NETWORK INFERENCE FROM COMPLEX SYSTEMS STEADY STATES OBSERVATIONS: THEORY AND METHODS

Hoi-To Wai[†]*, Anna Scaglione*[†]*, Baruch Barzel*[‡]*, Amir Leshem*[⋆]

[†]School of ECEE, Arizona State University, Tempe, AZ, USA.
[‡]Dept. of Math., Bar-Ilan University, Israel. [⋆]Faculty of Engg., Bar-Ilan University, Israel.
Emails: {htwai, Anna.Scaglione}@asu.edu, baruchbarzel@gmail.com, leshem.amir2@gmail.com

ABSTRACT

This paper presents new results on network inference from observations of steady state behaviors emerging from perturbations of complex networks dynamics. We focus on the estimation of network and flow parameters using a general regularized inference formulation, which is tackled numerically using the standard technique of alternating optimization. We argue that relying only on the steady states equations removes the requirement of precisely recording transient data, and allows to meaningfully combine data from multiple experiments. To provide theoretical benchmarks we study the relationship between topological and functional characteristics of the system and the divergence between the steady state behavior observed, to give rigorous performance benchmarks. Numerical results are presented on examples with social networks and gene regulatory networks to justify our claims.

***Index Terms*—** network inference, complex network systems, social networks, gene networks, network identifiability

1. INTRODUCTION

Our life is surrounded by complex networks — from gene regulatory networks to social networks, from power networks to transportation networks, etc.. It is not surprising that understanding complex network systems has attracted researchers from various disciplines. Among the related problems a challenging one is the *reverse engineering* of a complex network, by recovering the network dynamics that underpins its behavior.

This paper is concerned with the joint recovery of network and the complex system's dynamics from a small number of observed steady states. The steady states are the *stationary points* of a complex system that are easy to observe or measure. As the complex system may admit only a unique steady state, we consider the natural or synthetic perturbations introduced to the system as a mean to achieve identifiability. In fact, the steady states that emerge after such perturbations can be exploited for the inference process. This approach is akin to the standard system identification method involving probing and sensing in many signal processing and control problems, e.g., [1]. As the number of steady states observed is limited we propose a regularized inference model to be tackled using alternating optimization. The inference model has an interesting structure which can be decomposed naturally for computational speed up using parallel processors. We also address the issue of identifiability by reviewing a few related results. We demonstrate the efficacy of our approach in dynamics for social networks and gene networks. Importantly, we achieve better than state-of-the-art network inference performance using the steady states data alone.

This work is supported by NSF CCF-BSF 1714672.

Our work is related to a large body of prior work on network inference. We focus on the signal processing and statistics literature first. The notable examples include [2] which developed the graphical LASSO (gLASSO) method for *undirected* networks and [3] which generalizes gLASSO to the *temporally* nonlinear dynamics, a.k.a. graph filters [4–6]; also see the related work [7–11]. Note that our approach considers *directed* networks and *spatially* nonlinear dynamics. To this end, the most relevant work can be found in [12] which uses a kernel-based method to infer the directed network topology. However, their approach requires a relatively large amount of data that can be difficult to observe. Similar earlier work based on transient data was in [13, 14]. In addition, heuristics based on machine learning tools have been proposed in [15–17] with methods tailored to the gene network recovery.

2. DYNAMICS OF COMPLEX NETWORK SYSTEMS

We consider a complex network system whose inter-connectivity between its nodes can be described by a directed graph $G = (V, E)$, with $V = [n] := \{1, ..., n\}$ denoting the set of *agents* in the system and $E \subseteq V \times V$ the possible interactions between pairs of agents. The graph is directed, *i.e.,* $(i, j) \in E$ implies that it is possible for agent i to influence agent j but not vice versa. Moreover, we assume $(i, i) \notin E$ such that there is no self loop. We shall use the notion of *agents* as a generalization of the interacting parties in the complex systems — e.g., for social networks, an agent refer to an individual who befriend with others; for gene networks, the name *agent* refers to a gene which is regulated by another gene; for power networks, it is a bus injecting a net positive or negative amount of real power in the system, corresponding to generation and load respectively.

The interactions between agents differ from pairs to pairs. We model this heterogeneity using a weighted adjacency matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$. The matrix is defined on the graph G such that $A_{ij} \neq 0$ if and only if $(j, i) \in E$. Let $x_i(t)$ be the *state* of agent i at time t, and $\dot{x}_i(t) := dx(t)/dt$ [*resp.* $\ddot{x}_i(t) := d^2x(t)/dt^2$] be its first-order [*resp.* second-order] *time* derivative. Denote the complex network system as $\mathcal{C}_0$, its dynamics is described by the nonlinear system of differential equations —

$$f(\dot{x}_i(t), \ddot{x}_i(t)) = g_i(x_i(t)) + \sum_{j=1}^{n} A_{ij} h(x_i(t), x_j(t); \boldsymbol{\theta}_i) \ , \quad (1)$$

for all $i \in [n]$, where $g_i(\cdot)$ is some self-influence function, $\boldsymbol{\theta}_i$ is the parameter vector of the kernel functions $h(x, y; \boldsymbol{\theta}_i)$ that models the pairwise state interaction. The nonlinear function $f(x, y)$ serves as a measure of *stationarity* in the complex system such that $f(x, y) = 0$ if $x = y = 0$; in other words, the system is in *steady state* when the right hand side of (1) is zero for all $i \in [n]$. Formally, we say that $\bar{\boldsymbol{x}}$

is a steady state to the system $\mathcal{C}_0$ if it satisfies:

$$0 = g_i(\bar{x}_i) + \sum_{j=1}^{n} A_{ij} h(\bar{x}_i, \bar{x}_j; \boldsymbol{\theta}_i),\ i = 1, ..., n\ . \tag{2}$$

The *steady states* are the points of interest for many complex network systems, since they correspond the lowest energy states and are thus the stable points of the dynamics. Due to its stability, they are also the behaviors of the network that can be most easily observed.

This paper is concerned with inferring the parameters $\boldsymbol{\Theta} := (\boldsymbol{A}, \{\boldsymbol{\theta}_i\}_{i=1}^n)$ of the complex network system using its steady state data. As we shall reveal later, our approach is parametric such that it relies on fitting the parameters into the model (2). Note that the number of steady states is limited (or is even unique) for a number of complex network systems. This leads to an issue of extreme *rank deficiency* since the array of unknown parameters $\boldsymbol{\Theta}$ lives in a high-dimensional space, *i.e.,* there are $n^2 - n$ unknowns in $\boldsymbol{A}$ alone. To this end, a traditional approach in system identification is to perform a set of *perturbation experiments* in the system.

2.1. Perturbation Experiments

Consider introducing P different *perturbations* to the system $\mathcal{C}_0$. Let $p = 1, ..., P$, the perturbed system $\mathcal{C}_p$ admits a different steady state, denoted as $\boldsymbol{x}^p$, than $\bar{\boldsymbol{x}}$ while the sought parameters $\boldsymbol{\Theta}$ remain unchanged. In this way, we obtain a set of data that can be used to infer $\boldsymbol{\Theta}$ while alleviating the issue with rank deficiency.

In fact, performing perturbation experiments is a common approach for network inference in practice. The following perturbation models will be studied in this paper:

Type I — Injecting New Agents to $\mathcal{C}_0$. For this type of experiment, we introduce S new agents into the complex network system which hold the states of $\boldsymbol{z}^p := (z_j^p)_{j=1}^S$ during the pth experiment. This yields a perturbed system $\mathcal{C}_p$, whose steady state, $\boldsymbol{x}^p$, satisfies:

$$0 = \ell_i^{\text{type}-1}(\boldsymbol{A}, \boldsymbol{B}, \{\boldsymbol{\theta}_i\}_{i=1}^n; \boldsymbol{x}^p),\ i = 1, ..., n\ , \tag{3}$$

where

$$\begin{aligned}&\ell_i^{\text{type}-1}(\boldsymbol{A}, \boldsymbol{B}, \{\boldsymbol{\theta}_i\}_{i=1}^n; \boldsymbol{x}^p) := \\ &g_i(x_i^p) + \sum_{j=1}^{n} A_{ij} h(x_i^p, x_j^p; \boldsymbol{\theta}_i) + \sum_{j=1}^{S} B_{ij} h(x_i^p, z_j^p; \boldsymbol{\theta}_i)\ ,\end{aligned} \tag{4}$$

and B_{ij} is the interaction strength from the jth inserted agent to the ith agent. In control theory, this is similar to adding a control input to the complex network system. We remark that $\boldsymbol{B}$ is unknown in general and it should be estimated alongside with $\boldsymbol{\Theta}$ as a nuisance parameter during the network inference process.

This perturbation model is relevant in the context of opinion dynamics in social networks, where each of the inserted agent is a *stubborn agent* whose aim is to influence the social network. For this case, we also have $\sum_{j=1}^n A_{ij} + \sum_{j=1}^S B_{ij} = 1$ for all i such that the network dynamics is that of an average consensus process. Each of the kth experiment observed correspond to a discussion on a certain topic, and the inserted state $\boldsymbol{z}^k$ are precisely the opinions held by the stubborn agents. Interestingly, for the special case of linear DeGroot opinion dynamics, the observation rank will be upper bounded by the number of stubborn agents S [18, 19].

Type II — Knocking Out Agents in $\mathcal{C}_0$. In this perturbation model, we suppress/knock out an agent at a time in the complex network system.

This type of perturbation experiment is relevant to the common network inference experiments performed on gene regulatory networks (GRNs). Specifically, each agent corresponds to a gene in the GRN and it is possible to *knockdown* an agent through experiment techniques such as mutagenesis to knockout genes [20] and microarray experiments to profile them [21]. It is worth noting that the dataset collected from knockdown experiments are widely available, e.g., [21].

Let $\pi(p) \in \{1, ..., n\}$ be the agent knocked out at the pth experiment[1], the perturbed system $\mathcal{C}_p$ satisfies:

$$0 = \ell_i^{\text{type}-2}(\boldsymbol{A}, \{\boldsymbol{\theta}_i\}_{i=1}^n; \boldsymbol{x}^p),\ i = 1, ...\pi(p) - 1, \pi(p) + 1, ..., n\ , \tag{5}$$

where

$$\begin{aligned}&\ell_i^{\text{type}-2}(\boldsymbol{A}, \{\boldsymbol{\theta}_i\}_{i=1}^n; \boldsymbol{x}^p) := \\ &g_i(x_i^p) + \sum_{j=1, j\neq\pi(p)}^{n} A_{ij} h(x_i^p, x_j^p; \boldsymbol{\theta}_i) + A_{i,\pi(p)} h(x_i^p, 0; \boldsymbol{\theta}_i)\ ,\end{aligned} \tag{6}$$

Note that this is equivalent to imposing a boundary condition for $x_{\pi(p)}(t) = 0$ on the system of differential equations. Interestingly, for certain dynamics and parameters $\boldsymbol{\theta}_i$, it is possible to infer the partial support of $\boldsymbol{A}$ merely by comparing the steady states. This result can be summarized as follows.

Assumption 1 *For all i, we have $g_i(x) = x$ and the kernel function satisfies $h(x, y; \boldsymbol{\theta}) = h(x; \boldsymbol{\theta})$.*

Proposition 1 *[22] Under Assumption 1. Denote $[\boldsymbol{h}(\boldsymbol{x}; \boldsymbol{\theta})]_i := h(x_i; \boldsymbol{\theta}_i)$. If the perturbation in steady states,* i.e., *$\bar{\boldsymbol{x}} - \boldsymbol{x}^p$, is small and $\|(\boldsymbol{I} - \mathbf{e}_{\pi(p)}\mathbf{e}_{\pi(p)}^\top)\boldsymbol{A}\nabla_{\boldsymbol{x}}\boldsymbol{h}(\boldsymbol{x}^p; \boldsymbol{\theta})\|_2 < 1$, where $\nabla_{\boldsymbol{x}}\boldsymbol{h}(\boldsymbol{x}^p; \boldsymbol{\theta})$ is the Jacobian[2] of the function $\boldsymbol{h}(\boldsymbol{x}; \boldsymbol{\theta})$, then it holds that:*

$$\bar{\boldsymbol{x}} - \boldsymbol{x}^p \approx ([\bar{\boldsymbol{x}}]_{\pi(p)})\boldsymbol{e}_{\pi(p)} + ([\bar{\boldsymbol{x}}]_{\pi(p)})\nabla_{\boldsymbol{x}} h(\mathbf{0}; \boldsymbol{\theta})\boldsymbol{a}_{\pi(p)}^{\text{col}}\ , \tag{7}$$

where $\boldsymbol{e}_{\pi(p)}$ is the $\pi(p)$th coordinate vector, and $\boldsymbol{a}_{\pi(p)}^{\text{col}}$ is the $\pi(p)$th column vector of $\boldsymbol{A}$.

The result above (whose proof can be found in [22]) suggests that the support of $\boldsymbol{a}_{\pi(p)}^{\text{col}}$ can be revealed by merely inspecting and thresholding the difference vector $\bar{\boldsymbol{x}} - \boldsymbol{x}^p$. This is an example of the prior knowledge that we exploit in the network inference method proposed in the next section.

3. GENERAL INFERENCE WITH REGULARIZATION

To infer parameters $\boldsymbol{\Theta}$ of the complex network system from steady states data, we adopt a general regularized inference approach as:

$$\min_{\hat{\boldsymbol{\Theta}}}\ \textstyle\sum_{p=1}^P \mathcal{L}(\hat{\boldsymbol{\Theta}}; \boldsymbol{x}^p) + \mathcal{R}(\hat{\boldsymbol{\Theta}})\ \text{s.t.}\ \hat{A}_{ii} = 0,\ \forall\, i\ , \tag{8}$$

where $\mathcal{L}(\cdot; \cdot)$ is the loss incurred by the miss-match between $\hat{\boldsymbol{\Theta}}$ and the observed perturbed steady state. As the perturbed steady states are directly observed (potentially with some form of observation noise), we apply the least-square loss function[3]:

$$\mathcal{L}(\hat{\boldsymbol{\Theta}}; \boldsymbol{x}^p) = \textstyle\sum_{i=1}^n \left(\ell_i(\hat{\boldsymbol{\Theta}}; \boldsymbol{x}^p)\right)^2\ , \tag{9}$$

[1] Using a straightforward modification, it is possible to extend the model to include multiple knocked out agents in an experiment.

[2] Precisely, the Jacobian is given by the diagonal matrix $\nabla_{\boldsymbol{x}}\boldsymbol{h}(\boldsymbol{x}^p; \boldsymbol{\theta}) := \text{Diag}([..., \partial h(x; \boldsymbol{\theta}_i)/\partial x|_{x=x_i^p}, ...])$.

[3] As an extension, one could consider alternative loss functions for different observation models, e.g., logistics regression for binary input.

where $\ell_i(\cdot;\cdot)$ is defined in (4) or (6), depending on the type of perturbation experiments considered. Note that $\hat{\boldsymbol{\Theta}}$ encompasses two sets of parameters of the complex network system — *(i)* the network parameters, *i.e.,* $\hat{\boldsymbol{A}}$ (also the nuisance parameters modeled in Type-I experiments); *(ii)* the dynamics parameters, *i.e.,* $\{\hat{\boldsymbol{\theta}}_i\}_{i=1}^n$.

We are interested in the scenario when $P \ll n$. Note that in this case we are faced with the issue of rank deficiency as there are at least $n^2 - n$ unknowns in $\hat{\boldsymbol{A}}$ while only Pn observations are available. This suggests that an appropriate regularization function is necessary to guarantee recoverability for network inference using (8). We propose the following strategies for regularizing the network adjacency matrices (e.g., $\hat{\boldsymbol{A}}$, $\hat{\boldsymbol{B}}$) with different type of experiments:

- *(Type-I: New agents injection)* We apply:

$$\mathcal{R}^{\text{type}-1}(\hat{\boldsymbol{\Theta}}) = \rho\big(\|\text{vec}(\hat{\boldsymbol{B}})\|_1 + \|\text{vec}(\hat{\boldsymbol{A}})\|_1\big) \ , \qquad (10)$$

  for some $\rho > 0$. The regularizer simply imposes a sparse prior on the unknown network parameters, *i.e.,* the network is sparsely connected.

- *(Type-II: Knockout experiments)* Under Assumption 1 we apply:

$$\mathcal{R}^{\text{type}-2}(\hat{\boldsymbol{\Theta}}) = \rho\|\text{vec}(\hat{\boldsymbol{A}})\|_1 + \mathcal{I}_{\mathcal{S}}(\hat{\boldsymbol{A}}) \ , \qquad (11)$$

  where $\mathcal{I}_{\mathcal{S}}$ is an indicator function satisfying

$$\mathcal{I}_{\mathcal{S}}(\hat{\boldsymbol{A}}) = \begin{cases} \infty, & \text{if } [\hat{\boldsymbol{A}}]_{ij} \neq 0 \text{ for some } (i,j) \in \mathcal{S}, \\ 0, & \text{otherwise} \ , \end{cases} \qquad (12)$$

  and for some $\delta > 0$, we set

$$\mathcal{S} := \bigcup_{p=1}^{P} \Big\{ (i, \pi(p)) : i \in [n], \ \frac{\big[\bar{\boldsymbol{x}} - \boldsymbol{x}^p\big]_i}{[\boldsymbol{x}]_{\pi(p)}} < \delta \Big\} \ . \qquad (13)$$

  In addition to promoting sparsity in the network adjacency matrix $\hat{\boldsymbol{A}}$, we also exploit the partial support estimated by Proposition 1. Specifically, set $\mathcal{S}$ is an estimation for the locations of *zeros* in the sub-matrix $[\boldsymbol{A}]_{:,\pi([P])}$ as predicted by the proposition.

Due to the coupling between $\hat{\boldsymbol{A}}$ and $\hat{\boldsymbol{\theta}}_i$ as seen in the loss function (4) and (6), the inference problem (8) is non-convex in general. To handle this, we adopt the standard alternating optimization approach. Let k be the iteration number, we alternate between the two steps — *(i)* given $\{\hat{\boldsymbol{\theta}}_i^{(k)}\}_{i=1}^n$, problem (8) is convex and we solve

$$\widehat{\boldsymbol{A}}^{(k+1)} \in \arg\min_{\widehat{\boldsymbol{A}}} \ \mathcal{L}(\widehat{\boldsymbol{A}}, \{\boldsymbol{\theta}_i^{(k)}\}_{i=1}^n; \boldsymbol{x}^p) + \mathcal{R}(\hat{\boldsymbol{\Theta}}) \ ; \qquad (14)$$

(ii) given $\hat{\boldsymbol{A}}^{(k+1)}$ (and $\hat{\boldsymbol{B}}^{(k+1)}$), problem (8) is in general non-convex, yet we can update the dynamics parameters using the gradient method with a fixed step size:

$$\boldsymbol{\theta}_i^{(k+1)} = \boldsymbol{\theta}_i^{(k)} - \gamma \textstyle\sum_{p=1}^{P} \nabla_{\boldsymbol{\theta}_i} \mathcal{L}(\hat{\boldsymbol{A}}^{(k+1)}, \{\boldsymbol{\theta}_i^{(k)}\}_{i=1}^n; \boldsymbol{x}^p) \ , \qquad (15)$$

where $\gamma > 0$ is a step size. This procedure is a hybrid of alternating optimization with exact and inexact updates. Furthermore, it is guaranteed to converge to a stationary point of (8) under mild conditions on the loss functions [23].

We also remark that in the sub-problem for optimizing the network-related parameters $\widehat{\boldsymbol{A}}$, the problem (14) has a friendly decomposable structure. In particular, we observe that the loss functions in (4) and (6) are only dependent on the ith row of $\hat{\boldsymbol{A}}$ (or in addition $\hat{\boldsymbol{B}}$). When combined with the separability of the regularization functions in (10) and (11), this suggests that the inference problem can be decomposed into n independent sub-problems that can be solved in a parallel fashion. In fact, even with a naive implementation such as solving (14) by off-the-shelf software as `cvx` [], one could handle moderately sized network with $n \approx 1000$.

3.1. Identifiability Guarantees

We briefly review the identifiability guarantees for our general inference method. To derive meaningful results, we focus on the special case when the dynamics parameters $\{\boldsymbol{\theta}_i\}_{i=1}^n$ are known. Moreover, the observations are noiseless such that we impose the constraint $\mathcal{L}(\widehat{\boldsymbol{A}}, \{\boldsymbol{\theta}_i^{(k)}\}_{i=1}^n; \boldsymbol{x}^p) = 0$ in (14) [or we set the regularization parameter ρ to sufficiently large to enforce the former].

For Type-I experiments, a special case was analyzed in [18] with linear network dynamics, *i.e.,* when $h(x_i, x_j; \boldsymbol{\theta}) = x_j - x_i$ for all $\boldsymbol{\theta}$ and $g_i(x) = x$ for all i. We set $P = S$ such that the steady state data is rank-S and is the maximum rank achievable with S injected agents. Furthermore, [18] considers an *active sensing* setting where the *support* of $\boldsymbol{B}$ is known. The authors considered a regularization similar to (10) which takes the support of $\boldsymbol{B}$ into account and promotes sparsity on $\hat{\boldsymbol{A}}$. The results below hold when the support of $\boldsymbol{B}$ is drawn according to a random, ℓ-row regular pattern:

Theorem 1 *[18] If the number of injected agents S satisfies*

$$S \geq \beta(\alpha, \ell)n + \ell \ , \qquad (16)$$

where $\beta(\alpha, \ell)$ is a nonlinear function of the maximum degree fraction $\alpha := \max_i \|\boldsymbol{a}_i^{\mathsf{row}}\|_0/n$ which approaches a constant factor of α as $n \to \infty$. Then, as $n \to \infty$, the optimal solution to (14) *is unique satisfying $(\hat{\boldsymbol{A}}, \hat{\boldsymbol{B}}) = (\boldsymbol{A}, \boldsymbol{B})$ with probability one.*

The precise conditions are technical and are skipped in the interest of space. Interested readers are referred to [19] for a gentle description.

For Type-II experiments, the identifiability result is proven under Assumption 1 and a set of assumptions below:

Assumption 2 *(a) The approximation in Proposition 1 is exact and $\mathcal{S}$ finds correctly all the zeros in $[\boldsymbol{A}]_{:,\pi([P])}$; (b) matrix $\boldsymbol{A}$ is non-negative; (c) $\boldsymbol{h}(\boldsymbol{x}; \boldsymbol{\theta})$ has an exact first order Taylor approximation.*

Also, define $\mathcal{S}_i := \{j : (i,j) \in \mathcal{S}\}$ as the restriction of $\mathcal{S}$ to the indices of the ith row in $\boldsymbol{A}$. We have

Theorem 2 *[22] Under Assumption 1 and 2. Denote as G_i the bipartite graph induced by the outgoing edges from $\pi([P]) \setminus \mathcal{S}_i$ to all the nodes in V. If for all $i \in V$, G_i is an (α, δ)-unbalanced expander graph with left degree bounded in $[d_l, d_u]$ such that*

$$(1 + (d_l/d_u)\delta)\|\boldsymbol{a}_i^{\mathsf{row}}\|_0 \leq \alpha n, \ \ 2(d_l/d_u)\delta > \sqrt{5} - 1 \ , \qquad (17)$$

then solving (14) *with the additional constraint $\hat{\boldsymbol{A}} \geq 0$ yields a unique solution that $\hat{\boldsymbol{A}} = \hat{\boldsymbol{A}}$.*

Again, the precise conditions are technical and the interested readers are referred to the working paper [22] for details.

That said, it is possible to satisfy the conditions above by solving (14) using data from $P = \Omega(d_{\mathsf{max}})$ experiments. The required P needs not to grow with n. Moreover, the above conditions favor graphs that are *regular*, *i.e.,* when agents have similar degrees.

4. NUMERICAL RESULTS & CONCLUSIONS

This section presents numerical experiments to verify the efficacy of our proposed inference method under different network models. In particular, for synthetic networks,

Opinion Dynamics. We consider the experiments on synthetic networks. Particularly, the steady states data are collected from a complex network with $n = 1000$ agents and kernel functions:

$$h(x_i, x_j; \theta_i) = \exp(-\theta_i(x_i - x_j)^2) \cdot (x_j - x_i) \ , \qquad (18)$$

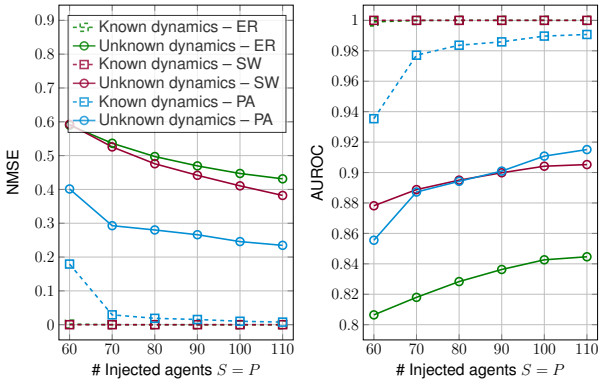

Fig. 1. **Network Inference on Opinion Dynamics**. We compare the normalized mean square error, and area under ROC curve, in recovering $\boldsymbol{A}$ versus the number of new agents injected S, under different network models.

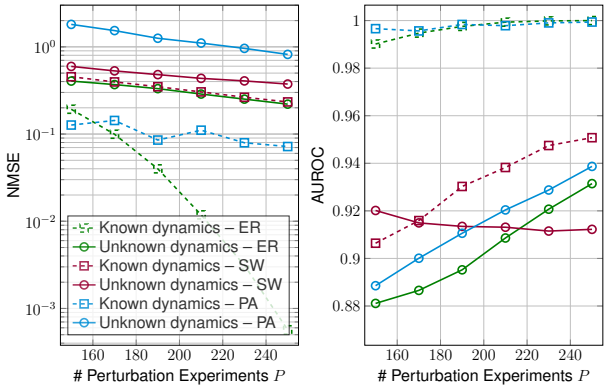

Fig. 2. **Network Inference on Gene Regulatory Dynamics**. Similar models are considered as in the case for opinion dynamics.

Methods	*E. coli* AUROC	*E. coli* AUPR	*S. cerevisiae* AUROC	*S. cerevisiae* AUPR
TIGRESS [16]	0.595	0.069	0.517	0.02
GENIE3 [15]	0.617	**0.093**	0.518	0.021
Proposed (top 100k)	**0.6823**	0.0508	**0.525**	0.02
iRafNet [17]	0.641	**0.112**	N/A	N/A
Proposed (top 500k)	**0.7573**	0.0574	**0.5734**	**0.0252**

⋆All values are computed from the top 100k predictions made. Exceptions are the iRafNet, GENIMS and proposed method (top 500k) in the last three rows, that are based on the top 200k, all, top 500k predictions, respectively.

Table 1. **GRN Recovery on Empirical Dataset.** We use the parameterization with $h(x;\boldsymbol{\theta}_i) = 0.75x^{b_1}/(1+b_2x^{b_1})$ where b_1, b_2 are unknown parameters. Proposed method (3rd row) uses the optimized parameters learnt for the two networks, *i.e.,* $b_1 = 0.047, b_2 = 0.5893$ for *E. coli* and $b_1 = 0.5571, b_2 = 0.3749$ for *S. cerevisiae*.

where $\theta_i > 0$ is a parameter controlling the openness of an agent to accept other's different opinions. Moreover, $g_i(x) = x$ for all i. The networks $\boldsymbol{A}$ are generated as — for Erdos-Renyi (ER) networks, the connection probability is set with $p_{\mathsf{ER}} = 1.01 \log n/n$; for small world (SW) networks, the networks are generated based on the algorithm in [24], set with a degree of $d_{\mathsf{SW}} = \lceil 0.5 \log n \rceil$ with a rewiring probability of 0.2; for preferential attachment (PA) networks, the networks are generated using [25] with a minimum degree of 2. For this configuration at $n = 1000$, the average number of edges in the ER and SW networks are comparable, while the PA networks have a smaller number of edges. Fixing the number of injected agents at S, at each experiment, the initial opinions of all the agents are randomly generated according to $\mathcal{U}[0,1]$. The support of the matrix $\boldsymbol{B}$ is constructed as a random $\ell = 5$ regular bipartite graph. The openness level of each agent is generated as $\theta_i \sim \mathcal{U}[1, 1.5]$ for $i \in \{1, ..., 500\}$ and $\theta_i \sim \mathcal{U}[3.5, 4]$ for $i \in \{501, ..., 1000\}$. We apply 10 iterations for the AO procedure, initiated with $\theta_i = 0.5$ for all i. The performance of network inference is presented in Fig. 1 for cases with unknown or known dynamics parameter.

As seen in the figure, the normalized mean square error (NMSE) and the area under ROC (AUROC) metrics improve with the number of injected agents in all types of networks. Note that $\mathsf{NMSE} \to 0$ indicates perfect identification of the network weights while $\mathsf{AUROC} \to 1$ indicates perfect recovery of the network topology. In particular, given the similar average edge density, the SW networks can be recovered with less number of injected agents than ER networks. This is due to the fact that the networks has less number of nodes with high degree. Lastly, the estimated dynamics parameters are matched with the ground truth. To illustrate this, for an instance of the recovery of an PA network with $S = 110$ injected agents, the recovered dynamics parameters satisfy $\mathbb{E}[\hat{\theta}_i] \approx 0.89$ for $i \in \{1, ..., 500\}$ and $\mathbb{E}[\hat{\theta}_i] \approx 2.4$ for $i \in \{501, ..., 1000\}$. This is compatible with our generation model.

Gene Dynamics. We first consider an application of our method on synthetic networks. Here, the dynamics equations satisfy Assumption 1 and the kernel function is given by $h(x;\boldsymbol{\theta}) = x^{b_1}/(b_2 + x^{b_1})$. The networks $\boldsymbol{A}$ are generated according to the same models described for opinion dynamics except that we set $p_{\mathsf{ER}} = 2 \log n/n$ and $d_{\mathsf{SW}} = \lceil \log n \rceil$ for ER and SW networks, respectively. The dynamics parameters for all agents are set as $b_1 = 0.5$ and $b_2 = 1$. The steady states of the network is computed using the Runge-Kutta 4th order method and a randomly chosen agent is knocked out in each perturbation experiment. Our results are presented in Fig. 2 for these synthetic networks. For ER and PA networks, we observe the same trend as in the case of opinion dynamics where the performance improves with the number of perturbation experiments observed. Interestingly, for SW networks, the AUROC performance deteriorates with P when the dynamics are unknown. This is potentially due to the lack of support for conditions in Theorem 2 with SW networks.

Finally, we apply our proposed method to two empirical datasets taken from the DREAM5 challenge [21]. The datasets are microarray experiments (where each record is the normalized gene expression) performed on *E. Coli* and *S. cerevisiae*. The *E. Coli* network has $n = 4511$ genes and the number of experiments (of the same conditions) is $P = 56$, while the *S. cerevisiae* network has $n = 5950$ genes and we have $P = 7$ experiments only. The results are presented in Table 1 where our method has outperformed the state-of-the-art methods. Details of our numerical experiments can be found in the technical report [22].

To conclude, this paper has described a general approach for the joint recovery problem of network and dynamics in complex networks, where only steady states data are used. We have also presented identifiability conditions required for perfect recovery of these networks. Numerical results show the benefits of our approach.

5. REFERENCES

[1] T. Söderström and P. Stoica, "System identification," 1989.

[2] J. Friedman, T. Hastie, and R. Tibshirani, "Sparse inverse covariance estimation with the graphical lasso," *Biostatistics*, vol. 9, no. 3, pp. 432–441, 2008.

[3] S. Segarra, A. G. Marques, G. Mateos, and A. Ribeiro, "Network topology inference from spectral templates," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 3, no. 3, pp. 467–483, 2017.

[4] D. I. Shuman, S. K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, "The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains," *IEEE Signal Processing Magazine*, vol. 30, no. 3, pp. 83–98, 2013.

[5] A. Sandryhaila and J. M. Moura, "Discrete signal processing on graphs," *IEEE transactions on signal processing*, vol. 61, no. 7, pp. 1644–1656, 2013.

[6] A. Ortega, P. Frossard, J. Kovačević, J. M. Moura, and P. Vandergheynst, "Graph signal processing," *arXiv preprint arXiv:1712.00468*, 2017.

[7] X. Cai, J. A. Bazerque, and G. B. Giannakis, "Inference of gene regulatory networks with sparse structural equation models exploiting genetic perturbations," *PLoS computational biology*, vol. 9, no. 5, p. e1003068, 2013.

[8] X. Dong, D. Thanou, P. Frossard, and P. Vandergheynst, "Learning laplacian matrix in smooth graph signal representations," *IEEE Transactions on Signal Processing*, vol. 64, no. 23, pp. 6160–6173, 2016.

[9] V. Kalofolias, "How to learn a graph from smooth signals," in *Artificial Intelligence and Statistics*, 2016, pp. 920–929.

[10] J. Mei and J. M. Moura, "Signal processing on graphs: Causal modeling of unstructured data," *IEEE Transactions on Signal Processing*, vol. 65, no. 8, pp. 2077–2092, 2017.

[11] H. E. Egilmez, E. Pavez, and A. Ortega, "Graph learning from data under laplacian and structural constraints," *IEEE Journal of Selected Topics in Signal Processing*, vol. 11, no. 6, pp. 825–841, 2017.

[12] Y. Shen, B. Baingana, and G. B. Giannakis, "Kernel-based structural equation models for topology identification of directed networks," *IEEE Transactions on Signal Processing*, vol. 65, no. 10, pp. 2503–2516, 2017.

[13] M. Timme, "Revealing network connectivity from response dynamics," *Physical review letters*, vol. 98, no. 22, p. 224101, 2007.

[14] W.-X. Wang, Y.-C. Lai, C. Grebogi, and J. Ye, "Network reconstruction based on evolutionary-game data via compressive sensing," *Physical Review X*, vol. 1, no. 2, p. 021021, 2011.

[15] V. A. Huynh-Thu, A. Irrthum, L. Wehenkel, P. Geurts *et al.*, "Inferring regulatory networks from expression data using tree-based methods," *PloS one*, vol. 5, no. 9, p. e12776, 2010.

[16] A.-C. Haury, F. Mordelet, P. Vera-Licona, and J.-P. Vert, "Tigress: trustful inference of gene regulation using stability selection," *BMC systems biology*, vol. 6, no. 1, p. 145, 2012.

[17] F. Petralia, P. Wang, J. Yang, and Z. Tu, "Integrative random forest for gene regulatory network inference," *Bioinformatics*, vol. 31, no. 12, pp. i197–i205, 2015.

[18] H.-T. Wai, A. Scaglione, and A. Leshem, "Active sensing of social networks," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 2, no. 3, pp. 406–419, 2016.

[19] ——, "Active sensing of social networks: Network identification from low rank data," in *Cooperative and Graph Signal Processing*, P. M. Djuric and C. Richard, Eds. Elsevier, 2018.

[20] G. Giaever, A. M. Chu, L. Ni, C. Connelly, L. Riles, S. Veronneau, S. Dow, A. Lucau-Danila, K. Anderson, B. Andre *et al.*, "Functional profiling of the saccharomyces cerevisiae genome," *nature*, vol. 418, no. 6896, p. 387, 2002.

[21] D. Marbach, J. C. Costello, R. Küffner, N. M. Vega, R. J. Prill, D. M. Camacho, K. R. Allison, A. Aderhold, R. Bonneau, Y. Chen *et al.*, "Wisdom of crowds for robust gene network inference," *Nature methods*, vol. 9, no. 8, p. 796, 2012.

[22] H.-T. Wai, A. Scaglione, U. Harush, B. Barzel, and A. Leshem, "Rids: Robust identification of sparse gene regulatory networks from perturbation experiments," *arXiv preprint arXiv:1612.06565*, 2016.

[23] M. Razaviyayn, M. Hong, and Z.-Q. Luo, "A unified convergence analysis of block successive minimization methods for nonsmooth optimization," *SIAM Journal on Optimization*, vol. 23, no. 2, pp. 1126–1153, 2013.

[24] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world'networks," *nature*, vol. 393, no. 6684, p. 440, 1998.

[25] A.-L. Barabási and R. Albert, "Emergence of scaling in random networks," *science*, vol. 286, no. 5439, pp. 509–512, 1999.