Network optimal retrieval of sparse perturbations for steady-state control

Krithika Krishnan[1], Tiange Shi[2], Satyam Kumar[1], Han Yu[3] and Rachael Hageman Blair[1,2]

[1]*Institute for Artificial Intelligence and Data Science, University at Buffalo,*
[2]*Department of Biostatistics, University at Buffalo,*
Buffalo, NY 14228, USA
www.buffalo.edu

[3]*Roswell Park Comprehensive Cancer Center*
Buffalo, NY 14263, USA
www.buffalo.edu

Corresponding Author:
Rachael Hageman Blair
E-mail: hageman@buffalo.edu

Prioritizing targeted perturbation experiments remains a central challenge in systems biology, where experimental constraints limit network manipulation. We introduce NORSP (Network Optimal Retrieval of Sparse Perturbations). This novel computational framework integrates network propagation with supervised subset selection to identify minimal perturbation sets that can shift a system from its initial to a desired steady state. NORSP leverages a sensitivity matrix derived solely from network topology, enabling control prediction without requiring full knowledge of system dynamics. Applicable to undirected, directed, and signed networks, NORSP accommodates a broad range of biological models and experimental scenarios. We validate its effectiveness using YBX1 knockdown transcriptomics data and 61 curated metabolic networks from the BioModels repository, demonstrating NORSP's robustness, scalability, and experimental relevance. Even under constraints that obscure true perturbations, the algorithm reliably infers alternative targets that achieve comparable control. Control is confirmed both in graphical approximations and through full dynamical model simulations. Overall, NORSP provides a practical and generalizable solution for steady-state control in complex biological systems, laying the foundation for multi-omics hypothesis generation and systems-level experimental design.

Keywords: Biological networks, systems biology, optimization, propagation

1. Introduction

Network analysis is central to systems biology, enabling the exploration of gene regulation, protein interactions, and metabolic control. Mathematical modeling has transformed this domain, offering powerful frameworks to simulate biological responses and drive hypothesis generation. Computation manipulations of network components, known as *in silico* experiments, allow for systematic investigation of system perturbation, guiding experimental design and

accelerating discovery. Despite their utility, computational models face a critical bottleneck: identifying which perturbations are most informative. Network flexibility and high dimensionality introduce uncertainty, while practical considerations such as experimental cost, biological feasibility, and node accessibility constrain the number of feasible interventions. Addressing this challenge requires control strategies that are both tractable and biologically grounded.

Recent advances have accelerated the integration of network modeling, causal inference, and perturbation data for the discovery of therapeutic targets. Integrated regulatory–metabolic frameworks couple probabilistic gene regulatory networks with constraint-based metabolic models to predict cell type–specific responses to interventions and prioritize therapeutic strategies.[1] Causal inference methods such as Causal Differential Networks (CDN) identify intervention targets by comparing causal graphs derived from observational and interventional single-cell data, outperforming existing approaches on Perturb-seq benchmarks.[2] Bayesian frameworks like BaCaDI further address unknown interventions by jointly inferring causal structures and targets.[3] Network-based approaches, including NETPERT, use perturbation theory to prioritize druggable intermediates,[4] while data-driven control methods enable optimal network manipulation without explicit dynamical models.[5] Together, these developments highlight emerging strategies that combine causal modeling, network theory, and data-driven methods to identify minimal and effective intervention sets in complex biological systems.

Beyond causal inference and control frameworks, network propagation provides a complementary strategy for modeling how perturbations diffuse through biological networks. By iteratively updating the state of each node based on its neighbors, propagation estimates the spread of information across a graph, effectively simulating how biological signals disseminate through molecular systems. This approach has been widely applied in systems biology for identifying disease-associated genes,[6,7] uncovering genetic associations,[8] characterizing biological networks,[9] and investigating cancer.[10,11] In this work, we build on the PRINCE (PRIoritizatioN and Complex Elucidation) algorithm,[12] which was originally developed to identify associations between protein complexes and human diseases using network propagation.

Building on propagation concepts, the DYNAMO (DYNamics-Agnostic Network MOdels) framework[13] extended network propagation to approximate perturbation responses in biological systems. DYNAMO introduced the key insight that a sensitivity matrix, derived solely from network topology, can serve as a surrogate for the Jacobian matrix of partial derivatives. This matrix quantifies the local influence of perturbations across the network and enables meaningful predictions when kinetic parameters or reaction mechanisms are unavailable. By leveraging structural information alone, DYNAMO outperformed traditional centrality-based metrics across several biological models.

More recently, the IGPON (Integrated Graph-based Perturbation Optimization using Network propagation) algorithm[14] used the sensitivity matrix to optimize perturbations through Broyden's quasi-Newton method, reframing network reprogramming as an unconstrained optimization problem without requiring explicit system dynamics. As a proof of concept, IGPON showed that propagation-based approximations can guide networks toward desired steady states. However, it assumes perturbations can be applied to all N nodes, which is rarely

feasible experimentally, as only a limited subset of nodes can typically be targeted.

This work introduces NORSP (Network Optimal Retrieval of Sparse Perturbations), a computational framework for identifying minimal, biologically actionable subsets of perturbations capable of steering a system toward a desired steady state. NORSP integrates network propagation with subset selection, reformulating network reprogramming as a tractable supervised learning problem. By using a sensitivity matrix derived solely from network topology to guide a greedy search, the method enables targeted intervention design without requiring detailed dynamical models.

In relation to existing strategies, NORSP fills a distinct methodological gap. Frameworks such as DYNAMO and IGPON prioritize individual nodes based on influence or centrality, whereas this approach identifies the minimal subset of nodes needed to achieve a specified perturbation. Causal inference methods like CDN and BaCaDI focus on inferring targets from perturbation data rather than determining how to intervene to control network states. By contrast, NORSP emphasizes control under experimental constraints, making it well suited for prioritizing limited, biologically feasible interventions.

NORSP is broadly applicable to undirected, directed, and mixed-direction networks. Its performance was evaluated using YBX1 knock-out gene expression data from breast cancer cell lines[15] and 61 curated metabolic models from BioModels.[16] Across these applications, the algorithm identified biologically meaningful targets in Wnt, T-cell receptor, and Jak–Stat signaling pathways, and reliably found minimal perturbation sets capable of shifting metabolic systems to desired steady states. Even when exact perturbations were not recovered, the NORSP identified alternative nodes that achieved comparable control, underscoring its potential as a scalable and practical tool for hypothesis generation in complex biological networks.

2. Materials and Methods

2.1. *Network Optimal Retrieval of Sparse Perturbations*

A network (graph), G, is defined by a set of nodes (vertices), V, and the edges, E, representing the connections between the nodes. Propagation is a process that simulates the flow of information through a network and can be used to estimate a node's influence. Mathematically, we define the graph adjacency matrix, $G = (V, E, w) \in \mathbf{R}^{N \times N}$, where w defines an edge weight. This work assumes that the connections are unweighted ($w_i = 1,\ \forall i$) and that G is a binary adjacency matrix. Undirected graphs are symmetric, whereas directed graphs are binary matrices with $g_{i,j} = 1$ if there is a directed edge between v_i and v_j.

A normalized version of the adjacency matrix that accounts for node degree[17] is defined as: $G' = D_1^{-1/2} G D_2^{-1/2}$, where D_1 and D_2 are the diagonal matrices with entries $d_1(i,i)$ and $d_2(i,i)$ as the sum of the absolute values of the rows and columns, respectively. This normalization creates a smoothing effect over the network.[18] The sensitivity matrix is defined as: $S = (1 - \alpha)(I - \alpha G')^{-1}$, where $\alpha \in [0, 1]$ denotes a tuning parameter for the propagation strength.[13] The sensitivity matrix captures the effect of node perturbations on other nodes in the network, and has been shown to be a good proxy for the Jacobian matrix.[13] Unlike the Jacobian matrix, which relies on partial derivatives of the system, the sensitivity matrix is estimated directly from the network structure (Figure 1A).

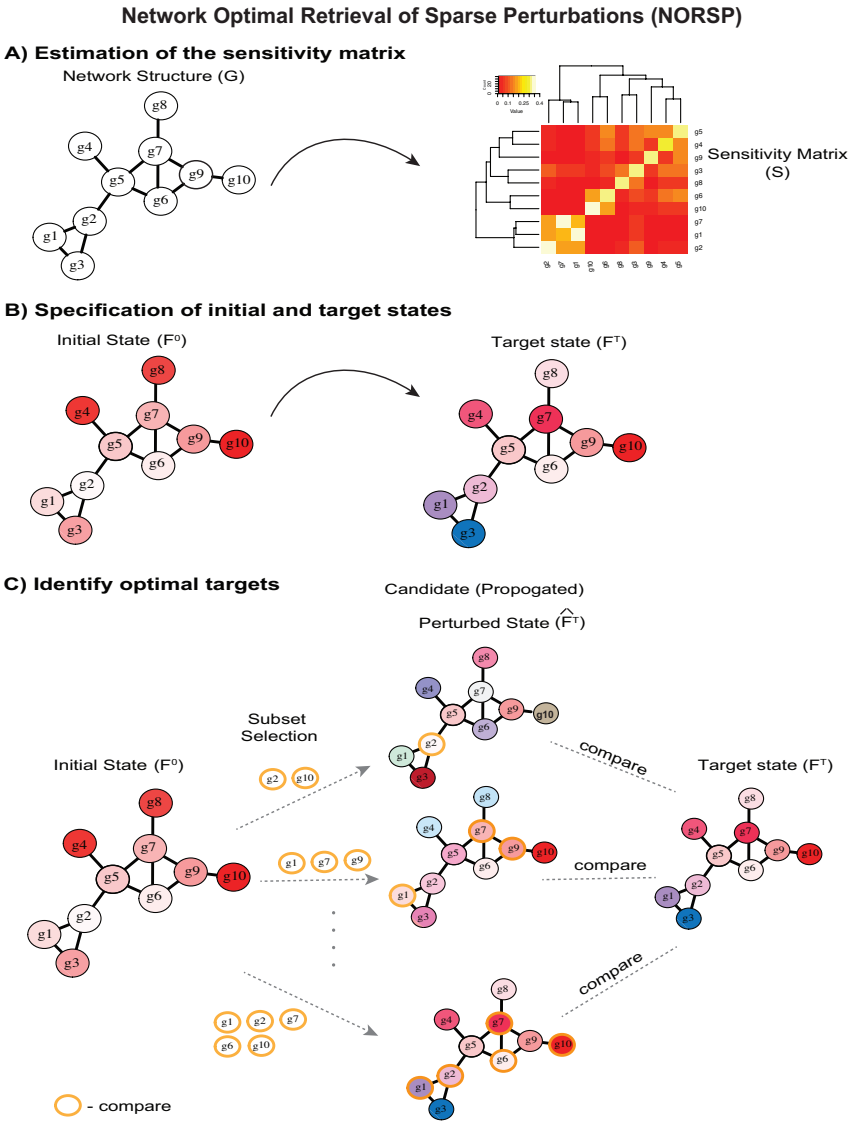

Fig. 1. Schematic of the NORSP framework. **(A)** The graph, G, serves as input and is used to generate a sensitivity matrix, S. **(B)** The optimization problem is to identify the optimal perturbations to drive an initial steady-state, F^0, to a target steady-state, F^T. **(C)** Subset selection generates sets of perturbed candidate states, $\hat{F}^T$, that are compared to the target, F^T, via error calculations. Candidate states are sorted in terms of proximity to the target to identify the optimal perturbation that drives the initial steady-state closest to the target.

This work aims to identify optimal perturbations that drive an observed network, F^0, to a target state, F^T (Figure 1B). This requires the selection of nodes to perturb, and an estimation of the nature of their perturbation. The methodology integrates machine learning via subset selection and network propagation, and we define it as: Network Optimal Retrieval of Sparse Perturbations (NORSP). Mathematically, we want to identify a perturbation vector, $\Delta \in \mathbf{R}^N$, such that:

$$F^T = F^0 + S\Delta. \tag{1}$$

In other words, we want to find a small change $S\Delta$, that will drive F^0 to the target F^T. Rewriting Equation 1, we obtain the following linear model:

$$F^T - F^0 = S\Delta + \epsilon. \tag{2}$$

Let the dependent variable be $y = F^T - F^0$, then the regression model is defined as:

$$y = S\Delta + \epsilon, \tag{3}$$

where $\epsilon \sim N(0, \sigma^2 I)$. The elements of the sensitivity matrix, $S(i, j)$, express the sensitivity of node v_i to small changes in node v_j.

A single or a small number of perturbations are typically possible in experimental systems. Thus, we aim to find a parsimonious solution to this problem that perturbs only a subset of nodes. NORSP uses subset selection to identify the solution that minimizes $\|y - S\Delta\|_2^2$.[19] This feature selection approach will yield the selected nodes for perturbation and the estimated regression coefficient that serves as the perturbation value. Forward subset selection is employed in this work, although alternative methods can be used. NORSP can be understood in Figure 1C, where different sets of nodes are selected for perturbation and propagated to a candidate perturbed state, $\hat{F}^T$. Candidate models arising from subset selection are again contrasted against the target state to identify the optimal solution. Model selection criteria, such as Bayesian Information Criteria (BIC) or adjusted R^2, can be used for selection. Notably, due to the nature of the graph representation, it is not possible to use data splitting or cross-validation without changing the structure.

2.2. *Simulations and Biological Data*

To rigorously assess NORSP's capabilities, we designed a series of simulations that evaluate its accuracy in perturbation recovery, performance across networks from different organisms and biological contexts and robustness to noise and structural misspecification.

Knockdown data: We applied NORSP to mean gene expression data for YBX1, a prooncogenic transcription factor, from the KnockTF database (GEO: GSE63565). The dataset includes 24 control and 24 knockdown samples, with expression profiles derived from log-transformed means. Significant KEGG pathways (Wnt, MAPK, and JAK–STAT) were selected to test NORSP's ability to identify perturbations that drive the system from control (F^0) to knockdown (F^T). YBX1 itself is not present in these pathways, reflecting downstream effects of the knockdown. Pathway graphs were imported using `KEGGgraph`, mapped via Entrez Gene IDs and gene symbols (`org.Hs.eg.db`), and trimmed to exclude unmapped nodes.

Evaluation on BioModels Repository: BioModels[16] are previously published peer-reviewed mathematical models of dynamical biochemical and cellular systems. The BioModels repository contained XML files in the Systems Biology Markup Language (SBML) format. BioModels were manually downloaded from the repository and read into Matlab using the `libSBML 5.19.0` API library.[20] The Biomodels include both the dynamical systems and the parameter specifications and initial concentration levels of biochemical species.

Simulations were performed by running each Biomodel to steady-state with the initializations provided in the database. The resulting concentration values represent an initial steady state, F^0. A perturbation was made to F^0, which sets the biochemical species with

the highest concentration to zero (a knock-out). The system was run to a new steady-state, F^T, representing the target state. The Jacobian matrices were estimated for each BioModel using finite differences. Following Santolini *et al.*,[13] the Jacobian estimates were used to construct the influence diagram that depicts the network representation we use with NORSP. For each Biomodel, three networks were constructed from the Jacobian: undirected, directed, and directed signed. Biomodels with unconnected nodes in the network were not considered. A total of 61 Biomodels, ranging across organisms and processes, were used for simulation. Network sizes ranged from 7 to 77; details are in the Supplementary Table (https://github.com/krithikakrishnan04/norsp). The inputs to NORSP include the network structure along with the initial and target states for each BioModel.

To evaluate the accuracy of node selection, we ranked the list of perturbed nodes identified in the optimal solution based on the order in which they were selected during the forward subset selection process. The first validation criterion assessed whether the actual perturbed node appeared in the optimal solution. The second criterion evaluated whether the perturbed gene was ranked first in the list. These metrics collectively characterize the effectiveness of NORSP's node selection strategy.

To assess control efficacy in BioModels, we evaluated whether NORSP-identified perturbations could drive the dynamical system toward its target steady state. This was done by reintroducing the top-ranked perturbation into each full dynamical model and simulating the system's response. This step mirrors experimental validation by examining how closely the perturbed system approaches the desired target state. If the top-ranked node matches the true perturbed node, F^T is recovered exactly. If not, alternate perturbations must be considered. To quantify control efficacy, we define a control performance metric: $F^{\text{ctr}}_{\text{dyn}} = \frac{\|F^T - \hat{F}^T_{\text{dyn}}\|_2}{\|F^T - F^0\|_2}$, which indicates movement toward the target state when $F^{\text{ctr}}_{\text{dyn}} < 1$, and divergence when $F^{\text{ctr}}_{\text{dyn}} > 1$. We also consider the relative prediction error of the dynamical model prediction, $F^{\text{err}}_{\text{dyn}} = \frac{\|F^T - \hat{F}^T_{\text{dyn}}\|_2}{\|F^T\|_2}$.

Sensitivity to noise and misspecification of the network structure: To assess the robustness of NORSP to observational noise and network misspecification, we generated scale-free networks with $N_n = 30$ nodes using a Barabási–Albert model. Two perturbation regimes were considered: sparse ($N_p = 5$ perturbed nodes) and dense ($N_p = 25$). Target states were generated by propagating ground-truth perturbation vectors through the sensitivity matrix. Additive Gaussian noise with mean zero and varying variance was then introduced either to the initial state (linear model) to mimic measurement error, or to the sensitivity matrix to mimic structural uncertainty. Noise levels ranged from 10^{-7} to 10^{-1}. For each noise level and condition, NORSP was applied to 100 simulated networks. Performance was quantified using the average relative error in recovering the target steady state ($\bar{F}^{\text{err}}$) and the average relative error in estimating the perturbation vector ($\bar{\Delta}^{\text{err}}$).

NORSP will be released to CRAN (https://cran.r-project.org/) as an open-source R package upon publication, providing an accessible platform for researchers and practitioners to apply the method across biological domains. The package will include simulation tools, visualization features, and support for custom network formats. Code for the simulations can be found on github (https://github.com/krithikakrishnan04/norsp).

3. Results

NORSP was evaluated across two main types of datasets: transcriptomic knockdown data and a diverse collection of 61 dynamic models from the BioModels repository. These datasets span a range of biological scales and modeling formalisms, allowing us to assess NORSP's generalizability and robustness. In both contexts, NORSP successfully identified sparse sets of perturbation targets capable of shifting the system from an initial state (F^0) to a target state (F^T), even under constraints typical of real-world experimental scenarios. The results demonstrate NORSP's ability to accommodate networks of varying sizes, topologies, and biological complexity, providing both accurate target approximations and interpretable, minimal intervention strategies.

We investigated pathway-level control using YBX1 knockdown data. Adjusted R^2 values were computed to assess the quality of the F^T approximation across different pathways. Overall, we found that F^T can be accurately approximated, with model fit naturally improving as the number of selected nodes increases. Using a selection threshold of 0.90, we achieved high accuracy across the three pathways while obtaining a parsimonious solution. Figure 2 shows the adjusted R^2 as a function of the selected nodes in the respective pathways. The selected nodes at the 0.90 threshold are indicated in red. Results indicate the relative error (F^{err}) of 0.0008 with 38 nodes selected from 115 in the Wnt signaling pathway (Figure 2A). The T cell receptor signaling pathway yielded a F^{err} of 0.0006 with 39 out of 113 nodes selected (Figure 2B), while the JAK-STAT signaling pathway showed a F^{err} of 0.0003 with 33 of 92 nodes selected (Figure 2C). In addition to model performance, the selected profiles help identify key targets for perturbation. These typically correspond to the earliest genes selected during the subset selection process and are thus prioritized for intervention.

In addition to the error rates, we also examined functions of the top-selected genes. In our application of NORSP, forward subset selection is used; thus, the prioritization is according to the order in which they are selected. In the Wnt signaling pathway, 18 of the 38 selected nodes were differentially expressed genes. The top three—MYC, DKK1, and PRKCA—are well-established components or downstream effectors of the pathway. MYC plays a central role in regulating genes essential for Wnt activation,[21–23] while DKK1 is a known inhibitor of Wnt signaling.[24,25] Notably, MYC has also been shown to suppress DKK1, promoting Wnt activation in breast cancer.[22,26] PRKCA encodes PKCα, which has been implicated in the negative regulation of Wnt signaling.[27] In the T-cell receptor signaling pathway, the top selected perturbation targets were CSF2, JUN, and MAPK9. Of these, CSF2 and JUN were among 10 enriched genes within the 39 selected nodes. In the JAK-STAT signaling pathway, SPRY1, MYC, and JAK1 emerged as top perturbation targets. All three were among 17 enriched genes out of the 33 selected. JAK1, in particular, is known to physically associate with the oncogene v-Abl, a key activator of the JAK-STAT pathway.[28]

NORSP was evaluated using metabolic models in the BioModel repository. To illustrate the methodology, we first present results from a single model before summarizing performance across all 61 systems. The metabolic model of the Calvin cycle (Figure 3A) was used to generate an influence network, serving as input to NORSP. The corresponding influence matrix, derived from the system's Jacobian, is shown in Figure 3B, with the perturbed node

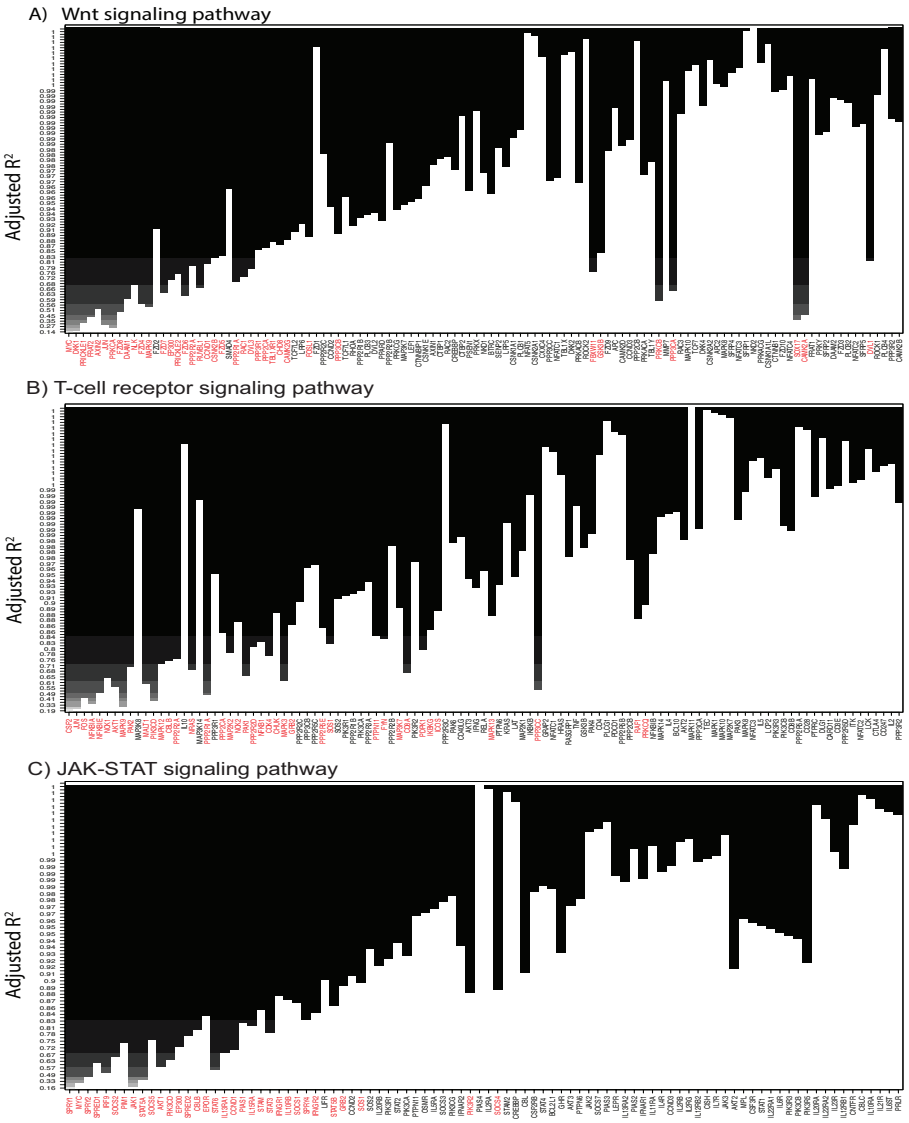

Fig. 2. NORSP results for YBX1 knockdown data, aiming to transition breast cancer cell lines (F^0) toward a healthy cell state (F^T). Subset selection results are shown for the top enriched KEGG pathways: **(A)** Wnt signaling, **(B)** MAPK signaling, and **(C)** JAK-STAT signaling. Adjusted R^2 is plotted as a function of the selected genes. Genes corresponding to the adjusted R^2 threshold are highlighted in red.

G6P, which was fixed at zero, circled in red. Using subset selection based on adjusted R^2, NORSP identified optimal perturbations across three types of influence networks: undirected, directed, and directed signed. In each case, the target node G6P was ranked among the top candidates (Figure 3C, shown in red). To better understand factors influencing network controllability, we examined structural properties of the BioModels under different graph representations—including node count, average degree, and clustering coefficients. However, these features did not consistently predict controllability outcomes (data not shown), suggesting

that control properties may depend on more complex, model-specific interactions.

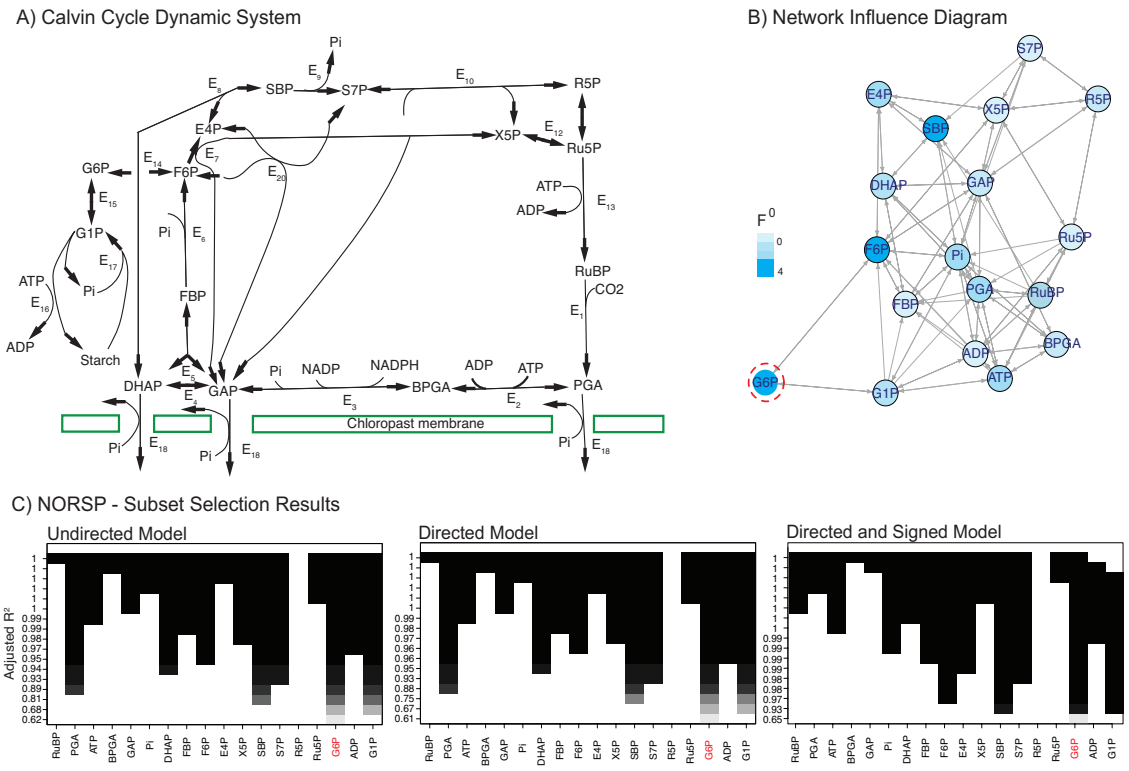

Fig. 3. An overview of NORSP applied to a single BioModel of the Calvin Cycle. **(A)** Schematic of the dynamic model representing the Calvin Cycle metabolism,[29] which is run to steady state and to obtain F^0, and perturbed to obtain a new target steady state, F^T. **(B)** The network influence diagram serves as the structural input to NORSP. The graph is extracted from the Jacobian and can be represented as undirected, directed, and signed. In this example, G6P is the perturbed node. **(C)** NORSP applied to the undirected, directed, and signed networks, and the corresponding model selection results using adjusted R^2 for target state approximation. The highest-ranked perturbation is G6P in all results, indicating that NORSP was able to identify the correct solution.

The results of the 61 BioModel simulations are summarized in Table 1. We evaluated graphical models derived from the influence matrix, which captures the network dynamics of each biological system, using the relative F error ratio. Three network propagation strategies were tested: undirected, directed, and directed signed. Model performance was assessed using two criteria: (1) adjusted R^2, and (2) the successful selection of the perturbed node among the top-ranked features, labeled as "target found." Under the adjusted R^2 criterion, 56, 57, and 55 models (out of 61) were correctly identified for the undirected, directed, and directed signed propagation methods, respectively (Table 1, Columns 1 and 2). This corresponds to 5, 4, and 6 misses, indicating a high success rate in selecting relevant perturbations across all three propagation strategies. We also examined the selected features to assess which models had identified the target as the top-ranked (Table 1, Column 3). We found that 48, 49, and

51 models correctly identified the perturbed node for the undirected, directed, and directed signed approaches, respectively. Finally, we counted the models for which the target was missed altogether, which were low 13, 12, and 10 (Table 1, Columns 4). These results highlight the robustness of the proposed framework in identifying meaningful perturbations across a broad range of biological models.

Table 1: Controllability results for the 61 Biomodels.
The F_{err}^{ctr} ratio is evaluates controllability (< 1) in the graphical (gm) and dynamical models (dyn).

Biomodel graph representation	Target found (# models)	Target ranked #1 (# models)	Target missed (# models)	Top hits in control graph ($F_{\text{gm}}^{\text{ctr}} <1$)	Top hits in control dyn. sys. ($F_{\text{dyn}}^{\text{ctr}} <1$)	Missed target in control dyn. sys. ($F_{\text{dyn}}^{\text{ctr}} <1$)
Undirected	56	48	13	48/48	35/48	9/13
Directed	57	49	12	49/49	33/49	11/12
Directed + Signed	55	51	10	51/51	36/51	10/10

A second mode of evaluation focused on system control, wherein the predicted perturbation was reintroduced into the system. Control was defined as the system's ability to approach the desired target state and was assessed both in the graphical model and the underlying dynamical system. When the target node was ranked as a top-priority candidate, control was consistently achieved across all models, as indicated by $F_{gm}^{ctr} < 1$ in the graphical model (Table 1, Column 5). Importantly, only the top-ranked perturbation of the selected set was introduced into the corresponding dynamical systems. System control was observed in the cases when the actual target was selected as the highest ranking (Table 1, Column 6). Notably, even in cases where the actual target was not ranked highest, the systems frequently remained controllable, especially when perturbations were derived from directed and signed-directed network models. This observation underscores the robustness of NORSP in identifying effective, experimentally actionable perturbations, even in the presence of uncertainty in target ranking.

To further assess NORSP's robustness, we simulated networks and introduced Gaussian noise either in the sensitivity matrix S or in the linear model. Table 2 summarizes the average prediction and perturbation errors across noise levels up to 10^{-1}. NORSP maintained low relative errors across both sparse and dense perturbation regimes when noise was confined to either the linear model or the sensitivity matrix. For small noise levels ($\epsilon \leq 10^{-3}$), both $\bar{F}_{\text{rel.err}}$ and $\bar{\Delta}_{\text{rel.err}}$ remained stable, typically below 0.01 and 0.07, respectively, demonstrating strong numerical stability. As noise increased to 10^{-1}, moderate increases in perturbation error were observed, most notably for the sensitivity matrix in the dense perturbation regime ($N_p = 25$), while target state errors remained low. These results indicate that NORSP is robust to moderate levels of measurement error and network misspecification, accurately recovering target states and perturbation patterns across a wide noise range.

Table 2: Noise added to the initial state of the system. Prediction errors of F_{target} were averaged over 100 simulations for networks with $N_n = 30$ nodes, under two perturbation regimes: sparse ($N_p = 5$) and dense ($N_p = 25$).

Noise-level ϵ	**Linear error**				**Sensitivity matrix error**			
	$N_p = 5$		$N_p = 25$		$N_p = 5$		$N_p = 25$	
	$\bar{F}^{\text{err}}$	$\bar{\Delta}^{\text{err}}$	$\bar{F}^{\text{err}}$	$\bar{\Delta}^{\text{err}}$	$\bar{F}^{\text{err}}$	$\bar{\Delta}^{\text{err}}$	$\bar{F}^{\text{err}}$	$\bar{\Delta}^{\text{err}}$
1e-07	0.007	0.069	0.023	0.171	0.006	0.053	0.023	0.174
1e-06	0.006	0.061	0.022	0.172	0.007	0.061	0.022	0.169
1e-05	0.007	0.060	0.022	0.166	0.006	0.062	0.022	0.168
1e-04	0.006	0.057	0.022	0.169	0.007	0.066	0.022	0.167
1e-03	0.000	0.065	0.022	0.178	0.007	0.065	0.022	0.165
1e-02	0.006	0.062	0.023	0.171	0.006	0.060	0.021	0.162
1e-01	0.006	0.056	0.022	0.164	0.007	0.288	0.024	0.257

4. Discussion

In this study, we introduced NORSP, a novel algorithm designed to identify sparse perturbations capable of guiding a system from an initial to a target steady state. By reframing the problem of network control as a subset selection task informed by network propagation, NORSP brings a practical approach to biological network optimization. It integrates graph-based sensitivity matrices with supervised learning, enabling direct prioritization of candidate perturbations without assuming full access to or manipulability of all network nodes.

A critical challenge in systems biology is to identify effective control strategies that are both interpretable and experimentally feasible. Traditional approaches, such as structural controllability theory,[30] offer important theoretical insights by identifying Minimum Driver Sets (MDSs)-the minimal number of nodes that must be controlled to steer the system. This framework has been widely applied in biological networks to classify nodes as critical, ordinary, or redundant,[31–33] and has informed analysis in domains such as metabolic regulation, disease gene identification, and signaling pathway dynamics.[34–36] Extensions of this theory have further characterized control modes and revealed emergent properties in complex systems.[37]

Despite its contributions, structural controllability also presents limitations. It often requires perturbation of a large proportion of nodes to achieve control, which is not compatible with experimental or clinical constraints.[38,39] It may also lack robustness in nonlinear systems, and typically does not estimate the magnitude or directionality of perturbations. These methods can be computationally expensive and may be less flexible in incorporating biological noise or measurement uncertainty.[40,41] Alternative strategies that involve iterative adjustment of system states, such as gradient-based approaches or local state feedback,[42,43] can also be computationally burdensome and sensitive to parameter initialization. NORSP is distinct from prior approaches that prioritize influential nodes (e.g., DYNAMO,[13] IGPON[14]) or rely on rich perturbation time-series data (e.g., CDN,[3] BaCaDI,[3] NETPERT[4]). Instead, NORSP operates in a realistic setting where only the network structure is available and perturbations must be experimentally actionable. Using gold-standard BioModels, knockdown data, and controlled

noise simulations, we showed that NORSP can recover true perturbations and effectively drive systems toward target states using only top-ranked candidates.

A significant strength of NORSP is the ability to find alternative paths towards perturbation when the true node target is not selected or is inaccessible. Simulation studies demonstrate that even when the actual perturbation is not among the top-ranked candidates, the system can still be driven toward the target state using alternate paths. This flexibility is particularly valuable in experimental contexts where not all network components are druggable or measurable. NORSP was also shown to be robust to misspecifications in the network structure and measurement error, which can be expected in models of biological systems. NORSP also expands the methodological toolkit for network control by using propagation-based sensitivity matrices as surrogates for the Jacobian. This allows the method to generalize across undirected, directed, and signed networks, accommodating the diverse structures found in biological systems. The use of subset selection aligns with a supervised learning framework, providing interpretability and scalability, especially when implemented with greedy forward selection or information criteria.

Nevertheless, several limitations warrant discussion. First, the propagation strength parameter must be manually tuned, as there is currently no theoretical guideline for optimal selection.[12,13] Second, while the subset selection framework simplifies the search space, the combinatorial nature of the optimization introduces heuristic dependencies. Furthermore, a threshold must be selected in subset selection according to model selection criteria. The use of forward subset selection can be viewed as modular, and alternatives could be considered. Future work could explore more sophisticated search strategies or Bayesian model averaging to enhance stability in parameter selection. Finally, an important limitation lies in the propagation assumption itself. Specifically, NORSP approximates the system Jacobian using a static network-derived sensitivity matrix. This simplification enables tractable optimization across large networks but does not capture nonlinear feedback, context-specific regulation, or time-dependent effects inherent to real biological dynamics. Future work could incorporate hybrid models that combine propagation-based inference with local dynamical refinement or explore adaptive sensitivity estimation from data.

Beyond theoretical utility, NORSP's targeted perturbation predictions may assist in prioritizing gene knockdowns or metabolic interventions for experimental validation. In particular, this framework could support early-stage drug screening by identifying mechanistically informative nodes whose control shifts replicate desired therapeutic outcomes. Its integration of sparse feasibility with network topology makes NORSP adaptable to both pathway-centric disease modeling and tissue-specific gene expression studies.

In summary, NORSP provides a principled yet practical framework for driving biological networks to target states using sparse, experimentally tractable interventions. By integrating ideas from network science, machine learning, and systems biology, it offers a new direction for hypothesis generation, experimental design, and therapeutic prioritization in complex biological systems. Its effectiveness across gene regulatory and metabolic networks highlights its versatility and potential as a tool for discovery.

References

1. T. Shi, H. Yu and R. H. Blair, Integrated regulatory and metabolic networks of the tumor microenvironment for therapeutic target prioritization, *Statistical Applications in Genetics and Molecular Biology* **22**, p. 20220054 (2023).
2. M. Wu, U. Padia, S. H. Murphy, R. Barzilay and T. Jaakkola, Identifying biological perturbation targets through causal differential networks, *arXiv preprint arXiv:2410.03380* (2024).
3. A. Hägele, J. Rothfuss, L. Lorch, V. R. Somnath, B. Schölkopf and A. Krause, Bacadi: Bayesian causal discovery with unknown interventions, in *International Conference on Artificial Intelligence and Statistics*, 2023.
4. M. C. Perrone, M. G. Lerner, M. Dunworth, A. J. Ewald and J. S. Bader, Prioritizing drug targets by perturbing biological network response functions, *PLoS Computational Biology* **20**, p. e1012195 (2024).
5. G. Baggio, D. S. Bassett and F. Pasqualetti, Data-driven control of complex networks, *Nature Communications* **12**, p. 1429 (2021).
6. S. Köhler, S. Bauer, D. Horn and P. N. Robinson, Walking the interactome for prioritization of candidate disease genes, *The American Journal of Human Genetics* **82**, 949 (2008).
7. E. O. Paull, D. E. Carlin, M. Niepel, P. K. Sorger, D. Haussler and J. M. Stuart, Discovering causal pathways linking genomic events to transcriptional states using tied diffusion through interacting events (tiedie), *Bioinformatics* **29**, 2757 (2013).
8. L. Cowen, T. Ideker, B. J. Raphael and R. Sharan, Network propagation: a universal amplifier of genetic associations, *Nature Reviews Genetics* **18**, 551 (2017).
9. G. Barel and R. Herwig, Netcore: a network propagation approach using node coreness, *Nucleic Acids Research* **48**, e98 (2020).
10. F. Vandin, E. Upfal and B. J. Raphael, Algorithms for detecting significantly mutated pathways in cancer, *Journal of Computational Biology* **18**, 507 (2011).
11. M. Hofree, J. P. Shen, H. Carter, A. Gross and T. Ideker, Network-based stratification of tumor mutations, *Nature Methods* **10**, p. 1108 (2013).
12. 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 (2010).
13. 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).
14. K. Krishnan, T. Shi, H. Yu and R. Blair, Integrated graph propagation and optimization with biological applications, in *Pacific Symposium on Biocomputing*, 2023.
15. 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).
16. M. Glont, T. V. N. Nguyen, M. Graesslin, R. Hälke, R. Ali, J. Schramm, S. M. Wimalaratne, V. B. Kothamachu, N. Rodriguez, M. J. Swat, R. Eils, H. Planatscher, E. Wingender, O. Wolkenhauer, H. Hermjakob, M. Hucka, N. Le Novere and R. S. Malik-Sheriff, Biomodels: expanding horizons to include more modelling approaches and formats, *Nucleic Acids Research* **46**, D1248 (2018).
17. M. Newman, *Networks* (Oxford university press, 2018).
18. 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 (2010).
19. A. Miller, *Subset selection in regression* (chapman and hall/CRC, 2002).
20. B. J. Bornstein, S. M. Keating, A. Jouraku and M. Hucka, LibSBML: an API library for SBML, *Bioinformatics* **24**, 880 (2008).

21. Y.-H. Hao, M. C. Lafita-Navarro, L. Zacharias, N. Borenstein-Auerbach, M. Kim, S. Barnes, J. Kim, J. Shay, R. J. DeBerardinis and M. Conacci-Sorrell, Induction of lef1 by myc activates the wnt pathway and maintains cell proliferation, *Cell Communication and Signaling* **17**, 1 (2019).
22. V. H. Cowling and M. D. Cole, Turning the tables: Myc activates wnt in breast cancer, *Cell Cycle* **6**, 2625 (2007).
23. B. Madan, N. Harmston, G. Nallan, A. Montoya, P. Faull, E. Petretto, D. M. Virshup *et al.*, Temporal dynamics of WNT-dependent transcriptome reveal an oncogenic wnt/myc/ribosome axis, *The Journal of Clinical Investigation* **128**, 5620 (2019).
24. A. Niida, T. Hiroko, M. Kasai, Y. Furukawa, Y. Nakamura, Y. Suzuki, S. Sugano and T. Akiyama, Dkk1, a negative regulator of wnt signaling, is a target of the β-catenin/tcf pathway, *Oncogene* **23**, 8520 (2004).
25. V. H. Cowling, C. M. D'Cruz, L. A. Chodosh and M. D. Cole, c-myc transforms human mammary epithelial cells through repression of the Wnt inhibitors DKK1 and SFRP1, *Molecular and Cellular Biology* **27**, 5135 (2007).
26. J. Li, I. Sarosi, R. C. Cattley, J. Pretorius, F. Asuncion, M. Grisanti, S. Morony, S. Adamu, Z. Geng, W. Qiu, P. Kostenuik, D. L. Lacey, W. S. Simonet, B. Bolon, X. Qian, V. Shalhoub, M. S. Ominsky, H. Z. Ke, X. Li and W. G. Richards, Dkk1-mediated inhibition of wnt signaling in bone results in osteopenia, *Bone* **39**, 754 (2006).
27. J. Gwak, M. Cho, S. Gong, J. Won, D. Kim, E. Kim, S. S. Lee, M. Kim, T. K. Kim, J. Shin and S. Oh, Protein-kinase-c-mediated β-catenin phosphorylation negatively regulates the wnt/β-catenin pathway, *Journal of Cell Science* **119**, 4702 (2006).
28. N. N. Danial, A. Pernis and P. B. Rothman, Jak-stat signaling induced by the v-abl oncogene, *Science* **269**, 1875 (1995).
29. M. Poolman, H. Assmus and D. Fell, Applications of metabolic modelling to plant metabolism, *Journal of Experimental Botany* **55**, 1177 (2004).
30. C.-T. Lin, Structural controllability, *IEEE Transactions on Automatic Control* **19**, 201 (1974).
31. Y.-Y. Liu, J.-J. Slotine and A.-L. Barabási, Controllability of complex networks, *Nature* **473**, 167 (2011).
32. S. Nie, X. Wang, H. Zhang, Q. Li and B. Wang, Robustness of controllability for networks based on edge-attack, *PloS One* **9**, p. e89066 (2014).
33. C.-L. Pu, W.-J. Pei and A. Michaelson, Robustness analysis of network controllability, *Physica A: Statistical Mechanics and its Applications* **391**, 4420 (2012).
34. E. Kawakami, V. K. Singh, K. Matsubara, T. Ishii, Y. Matsuoka, T. Hase, P. Kulkarni, K. Siddiqui, J. Kodilkar, N. Danve, I. Subramanian, M. Katoh, Y. Shimizu-Yoshida, S. Ghosh, A. Jere and H. Kitano, Network analyses based on comprehensive molecular interaction maps reveal robust control structures in yeast stress response pathways, *NPJ Systems Biology and Applications* **2**, 1 (2016).
35. A. Vinayagam, T. E. Gibson, H.-J. Lee, B. Yilmazel, C. Roesel, Y. Hu, Y. Kwon, A. Sharma, Y.-Y. Liu, N. Perrimon and A. Barabasi, Controllability analysis of the directed human protein interaction network identifies disease genes and drug targets, *Proceedings of the National Academy of Sciences* **113**, 4976 (2016).
36. V. Ravindran, V. Sunitha and G. Bagler, Identification of critical regulatory genes in cancer signaling network using controllability analysis, *Physica A: Statistical Mechanics and its Applications* **474**, 134 (2017).
37. T. Jia, Y.-Y. Liu, E. Csóka, M. Pósfai, J.-J. Slotine and A.-L. Barabási, Emergence of bimodality in controlling complex networks, *Nature Communications* **4**, 1 (2013).
38. X. Liu and L. Pan, Detection of driver metabolites in the human liver metabolic network using structural controllability analysis, *BMC Systems Biology* **8**, 1 (2014).

39. V. Ravindran, J. C. Nacher, T. Akutsu, M. Ishitsuka, A. Osadcenco, V. Sunitha, G. Bagler, J.-M. Schwartz and D. L. Robertson, Network controllability: Viruses are driver agents in dynamic molecular systems, *bioRxiv* , p. 311746 (2018).
40. M. Li, H. Gao, J. Wang and F.-X. Wu, Control principles for complex biological networks, *Briefings in Bioinformatics* **20**, 2253 (2019).
41. R. M. D'Souza, M. di Bernardo and Y.-Y. Liu, Controlling complex networks with complex nodes, *Nature Reviews Physics* **5**, 250 (2023).
42. S. P. Cornelius, W. L. Kath and A. E. Motter, Realistic control of network dynamics, *Nature Communications* **4**, p. 1942 (2013).
43. M. R. Yousefi and E. R. Dougherty, Intervention in gene regulatory networks with maximal phenotype alteration, *Bioinformatics* **29**, 1758 (2013).