  $z_{kr} = X_{kr}^T v_{k,-r}$  and the terms that do not depend on  $\beta_{kr}$  have been dropped. The minimizer of (3.7) is [21]

$$\beta_{kr}^{(i)} = z_{kr}^{(i-1)} I(\|z_{kr}^{(i-1)}\| \ge \sqrt{\lambda_k})$$
(3.8)

After step d we obtain  $c_k > 0$  using the multiplicative update

$$c_{k}^{(i)} = c_{k}^{(i-1)} \frac{\delta(\mathbf{1}^{T} e_{k}^{(i-1)})_{+}}{T c_{k}^{(i-1)} + \frac{\lambda_{0}}{2} \Sigma_{j=1}^{d} I_{kj}}$$
(3.9)

where  $e_k^{(i-1)}=v_k^{(i-1)}+c_k^{(i-1)}\mathbf{1}$  and  $(x)_+=\max(\epsilon,x)$  for a positive number  $\epsilon\ll 1.$ 

The  $\beta_k$ -update step (3.8) for r = 1, ..., d and  $c_k$ -update step (3.9) can be solved iteratively until convergence.

#### 3.3. Computational Details

 $X_k, k = 1, ..., d$  are known and need to be computed only once. The system identification involves d separate optimizations which can be done concurrently.

Note that in (3.7),  $z_{kr} = X_{kr}^T v_{k,-r} = X_{kr}^T (v_k + X_{kr}\beta_{kr}) = X_{kr}^T v_k + \beta_{kr} = \gamma_{kr} + \beta_{kr}$ . Then, (3.8) becomes

$$\beta_{kr}^{(i)} = (\gamma_{kr}^{(i-1)} + \beta_{kr}^{(i-1)})I(\|(\gamma_{kr}^{(i-1)} + \beta_{kr}^{(i-1)})\| \ge \sqrt{\lambda_k})$$

For a given k and starting values of  $c_k$ ,  $\beta_k$ , the algorithm proceeds in a cyclic descent fashion to update  $\beta_k$  and  $c_k$ . A stopping criterion such as  $|J_k(\beta_k^{(i)}) - J_k(\beta_k^{(i-1)})| < \alpha$  f or some fixed positive number  $\alpha \ll 1$  can be used.

#### 3.4. Convergence

The analysis of convergence of the algorithm is non-trivial and will be pursued elsewhere. The modification of the algorithm from a classic  $l_1$  penalized least squares problem to one where one paprameter, c, is not penalized means that the whole problem of convergence would have to be revisited from scratch. That will need a whole paper by itself. As indicated in the data analysis section we have not experienced convergence problems.

#### 4. DATA ANALYSIS

We analyze genomic data from a study [22] to identify causal interactions in transcriptional regulatory networks (TRNs) in embryonic stem cells. The data are the co-ordinates of the binding sites of 13 TFs and 2 TRs.

We show a comparison of the  $l_0$ -LS and  $l_1$ -LS [11] algorithms. We confirm using point process data what [9, 21] have already demonstrated using analog data; namely the superiority of  $l_0$ -LS over the  $l_1$ -LS.

The point process data are obtained from chromosome 1 following the approach in [3].

- (a) We set  $\delta = 100$  base pairs.
- (b) Transcription elements with fewer than 100 counts were dropped.

A raster plot of the downstream occurrences of 12 TFs and 1 TR in chromosome 1 is shown in Fig. 1.

We use a standard least squares solution under non-negativity constraints<sup>1</sup> in two ways. Firstly, to initialize the iterations; secondly to scale the  $l_0$ -LS and  $l_1$ -LS criteria by the standard least squares criterion. The scaled criteria are denoted by  $\hat{J}_0$  and  $\hat{J}_1$  respectively and the scaled criterion iterates are denoted by  $\hat{J}^{(i)}$ . The stopping criterion is  $|\hat{J}^{(i)} - \hat{J}^{(i-1)}| < \alpha$  with threshold value  $\alpha = 10^{-4}$ .

In order to compare the two algorithms:

(a) we expand the HIR  $g_{k,j}(.), k, j = 1, ..., d$  in Laguerre polynomials with p = 2 terms;

(b) the  $\beta_{kj,o}$  and  $\lambda_o$  parameters are determined jointly using the Bayesian Information Criterion (BIC).

The BIC heat maps of  $\lambda_o$  vs.  $\beta_{kj,o}$  for  $l_0$ -LS and  $l_1$ -LS are shown in Fig. 2 and Fig. 3 respectively. Using the BIC minimizer in each case the  $\hat{J}_0$  and  $\hat{J}_1$  iterates are shown Fig. 4.

#### 4.1. Discussion of Results

We find that the BIC minimizer for the  $l_0$ -LS method is  $(\lambda_o, \beta_{kj,o}) = (9.5, 100)$  and for the  $l_1$ -LS method is  $(\lambda_o, \beta_{kj,o}) = (57, 80)$ . We can clearly see from Fig. 4 that the  $l_0$ -LS reduces the criterion more than the  $l_1$ -LS. An interesting feature is the oscillatory behaviour before flattening. This occurs when some of the starting values of  $\beta_k$  are zero. With non-zero starting values the scaled criterion drops monotonically before flattening.

The TRN constructed using the  $l_0$ -LS algorithm (shown in Fig. 5) has 62 links ( $\equiv 13 + 2 \times 62 = 137$  parameters). The 5 TFs: Oct4, Zfx, c-Myc, n-Myc and STAT3 do not have self-exciting links which have been omitted for the remaining TFs to avoid clutter. There are 15 bi-directional interactions. These are indicated by undirected links. We find that Fig. 5 shares greater similarity with the network constructed from experimental study [22] compared to the TRN with 99 links ( $\equiv 13 + 2 \times 99 = 211$  parameters) constructed using the  $l_1$ -LS algorithm [10].

[22] found that suz12 did not show strong association with any of the TFs. The links connecting suz12 in Fig. 5 are possibly weak interactions with the TFs. [22] identified 2 clusters with strong association within the clusters. The first cluster includes Nanog, Oct4, Sox2 and STAT3 while the second cluster includes c-Myc, n-Myc, Zfx and E2f1. The experimental finding is consistent with our analysis with one major difference. The study finds strong association Sox $\leftrightarrow$ Oct4 and Sox2 $\rightarrow$ Nanog. Since the study infers interactions in a pairwise manner it is not enough to conclude the existence of a direct link. Our analysis suggests that the interaction is most likely mediated via E2f1 and the clusters are not as isolated as suggested in [22].



Fig. 1. Raster Plot of Downstream Occurrences of 12 TF and 1 TR in Chromosome 1.

#### 5. CONCLUSIONS

In this paper we have developed for the first time a sparse multivariate point process network identification method based on a vector or group  $l_0$ -norm penalty. This is not a minor modification of our previous  $l_1$  based methods; the nonconvexity required a complete redesign of the algorithm. A BIC based tuning parameter selection method was developed. We presented a cyclic descent algorithm that guarantees positivity of the point process stochastic intensity function. We have found the cyclic descent approach robust in the

<sup>&</sup>lt;sup>1</sup>Positivity of the stochastic intensity function is guaranteed with  $\beta_k \ge 0$ and  $c_k > 0$ . Negative entries of  $\beta_k$  are set to zero and a negative  $c_k$  value is set to a small positive number.



**Fig. 2.**  $l_0$ -LS: BIC Heat Map  $\lambda_o$  vs.  $\beta_{kj,o}$ . BIC Minimizer  $(\lambda_o, \beta_{kj,o}) = (9.5, 100)$ .



**Fig. 3.**  $l_1$ -LS: BIC Heat Map  $\lambda_o$  vs.  $\beta_{kj,o}$ . BIC Minimizer  $(\lambda_o, \beta_{kj,o}) = (57, 80)$ .

high-dimensional setting and have demonstrated empirically that our proposed algorithm has good convergence rate. Based on this an efficient method for topology identification of interacting Hawkes processes has been developed. The algorithm was tested on some genomic data relating to mouse embryonic stem cells to reconstruct the transcriptional regulatory network (TRN). A comparison of results with a corresponding sparsity method based on the  $l_1$ -norm has shown the superiority of the proposed  $l_0$  algorithm for constructing a sparser TRN.



**Fig. 4**. Scaled Criterion Iterates of the  $l_0$ -LS and  $l_1$ -LS.



**Fig. 5.** TRN of the Mouse Embryonic Stem Cells Constructed using the  $l_0$ -LS.

#### 6. REFERENCES

- P.O. Perry and P.J. Wolfe, "Point process modeling for directed interaction networks," *Jl. Royal Stat. Soc.: Series B*, vol. 75, pp. 821–849, 2013.
- [2] K. Zhou, H. Zha, and L. Song, "Learning social infectivity in sparse low-rank networks using multi-dimensional Hawkes processes," in *AISTATS*, 2013, pp. 641–649.
- [3] L. Carstensen et al., "Multivariate Hawkes process models of the occurrence of regulatory elements," *BMC Bioinformatics*, vol. 11(1), 2010.
- [4] Y. Ait-Sahalia, J Cacho-Diaz, and RJA Laeven, "Modelling financial contagion using mutually exciting jump processes," Tech. Rep., Dept. Economics, Princeton Univ., 2010.
- [5] A. Bolstad, B.D. Van Veen, and R. Nowak, "Causal network inference via group sparse regularization," *IEEE Trans. Sig. Proc.*, vol. 59, pp. 2628–2641, 2011.
- [6] BM. Sanandaji, TL. Vincent, and MB. Wakin, "Compressive topology identification of interconnected dynamic systems via clustered orthogonal matching pursuit," in *Proc. IEEE Conf. Dec. Control.* IEEE, 2011, pp. 174–180.
- [7] M. Ayazoglu, M. Sznaier, and N. Ozay, "Blind identification of sparse dynamic networks and applications," in *Proc. IEEE Conf. Dec. Control.* IEEE, 2011, pp. 2944–2950.
- [8] S. Haufe, K. Muller, G. Nolte, and N. Kramer, "Sparse causal discovery in multivariate time series," in *NIPS*, 2008, pp. 97– 106.
- [9] AJ Seneviratne and V Solo, "On vector l<sub>0</sub> penalized multivariate regression," in *Proc. IEEE ICASSP*, 2012.
- [10] S.A. Pasha and V. Solo, "Topology identification of dynamic point process networks," in *Proc. IEEE ICASSP*, 2014, pp. 375–378.
- [11] S.A. Pasha and V. Solo, "Topology identification for sparse dynamic point process networks," in *Proc. IEEE CDC*, 2014, pp. 1786–1791.

- [12] S.A. Pasha and V. Solo, "Distributed topology identification for point process dynamic networks," in *Proc. IEEE ICASSP*, 2015, pp. 3681–3685.
- [13] S.A. Pasha and V. Solo, "Distributed topology identification for sparse point process dynamic networks," in *Proc. IEEE CDC*, 2015, pp. 3379–3384.
- [14] V. Solo, "High dimensional point process system identification: PCA and dynamic index models," in *Proc. IEEE CDC*, 2006, pp. 829–833.
- [15] D.J. Daley and D. Vere-Jones, *An Introduction to the Theory* of *Point Processes*, Springer, NY, 1988.
- [16] AG Hawkes, "Spectra of some self-exciting and mutually exciting point processes," *Biometrika*, vol. 58(1), pp. 83–90, 1971.
- [17] A.G. Hawkes, "Point spectra of some mutually exciting point processes," J. Roy. Statist. Soc., vol. 33(3), 1971.
- [18] B Wahlberg, "System identification using Laguerre models," *IEEE Trans. Autom. Control*, vol. 36(5), 1991.
- [19] Y. Ogata and H. Akaike, "On linear intensity models for mixed doubly stochastic Poisson and self-exciting point processes," J. R. Stat. Soc. Ser. B. Stat. Methodol., 1982.
- [20] V. Solo and A. Pasha, "A test for independence between a point process and an analogue signal," *Jl. Time Series Analysis*, vol. 33(5), pp. 824–840, 2012.
- [21] AJ Seneviratne and V Solo, "Topology identification of a sparse dynamic network," in *Proc. IEEE CDC*, 2012.
- [22] X. Chen et al., "Integration of external signaling pathways with the core transcriptional network in embryonic stem cells," *Cell*, vol. 133, pp. 1106–1117, 2008.