Integrated Graph Propagation and Optimization with Biological Applications

Krithika Krishnan†, Tiange shi‡, Han Yuↀ, Rachael Hageman Blair† ‡

†Institute of Artificial Intelligence and Data Science, University at Buffalo. ‡ Department of Biostatistics, University at Buffalo.

ↀ Department of Biostatistics and Bioinformatics, Roswell Park Comprehensive Cancer Center.

Introduction

The ability to estimate a sensitivity matrix from propagation on the structure has important implications for network optimization, which to the author's knowledge, has not been explored. This work develops the first optimization framework that leverages the sensitivity matrix to identify optimal perturbation patterns to drive a network to a target steady-state. A novel approach, Integrated Graph Propagation, and OptimizatioN (IGPON), were developed, which casts the problem as an unconstrained optimization that minimizes the difference between a current network state and a desired target network state.

Methods

Integrated Graph Propagation with OptimizatioN (IGPON)

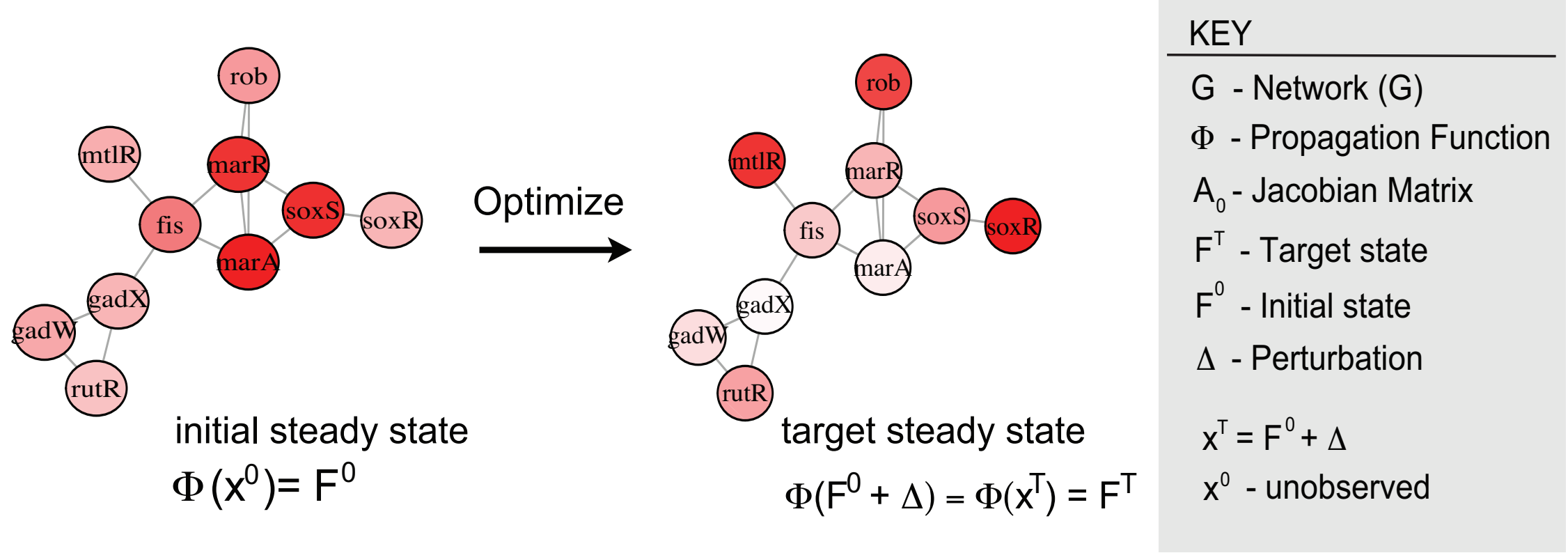

Figure 1: Schema of the integrated graph propagation and optimization with biological applications (IGPON) method. IGPON drives an observed initial steady-state of the network, F^0, to a target steady-state, F^T, through the identification of an optimal perturbation, $\boldsymbol{\Delta}$, such that $\boldsymbol{\Phi}(F^0 + \boldsymbol{\Delta}) = F^T$.

- Network structure, *G*, can be directed or undirected and contains *n* nodes (Figure 2).
- Structure is known *a priori* as either inferred from data or specified by an expert or database.
- Let *Φ(·)* be the output of propagation (PRINCE algorithm).
- Sensitivity matrix plays the role of the initial Jacobian, A_0 (Figure 2B), and is estimated directly using graph propagation, *Φ(G)* (Figure 1), and the state of the network *Φ(x) = F*.
- Influence score at iteration *t* is given as

$$F^t := \alpha G' F^{t-1} + (1 - \alpha) \cdot Y$$

where α is a diffusion constant that score enforces smoothness over the network, and *Y* is an initial set of scores, F^0 [1].

Unconstrained minimization problem is defined as:

$$\min_x ||\Phi(x) - F^T||_2$$

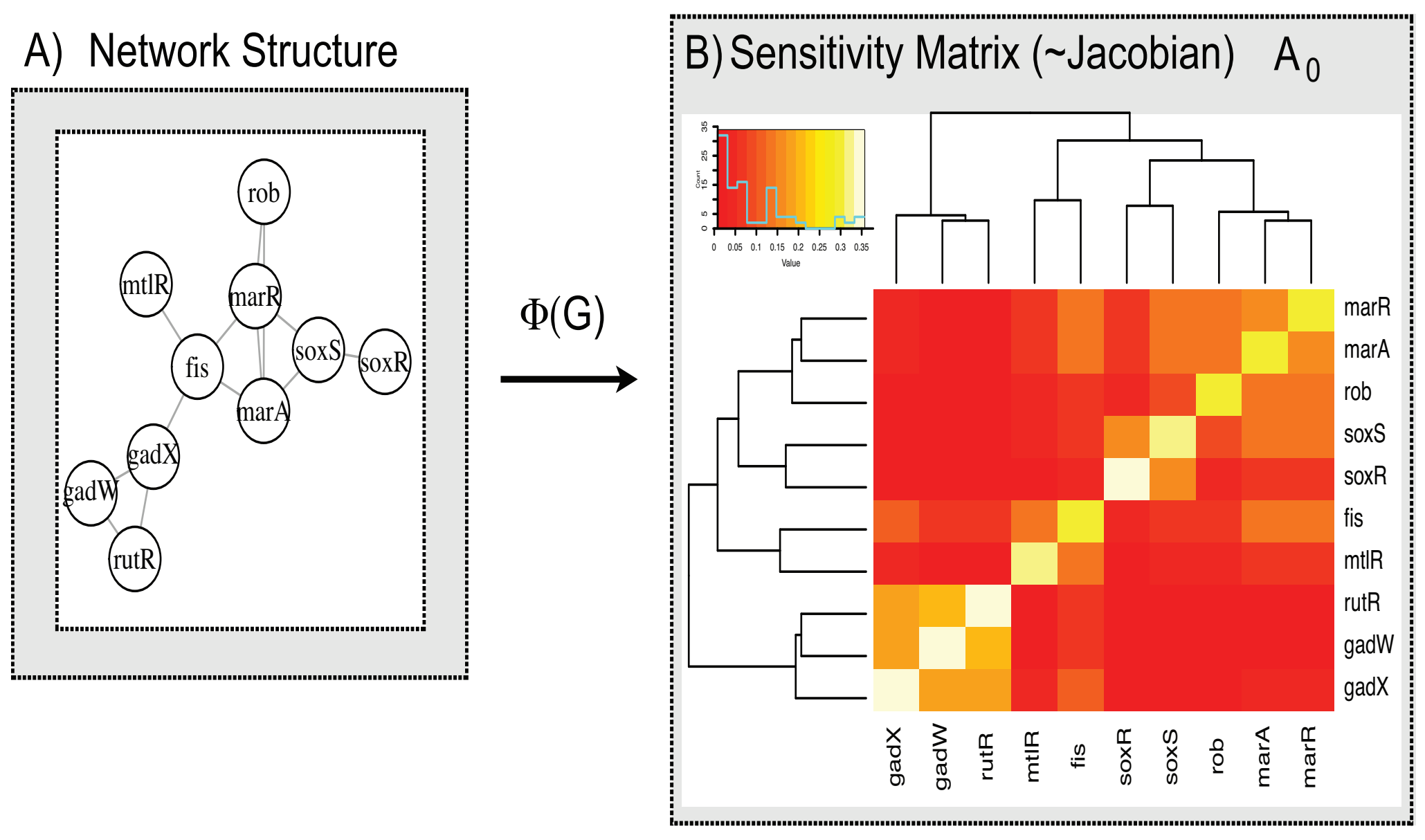

Figure 2: (A) The structure of the network (graph), G, is assumed to be given. (B) The sensitivity matrix [2] derived through graph propagation, $\boldsymbol{\Phi}$(G), on the network structure, serves as the initial Jacobian, A_0.

Simulation and Biological datasets

IGPON was applied to the simulated random graphs and data from the DREAM4 *in silico* challenge.

1. Random graphs were generated with 50 and 150 nodes using the Barábasi-Albert model, $x_0 \sim U[0, 1]$ (Figure 3). The probability of an edge between two arbitrary vertices was set at *p = 0.10*.
2. DREAM networks of 10 and 98 nodes were generated using DREAM4 data.

Knockout dataset

- Gene expression data was utilized from the KnockTF database [3] for two experimental conditions.
- Knockout data for transcription factor signal transducer and activator of transcription 6 (STAT6) was extracted from the database.
- The gene expression data contained wild-type controls (N = 27) and STAT6 knockout (N = 27).

Results

- IGPON was tested on simulations of random graphs (Figure 4), DREAM4 networks (Figure 4), and data from a knockout database [3] with node and edge sizes 52 x 162, 53 x 147, and 62 x 75, respectively (Table 1).
- Systematic white noise (10% - 50%) was added to the initial sensitivity matrix to represent misspecifications in the network structure (Table 1 and Figure 3, 4).

Pathway Name	Graph Type	Number of Iterations			
		0% noise	10% noise	25% noise	50% noise
HH	Directed	2	69	154	278
	Undirected	2	85	178	338
IL-17	Directed	2	72	162	335
	Undirected	2	87	202	388
p53	Directed	2	79	199	383
	Undirected	2	90	230	390

Table 1: Convergence of Biological Pathways to Target States

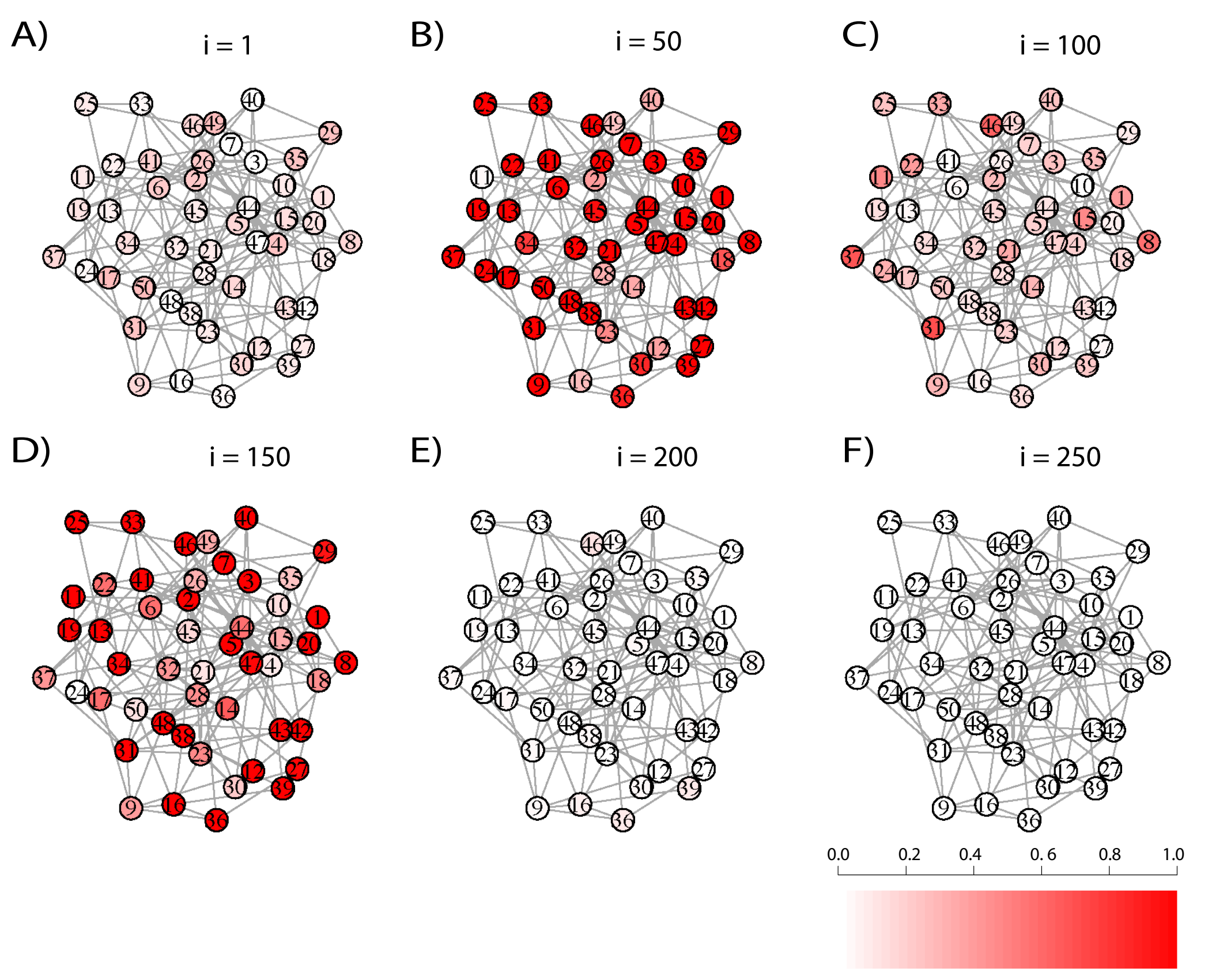

Figure 3: Node convergence profiles for the simulated 50 node network with 25% noise added to the Jacobian at select IGPON iterations *k*. The coloring of a node *i* corresponds to the relative error, at iteration (A) *k = 1,* (B) *k = 50,* (C) *k = 100,* (D) *k = 150,* (E) *k = 200,* (F) *k = 250.*

- IGPON was able to drive expression profiles to target states in the HH, IL-17, and p53 pathways.
- In both directed and undirected representations, convergence was achieved across all noise levels.

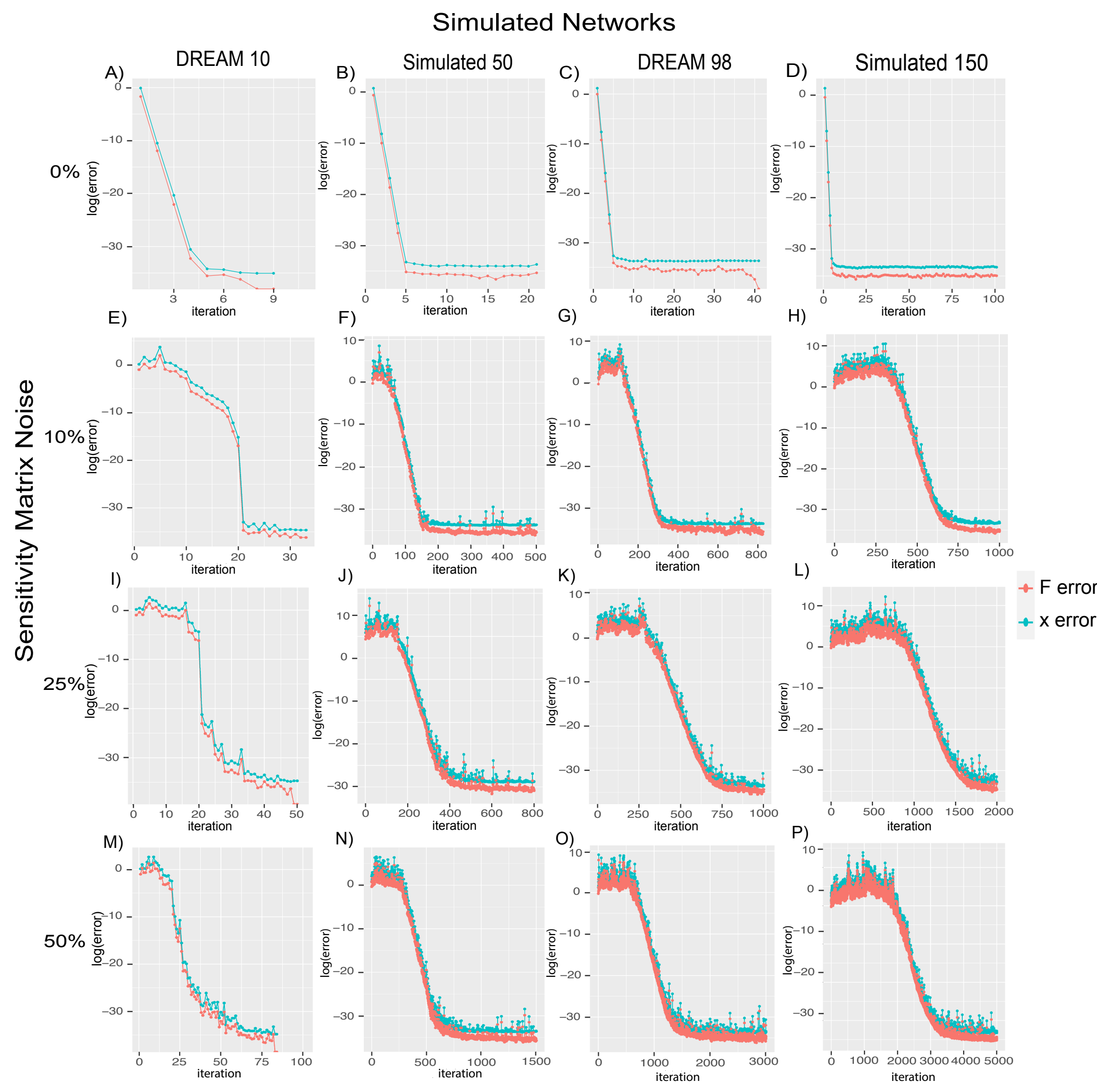

Figure 4: Convergence profiles of the log(error) for *F* (coral) and *x* (blue). Simulated graphs are ordered according to size (columns): columns 1 (N = 10), column 2 (N = 50), column 3 (N = 98), and column 4 (N = 150). The rows represent the noise level added to the sensitivity matrix in the optimization. (A-D) No noise is added (E-H) 10%, (I-L) 25%, and (M-P) 50%.

Conclusion

- A distinguishing feature of this method is that the optimization relies on using two primary ingredients: a parameterized network structure and target node states.
- IGPON bypasses the need for complex forms of biochemical reactions and derivatives.
- IGPON is an effective way to optimize directed and undirected networks that are also robust to noise in the sensitivity matrix that reflects potential misspecification in the structure.
- IGPON embeds propagation into an optimization that can drive an undirected or directed graph to a desired steady-state using graph structure and integrated propagation.
- IGPON works directly with a network structure and does not rely on complex parameterizations.
- Predicting optimal perturbations to drive biological systems to any desired state is a promising area of research in biological and genetic engineering.

References

1. O. Vanunu, O. Magger, E. Ruppin, T. Shlomi and R. Sharan, Associating genes and protein complexes with disease via network propagation, PLoS Computational Biology 6, p.e1000641 (January 2010).
2. M. Santolini and A.-L. Barabási, Predicting perturbation patterns from the topology of biological networks, Proceedings of the National Academy of Sciences 115, E6375 (2018).
3. C. Feng, C. Song, Y. Liu, F. Qian, Y. Gao, Z. Ning, Q. Wang, Y. Jiang, Y. Li, M. Li, J. Chen, J. Zhang and C. Li, KnockTF: a comprehensive human gene expression profile database with knockdown/knockout of transcription factors, Nucleic Acids Research 48, D93 (2020).