Computational Modeling of Host-Pathogen Interaction Networks for Identifying Multi-Target Antiviral Therapeutic Candidates

Shona McKinnon
Department of Electrical Engineering and Computer Science, University of Kansas, Lawrence, KS, USA.
mckinnon569@ku.edu

Federigo Manacorda
Department of Computer Science, University of Alabama at Birmingham, Birmingham, AL, USA.
hellofederigo@uab.edu

Hans-Detlef Fechner
Department of Computer Science, University of North Texas, Denton, TX, USA.
hellohansdetlef@unt.edu

Abstract

Antiviral therapeutic development faces a persistent mismatch between the reductionist logic of single-target inhibition and the networked, adaptive behavior of host-pathogen systems. Viruses frequently exploit redundant host pathways, evade direct-acting agents through mutation, and produce context-dependent effects that are poorly captured by target-centric ranking. This paper develops a computational modeling framework for host-pathogen interaction networks that converts heterogeneous interaction evidence into a weighted bipartite network, propagates pathogen-induced perturbations through the host protein interaction system, and formalizes multi-target antiviral selection as a constrained multi-objective problem. The framework combines evidence-aware edge construction, network diffusion, residual perturbation estimation, target-engagement modeling, community-level coverage, and first-order uncertainty propagation. Candidate combinations are evaluated not only by predicted neutralization of pathogen-induced host perturbation but also by off-target burden, combination size, and robustness to interaction uncertainty. A robust decision rule selects target sets that perform well across plausible edge-weight perturbations, reducing dependence on a single reconstructed interactome. The central contribution is a mechanistically interpretable selection logic that connects host-pathogen network topology to combination therapy design rather than merely ranking individual proteins. The framework is intended to support hypothesis generation for host-directed antiviral combinations, provide a defensible basis for prioritizing experimental tests, and clarify where network incompleteness, context specificity, and causal ambiguity limit computational inference.

Keywords

host-pathogen interaction networks; network medicine; multi-target antivirals; network propagation; multi-objective optimization; drug repurposing

1. Introduction

Network biology has established that cellular functions emerge from organized interactions rather than from isolated molecular entities [1]. Systems-level approaches therefore treat proteins, genes, and metabolites as components of a structured interaction system whose topology constrains information flow, redundancy, and response to perturbation [2]. In parallel, high-throughput protein interaction maps have made it possible to reconstruct host-pathogen interfaces at increasing scale, although these maps remain uneven in coverage and confidence [3]. Specialized resources for virus-host interactions have consolidated such evidence and made proteome-wide host-pathogen networks accessible for computational analysis [4]. Drug-target networks add a therapeutic layer by linking compounds to their known or predicted molecular targets, revealing that many successful drugs act through multiple targets rather than through a single dominant node [5]. Network pharmacology extends this observation into a design principle: therapeutic effects can be understood as perturbations of networked systems, not merely as local inhibition events [6]. Network medicine further connects disease modules, pathobiological mechanisms, and drug action within a shared topological language [7].

These ideas are directly relevant to antiviral discovery. Multi-target interventions are attractive when pathogens exploit redundant host functions, when resistance mutations reduce the effectiveness of direct-acting antivirals, or when a single virus protein interfaces with several host subsystems [8]. Network-based function prediction and propagation methods provide ways to infer the consequences of perturbing one component across a neighborhood of interacting proteins [9,10]. Control-theoretic analyses of complex networks clarify why a small number of strategically placed interventions can, in principle, reshape system dynamics, while also showing that topological centrality alone does not determine controllability [11]. Studies of synthetic rescue demonstrate that compensatory network relationships can be exploited to identify combinations that restore function when single perturbations fail [12]. Multi-target drug discovery has developed complementary arguments for combination design in complex diseases, emphasizing balanced modulation of multiple nodes and the need to control toxicity [13]. The COVID-19 pandemic accelerated the construction of severe acute respiratory syndrome coronavirus 2 host interaction maps and drug-repurposing hypotheses [14], as well as integrated virus-host-drug networks that connect pathogen proteins, host factors, and candidate compounds [15].

A specific gap remains between network reconstruction and combination selection. Many analyses produce ranked lists of host factors or individual drug targets, but they do not explicitly formalize how a set of targets should be chosen to neutralize a distributed pathogen-induced perturbation under uncertainty. Ranking methods also tend to conflate topological importance with therapeutic value, even though a highly connected host protein may be essential, toxic, or poorly druggable. What is needed is a modeling logic that treats a multi-target antiviral candidate as a constrained intervention on a host-pathogen network, evaluates residual perturbation after partial target engagement, accounts for off-target burden, and admits uncertainty in interaction evidence. This paper develops such a framework. Its contribution is not an empirical claim about a particular virus or compound, but a technically specified computational model for hypothesis generation and candidate prioritization. The framework integrates evidence-weighted network construction, diffusion-based perturbation analysis, multi-objective target-set selection, community-level coverage, and robustness analysis. It is intended to make assumptions explicit, separate mechanistic inference from

observed evidence, and provide a reproducible conceptual basis for subsequent experimental evaluation.

2. Modeling Framework for Host-Pathogen Interaction Networks

2.1 Evidence-weighted network construction

The input is a host protein interaction network and a set of pathogen proteins that physically or functionally interface with host components. Let H denote host proteins and P denote pathogen proteins. Each candidate host-pathogen edge (i,j) with i in H and j in P may be supported by several evidence channels, such as affinity purification, yeast two-hybrid assays, literature curation, or computational prediction. Because these channels differ in reliability and false-positive behavior [3,4], raw binary edges should not be treated as equivalent. A reliability-weighted noisy-OR model provides a transparent construction rule:

$$w_{ij} = 1 - \prod_k \left(1 - a_{ij}^{(k)} r_k\right)$$

Here, w_{ij} is the integrated confidence that host protein i interacts with pathogen protein j; $a_{ij}^{(k)}$ is a binary indicator that evidence channel k supports the edge; and r_k is the channel-specific reliability in the interval zero to one. The product runs over all evidence channels. The noisy-OR form assumes that independent evidence sources can jointly increase confidence while no single source is treated as definitive. This is preferable to simple edge union because it preserves the distinction between an interaction observed in one assay and an interaction corroborated across several. The resulting matrix W defines a weighted bipartite host-pathogen network. Host-host interactions are represented separately by a weighted adjacency matrix A, with entries reflecting confidence in physical or functional association. Both matrices are treated as evidence-bearing reconstructions, not as complete descriptions of the infected cell.

2.2 Perturbation propagation through the host network

A virus does not perturb host physiology only at direct binding sites. Pathogen proteins can alter host complexes, signaling modules, and regulatory neighborhoods, so the functional impact of an interaction must be propagated through host connectivity [10]. Let p be a vector of pathogen-induced seed perturbations on host proteins. A natural seed value is $p_i = \sum_j w_{ij}$, which aggregates pathogen interactions incident on host protein i. To model how those seeds spread through the host network, the normalized host adjacency A_n is used in a resolvent diffusion:

$$y = (I - \beta A_n)^{-1} p$$

In this expression, y is the vector of propagated host perturbation scores; I is the identity matrix; β is a diffusion parameter; and A_n is the row- or symmetric-normalized host adjacency matrix. The resolvent is well defined when β is smaller than the reciprocal of the spectral radius of A_n. Larger β values produce broader propagation, whereas smaller values concentrate impact near direct interaction sites. The model assumes approximately linear, stationary propagation during the initial perturbation regime. That assumption is a deliberate simplification: it permits interpretable attribution from pathogen seeds to host neighborhoods without requiring time-resolved infection data. Nonlinear or dynamic models can be layered onto the same network representation, but the linear resolvent provides a defensible first-order screening device.

2.3 Multi-target intervention model

A multi-target antiviral candidate is a set T of host or pathogen targets whose modulation is intended to reduce pathogen-induced host perturbation. For each target j and host protein i, let c_{ij} represent the fraction of target j's effect that acts on host node i, and let q_j represent an attainable engagement or efficacy parameter for target j in the relevant tissue context. Both are bounded between zero and one. The residual host perturbation after applying the target set T is modeled as:

$$R(T) = \sum_i y_i^2 \prod_j \in T\left(1 - c_{ij} q_j\right)$$

The squared propagated perturbation y_i^2 gives greater weight to host proteins whose pathogen-induced deviation is large. The product term reduces the residual contribution of host node i when at least one selected target can engage a pathway affecting that node. This formulation reflects partial and overlapping target effects: a node is not switched fully off unless a target has complete engagement and complete coverage, and multiple targets can compensate for incomplete single-target activity. The model does not assume that target effects are independent in a biological sense; rather, it uses a multiplicative residual form as a conservative approximation for overlapping neutralization of a shared host perturbation.

Candidate selection should not minimize residual perturbation alone. A combination that requires many targets, or that engages host proteins with broad physiological roles, may have unacceptable toxicity. Let O(T) be an off-target or host-toxicity burden score assembled from expression breadth, essentiality, known adverse-effect associations, and tissue-specific constraint. Let R_0 be the residual perturbation without intervention, and let |T| be the number of selected targets. A scalar utility for a candidate set can be written as:

$$U(T) = R_0 - R(T) - \lambda O(T) - \mu |T|$$

Here, λ and μ are nonnegative trade-off parameters that penalize off-target burden and combination size, respectively. The equation is an objective function for candidate ranking, not a fitted clinical model. Its parameters should be chosen according to a prespecified decision policy, such as a maximum tolerated off-target burden or a preferred combination size, and their sensitivity should be reported. The optimization is subject to a cardinality constraint $|T| \leq K$, where K is the maximum number of targets considered practical for a development program, and to eligibility constraints that exclude targets with insufficient evidence, unacceptable essentiality, or no plausible modulation strategy.

2.4 Community-level coverage and combination synergy

Pathogen-induced perturbations often concentrate in host modules rather than distributing uniformly across the interactome. Community structure therefore offers a useful intermediate level between single proteins and the whole network. Let C_c denote the c-th host community or functional module, and let κ_c be the number of target-accessible nodes in that community that are needed to achieve meaningful coverage. A community coverage score can be defined as:

$$S(T) = \sum c \, min(1, |T \cap C_c|/\kappa_c) - \theta \sum_j \in T \, b_j$$

In this expression, |T∩C_c| is the number of selected targets in community c; b_j is a target-specific penalty for broad or undesirable host effects; and θ scales that penalty. The min function saturates coverage, preventing a combination from receiving disproportionate

credit for over-targeting one module while leaving other perturbed modules untouched. Community-based synergy models for complex systems, including network-community approaches to combination identification [16], provide a relevant conceptual precedent: combination value can arise from coordinated coverage of distinct system components rather than from the sum of individual target strengths. The present formulation does not assert a specific biological synergy mechanism. It operationalizes the weaker and more defensible claim that a multi-target set should be evaluated by how well it covers perturbed host modules while controlling collateral burden.

2.5 Uncertainty propagation and robust selection

Host-pathogen networks contain false positives, false negatives, and context-dependent edges. A candidate ranking that changes drastically under small edge perturbations is not a reliable basis for experimental prioritization. Suppose each edge weight w_e has an uncertainty σ_e. A first-order approximation of the variance of propagated host perturbation is:

$$Var(y_i) \approx \sum_e (\partial y_i/\partial w_e)^2 \, {\sigma_e}^2$$

The sum runs over network edges, and $\partial y_i/\partial w_e$ is the sensitivity of the propagated score at host node i to edge e. This approximation assumes approximately independent edge errors and a differentiable propagation regime. It is most appropriate for ranking uncertainty under modest perturbations, not for exact probabilistic inference. The sensitivity terms can be obtained from the resolvent expression or approximated numerically by edge-weight resampling. Candidate selection can then use a risk-adjusted decision rule:

$$T* = argmax_T \{E[U(T)] - \eta \sqrt{Var[U(T)]}\}$$

Here, E[U(T)] is the expected utility under the edge-uncertainty model; Var[U(T)] is the corresponding utility variance; and η controls risk aversion. Larger η favors target sets whose predicted advantage persists across plausible network reconstructions. This robust rule is preferable to selecting the single highest-scoring set from one reconstructed network because it makes uncertainty explicit and reduces overconfidence in sparse interaction evidence.

3. Candidate Prioritization and Decision Logic

The computational workflow begins with assembly of host-host and host-pathogen evidence, followed by reliability-weighted edge construction and normalization. Seed perturbation scores are computed from pathogen interaction weights and propagated through the host network. Candidate targets are then filtered by evidence quality and biological constraints before combination search. Because exhaustive enumeration of all target sets becomes infeasible as the candidate pool grows, the selection problem can be solved by integer programming for the linearized objective or by greedy, beam-search, or Pareto-based heuristics when nonlinear community terms are included. The search should retain a Pareto frontier rather than a single scalar optimum, because the trade-offs among residual perturbation, off-target burden, combination size, and robustness are policy choices that require transparent reporting.

Several constraints are essential to avoid biologically implausible recommendations. Host targets with broad essentiality or systemic roles should be penalized unless the therapeutic context justifies transient or tissue-restricted modulation. Targets without a plausible modulation route, such as host proteins lacking suitable binding pockets or available chemical matter, should receive lower priority even if their network impact is high. Pathogen targets should be evaluated for conservation and resistance potential, because a multi-target strategy

that concentrates on a single mutable pathogen protein may retain the weakness of direct-acting monotherapy. The model should also distinguish between targets whose engagement reduces pathogen-induced perturbation and targets that merely occupy central network positions. Centrality can be informative, but it is not equivalent to therapeutic efficacy [11].

Sensitivity analysis should accompany every ranking. At minimum, the analysis should examine changes in the selected target set under edge-weight perturbation, reliability-channel removal, alternative diffusion strengths, different cardinality limits, and alternate off-target penalties. A target that remains in the robust set across these conditions has stronger computational support than a target that appears only under a narrow parameterization. The uncertainty terms in equations (6) and (7) are intended to make this assessment part of the objective rather than an afterthought. The output is not a validated drug combination. It is a ranked, mechanistically annotated set of candidate interventions whose assumptions can be examined and whose predictions can be tested experimentally.

4. Technical Discussion

The proposed framework differs from purely target-centric scoring in that it evaluates combinations as interventions on a distributed perturbation. Single-target ranking can be obtained as a special case when K equals one, but the multi-target formulation exposes trade-offs that are hidden in individual rankings. It also differs from supervised drug-repurposing models that learn from existing compound-disease associations. Such models can be powerful when large labeled training sets are available, but they depend on historical discovery patterns and may reproduce existing biases. The present framework is mechanistic and assumption-driven: it specifies how evidence is weighted, how perturbations diffuse, how target effects combine, and how uncertainty enters selection. This makes it suitable for early-stage hypothesis generation and for comparing competing biological assumptions, even when labeled antiviral combination data are sparse. Network-based repurposing studies for coronaviruses have demonstrated the practical value of connecting host interactomes to candidate drugs [17], but the selection of target combinations still requires an explicit decision rule of the kind developed here.

The model's biological credibility depends on the quality and context specificity of its inputs. Viral proteins are expressed at different times, in different subcellular compartments, and at different abundances depending on cell type and infection stage. A static network aggregates these conditions and may therefore overstate the relevance of interactions that occur only in one setting. Multilevel proteomic studies of SARS-CoV-2 and SARS-CoV have shown that host perturbations extend across protein abundance, phosphorylation, and interaction states [18], which means that a single physical interaction map captures only part of the relevant biology. The framework can incorporate context-specific weights when transcriptomic or proteomic data are available, but it should not pretend that a generic network is a complete infection model. Where context data are absent, the output should be interpreted as a hypothesis about possible mechanisms rather than as a prediction of clinical response.

A further limitation concerns causal direction. Host-pathogen edges may represent binding, modification, dependence, or indirect functional association. Viral immune modulators can perturb host networks through common and unique strategies [19], and an edge that is statistically associated with infection may not be a productive intervention point. The propagation model treats edges as conduits of perturbation, which is appropriate for screening but not sufficient for causal inference. Candidate targets should therefore be

annotated with the type of evidence supporting their role and with the expected direction of modulation. Inhibiting a host factor that supports viral replication is mechanistically different from activating an antiviral restriction factor, and the residual perturbation formula must be parameterized accordingly. This distinction prevents the model from treating all network edges as equivalent and allows experimental follow-up to test the proposed direction of effect.

Finally, the framework should be evaluated against credible alternatives rather than against no baseline. Useful comparisons include degree-based ranking, betweenness-based ranking, disease-module overlap, diffusion-only ranking, and unconstrained multi-target selection. These comparisons do not require clinical claims; they test whether the added constraints and uncertainty logic change candidate sets in interpretable ways. Network medicine frameworks for COVID-19 drug repurposing have shown how proximity and module relationships can generate testable hypotheses [20], and the present approach extends that logic to the specific problem of combination selection under partial engagement and off-target burden. The most informative validation would combine target perturbation, host-pathogen interaction assays, and functional infection measurements in relevant cell systems. Such validation is outside the scope of the present modeling paper, but the framework is designed to produce falsifiable predictions that can be tested in those settings.

5. Conclusion

This paper has developed a computational modeling framework for identifying multi-target antiviral therapeutic candidates from host-pathogen interaction networks. The framework treats interaction evidence as weighted and uncertain, propagates pathogen-induced perturbations through the host protein network, and defines a multi-target candidate as a constrained intervention that reduces residual host perturbation while penalizing off-target burden and combination size. Community-level coverage introduces an intermediate organizational layer between single targets and whole-network effects, and first-order uncertainty propagation supports a risk-adjusted selection rule. The central contribution is a coherent selection logic that connects network topology, partial target engagement, toxicity constraints, and evidence uncertainty. The framework does not claim that any specific target set is clinically effective; it provides a transparent method for generating and prioritizing hypotheses. Its practical value lies in making assumptions explicit, separating mechanistic inference from observed evidence, and producing candidate combinations that can be subjected to focused experimental testing. By treating antiviral combination design as a network intervention problem rather than a ranking exercise, the approach offers a more defensible computational foundation for host-directed therapeutic discovery.

References

1. Barabási, A.-L., & Oltvai, Z. N. (2004). Network biology: Understanding the cell's functional organization. Nature Reviews Genetics, 5(2), 101–113. https://doi.org/10.1038/nrg1272

2. Kitano, H. (2002). Computational systems biology. Nature, 420(6912), 206–210. https://doi.org/10.1038/nature01254

3. von Mering, C., Krause, R., Snel, B., Cornell, M., Oliver, S. G., Fields, S., & Bork, P. (2002). Comparative assessment of large-scale data sets of protein–protein interactions. Nature, 417(6887), 399–403. https://doi.org/10.1038/nature750

4. Navratil, V., de Chassey, B., Meyniel, L., Delmotte, S., Gautier, C., André, P., Lotteau, V., & Rabourdin-Combe, C. (2009). VirHostNet: A knowledge base for the management and the analysis of proteome-wide virus–host interaction networks. Nucleic Acids Research, 37(Database issue), D661–D668. https://doi.org/10.1093/nar/gkn794

5. Yildirim, M. A., Goh, K.-I., Cusick, M. E., Barabási, A.-L., & Vidal, M. (2007). Drug-target network. Nature Biotechnology, 25(10), 1119–1126. https://doi.org/10.1038/nbt1338

6. Hopkins, A. L. (2008). Network pharmacology: The next paradigm in drug discovery. Nature Chemical Biology, 4(11), 682–690. https://doi.org/10.1038/nchembio.118

7. Barabási, A.-L., Gulbahce, N., & Loscalzo, J. (2011). Network medicine: A network-based approach to human disease. Nature Reviews Genetics, 12(1), 56–68. https://doi.org/10.1038/nrg2918

8. Csermely, P., Korcsmáros, T., Kiss, H. J. M., London, G., & Nussinov, R. (2013). Structure and dynamics of molecular networks: A novel paradigm of drug discovery. Pharmacology & Therapeutics, 138(3), 333–408. https://doi.org/10.1016/j.pharmthera.2013.01.016

9. Sharan, R., Ulitsky, I., & Shamir, R. (2007). Network-based prediction of protein function. Molecular Systems Biology, 3, 88. https://doi.org/10.1038/msb4100129

10. Cowen, L., Ideker, T., Raphael, B. J., & Sharan, R. (2017). Network propagation: A universal amplifier of genetic associations. Nature Reviews Genetics, 18(9), 551–562. https://doi.org/10.1038/nrg.2017.38

11. Liu, Y.-Y., Slotine, J.-J., & Barabási, A.-L. (2011). Controllability of complex networks. Nature, 473(7346), 167–173. https://doi.org/10.1038/nature10011

12. Motter, A. E., Gulbahce, N., Almaas, E., & Barabási, A.-L. (2008). Predicting synthetic rescues in metabolic networks. Molecular Systems Biology, 4, 168. https://doi.org/10.1038/msb.2008.1

13. Ramsay, R. R., Popovic-Nikolic, M. R., Nikolic, K., Uliassi, E., & Bolognesi, M. L. (2018). A perspective on multi-target drug discovery and design for complex diseases. Clinical and Translational Medicine, 7, 3. https://doi.org/10.1186/s40169-017-0181-2

14. Gordon, D. E., Jang, G. M., Bouhaddou, M., Xu, J., Obernier, K., White, K. M., O'Meara, M. J., Rezelj, V. V., Guo, J. Z., Swaney, D. L., Tummino, T. A., Hüttenhain, R., Kaake, R. M., Richards, A. L., Tutuncuoglu, B., Foussard, H., Batra, J., Haas, K., Modak, M., ... Krogan, N. J. (2020). A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. Nature, 583(7816), 459–468. https://doi.org/10.1038/s41586-020-2286-9

15. Sadegh, S., Matschinske, J., Blumenthal, D. B., Galindez, G., Kacprowski, T., List, M., Nasirigerdeh, R., Oubounyt, M., Pichlmair, A., Rose, T. D., Salgado-Albarrán, M., Späth, J., Stukalov, A., Wenke, N. K., Yuan, K., Pauling, J. K., & Baumbach, J. (2020). Exploring the SARS-CoV-2 virus-host-drug interactome for drug repurposing. Nature Communications, 11, 3518. https://doi.org/10.1038/s41467-020-17189-2

16. HerbSyner_Finder: a network community-based model for identifying synergistic combinations from herbal medicines and complex systems. Targetome. 2026;2(2):0-0. doi: 10.48130/targetome-0026-001

17. Zhou, Y., Hou, Y., Shen, J., Huang, Y., Martin, W., & Cheng, F. (2020). Network-based drug repurposing for novel coronavirus 2019-nCoV/SARS-CoV-2. Cell Discovery, 6, 14. https://doi.org/10.1038/s41421-020-0153-3

18. Stukalov, A., Girault, V., Grass, V., Karayel, O., Bergant, V., Urban, C., Haas, D. A., Huang, Y., Oubraham, L., Wang, A., Hamad, M. S., Piras, A., Hansen, F. M., Tanzer, M., Pichlmair, A., & Superti-Furga, G. (2021). Multilevel proteomics reveals host perturbations by SARS-CoV-2 and SARS-CoV. Nature, 594(7862), 246–252. https://doi.org/10.1038/s41586-021-03493-4

19. Pichlmair, A., Kandasamy, K., Alvisi, G., Mulhern, O., Sacco, R., Habjan, M., Binder, M., Stefanovic, A., Eberle, C. A., Goncalves, A., Buerckstuemmer, T., Binder, B., Baumann, C. L., & Superti-Furga, G. (2012). Viral immune modulators perturb the human molecular network by common and unique strategies. Nature, 487(7408), 486–490. https://doi.org/10.1038/nature11289

20. Gysi, D. M., Do Valle, Í., Zitnik, M., Ameli, A., Gan, X., Varol, O., Ghiassian, S. D., Patten, J. J., Davey, R. A., Loscalzo, J., & Barabási, A.-L. (2021). Network medicine framework for identifying drug-repurposing opportunities for COVID-19. Proceedings of the National Academy of Sciences of the United States of America, 118(19), e2025581118. https://doi.org/10.1073/pnas.2025581118