A Graph-Based Systems Biology Framework for Prioritizing Multi-Compound Interventions in Complex Disease Mechanisms

Cyprian Gajowiak
Department of Computer Science and Engineering, University at Buffalo, Buffalo, NY, USA.
cyprian27@buffalo.edu

Mieszko Polakiewicz
Department of Computer Science, Binghamton University, Binghamton, NY, USA.
mieszkop@binghamton.edu

Wolfram Ruppersberger
Department of Electrical Engineering and Computer Science, University of Kansas, Lawrence, KS, USA.
contactwolfram@ku.edu

Abstract

Complex diseases are rarely explained by a single molecular lesion, and interventions that modulate one target often produce limited or transient benefit. Network and systems-biology approaches have clarified how disease phenotypes emerge from distributed molecular interactions, but the prioritization of multi-compound interventions remains methodologically fragmented. Pairwise synergy models are not designed to evaluate mechanism coverage across several disease modules, while single-compound network methods do not represent combination effects, redundancy, or cumulative safety burden. This manuscript develops a graph-based systems-biology framework that treats multi-compound prioritization as a constrained, mechanism-aware ranking problem. The framework integrates compounds, molecular targets, pathways, and disease modules into a heterogeneous intervention graph; converts heterogeneous evidence into calibrated edge weights; propagates compound-set effects through target engagement and module perturbation; and scores candidate sets using mechanism coverage, redundancy penalties, and off-target burden. Synergy is represented as complementarity at the target and module level rather than as an isolated pairwise phenotype. Uncertainty is propagated from edge evidence to candidate rankings, and the final decision rule balances expected benefit against evidence variance and model complexity. Analytical assessment shows where the framework is likely to outperform target-centric or signature-only strategies, especially when disease mechanisms are distributed and combinations must avoid overlapping or antagonistic effects. It also identifies boundary conditions involving module resolution, topological bias, dose context, and causal interpretation. The contribution is a transparent prioritization architecture for preclinical and translational hypothesis generation, not a validated clinical decision tool.

Keywords

graph-based systems biology; multi-compound interventions; network pharmacology; disease modules; synergy prioritization; uncertainty propagation

1. Introduction

Complex diseases such as metabolic syndrome, inflammatory disorders, and neurodegenerative conditions involve multiple interacting molecular processes. The systems-biology view treats disease as a perturbation of networked cellular functions rather than as a defect in one isolated molecule [1], [2]. Network pharmacology extended this perspective to drug discovery by linking compounds to targets, targets to pathways, and pathways to phenotypic outcomes [3]. Drug-target networks showed that approved compounds often act on several proteins, and that target families are not randomly distributed across the druggable genome [4]. Molecular network analyses further demonstrated that disease-related genes tend to cluster in local neighborhoods and that network proximity can inform drug efficacy and adverse-effect prediction [5].

These advances motivate multi-compound interventions, but they do not by themselves solve the prioritization problem. Network-based combination prediction has used target proximity, pathway overlap, and expression reversal to rank candidate pairs and larger sets [6], [7]. Classical synergy models provide quantitative frameworks for detecting departure from additive effects, including the median-effect principle, Loewe additivity, and Bliss independence [8], [9], [10]. Such models are essential for experimental combination studies, yet they describe observed dose-response behavior rather than the distributed mechanism structure that makes a combination biologically plausible before testing. A compound pair can be statistically synergistic in one cell context while acting through redundant pathways, poor target coverage, or unacceptable off-target burden. Conversely, a combination with modest pairwise synergy may be valuable because it covers several disease modules while preserving safety margins.

The gap addressed here is the absence of a coherent graph-based prioritization framework that connects molecular mechanism, combination effect, safety burden, and evidence uncertainty for multi-compound interventions. Most existing strategies optimize one of these dimensions. Target-centric approaches rank compounds by affinity or pathway annotation but do not represent set-level complementarity. Expression-signature approaches align compound-induced transcriptional changes with disease signatures but can conflate mechanism with phenotypic similarity. Pairwise synergy models are experimentally interpretable but scale poorly to higher-order sets and depend on context-specific dose matrices. A framework is needed that can rank sets of compounds before experimental screening while making its assumptions, failure modes, and uncertainty explicit.

This manuscript proposes a graph-based systems-biology framework for prioritizing multi-compound interventions in complex disease mechanisms. The central contribution is a mechanism-aware ranking architecture in which compounds, targets, pathways, and disease modules form a heterogeneous graph; compound-set effects are propagated through target engagement and module perturbation; synergy is represented as nonredundant mechanism coverage rather than only as a pairwise interaction term; and candidate sets are optimized under safety, size, and evidence-uncertainty constraints. The framework is deliberately analytical: it specifies how evidence should be structured and combined, which quantities determine ranking, and where inferential limits arise. It is intended to support preclinical hypothesis generation and combination repurposing, not to substitute for experimental dose-response validation or clinical evidence.

2. Biological and Formal Problem Definition

A disease module is a connected or functionally coherent set of molecules whose collective perturbation contributes to a disease phenotype. Modules may be derived from protein interaction neighborhoods, co-expression clusters, pathway annotations, or genome-wide association signals. The module concept is useful because it converts a long list of disease-associated genes into a smaller number of mechanistic units that can be evaluated for intervention coverage. It also introduces a resolution problem: large modules may combine distinct processes, whereas small modules may fragment a single mechanism into redundant units. The framework therefore treats module construction as an explicit modeling choice rather than as a neutral preprocessing step.

Formally, let C denote a set of candidate compounds, T a set of molecular targets, P a set of pathways, and M a set of disease modules. The intervention graph is defined as a heterogeneous graph whose vertices belong to these four types and whose edges encode evidence or functional relationships.

$$G = (V, E, W), V = C \cup T \cup P \cup M$$

Here V is the vertex set, E is the edge set, and W assigns weights to edges. Edges may connect compounds to targets, targets to pathways, pathways to modules, targets directly to modules, or compounds to modules when phenotypic evidence is available. The graph is heterogeneous because vertex types have different biological meanings; a path through compound-target-pathway-module is not equivalent to a path through compound-target-module. This distinction prevents the model from treating all associations as interchangeable and allows evidence type to influence ranking.

The framework assumes that candidate compounds have at least one annotated or predicted target, that disease modules can be defined from prior biological knowledge or expression data, and that edge evidence can be represented with a confidence weight. It does not assume that all edges are causal. Instead, edge weights express the strength of mechanistic support, and causal claims require separate validation. This distinction is important because network proximity can reflect shared membership in a biological process, physical interaction, or database bias. The proposed framework uses graph structure to organize hypotheses, not to infer causality from connectivity alone.

3. Methodological Framework

3.1. Heterogeneous evidence integration

The first component converts heterogeneous evidence into comparable edge weights. Compound-target evidence may include binding affinity, predicted interactions, transcriptional reversal, and pathway-level annotations. Disease-module evidence may include genetic association, differential expression, and pathway enrichment. These evidence types are not commensurable in their raw form, so the framework uses a bounded fusion function rather than a simple sum of database scores. A logistic form is convenient because it constrains weights to the unit interval and allows evidence contributions to saturate.

$$w(c,t) = \sigma\big(\alpha \cdot b(c,t) + \beta \cdot e(c,t) + \gamma \cdot p(c,t)\big)$$

In this expression, w(c,t) is the compound-target edge weight, b(c,t) is a binary or graded binding evidence term, e(c,t) is an expression-reversal or functional-association term, and p(c,t) is a pathway or prior-knowledge term. The symbol σ denotes a logistic transform, and α, β, and γ are nonnegative parameters that determine the relative contribution of each evidence

class. The parameters may be elicited from domain experts, bounded by evidence-quality categories, or calibrated in a specific application. The equation is a modeling choice, not an estimated result. Its purpose is to make evidence fusion explicit and auditable, so that a high weight cannot arise from an untraceable combination of weak annotations.

Expression signatures can inform module construction and compound-response similarity [11], and gene-set enrichment can summarize pathway-level effects [12]. Protein interaction resources provide scaffold connectivity [13], while drug databases provide target and indication annotations [14]. These resources differ in coverage and error structure. A compound may be well annotated in one database and absent from another; a disease module may be enriched for well-studied genes simply because they have been investigated more often. The framework therefore treats missing edges as unknown rather than as evidence of no relationship. This choice reduces the risk that poorly studied mechanisms are systematically deprioritized.

3.2. Module perturbation and compound-set effects

The second component translates compound-target engagement into module-level perturbation. For a candidate set S, the effect on module m depends on which targets in that module are engaged, how strongly they are engaged, and how those targets relate to the disease state. The following expression defines a signed module perturbation score.

$$\Delta(m,S) = \sum_{t \in T(m)} z(t) \cdot \varphi(t,S) \cdot r(m,t)$$

Here T(m) is the set of targets assigned to module m, z(t) is a signed disease-alignment weight for target t, φ(t,S) is the compound-set effect on that target, and r(m,t) is a module-membership or regulatory weight. A positive Δ(m,S) indicates a shift toward disease-promoting activity under the chosen sign convention, whereas a negative value indicates reversal. The sign convention must be fixed before ranking because targets can be disease-promoting in one context and protective in another. Module membership weights allow a target to contribute to more than one module, which is realistic for pleiotropic signaling proteins.

The compound-set effect φ(t,S) should reflect saturating target engagement rather than a linear sum of compound concentrations. A simple monotone form is:

$$\varphi(t,S) = 1 - \prod_{c \in S} \left(1 - \eta(c,t) \cdot x(c)\right)$$

In this expression, η(c,t) is the fractional target modulation produced by compound c at target t under the adopted scale, and x(c) is a normalized exposure or dose variable. The product term represents the fraction of target activity not yet modulated by the set, so φ(t,S) approaches one when the set saturates the target and remains near zero when engagement is weak. The equation assumes monotone, roughly independent target engagement within the modeled range. It does not represent antagonism, target-mediated clearance, or dynamic adaptation; those effects require additional terms or experimental calibration. Its value is that it prevents a combination from receiving credit for repeatedly modulating the same target through redundant compounds.

For ranking, the relevant quantity is not perturbation alone but beneficial reversal. For disease-promoting modules, R(m,S) can be defined as the positive part of −Δ(m,S), with the sign reversed for protective modules. This transformation ensures that compounds which

worsen a disease-promoting module are not rewarded merely because they change it. The framework therefore separates directional mechanism from absolute activity, a distinction that is often lost when expression signatures are compared by correlation alone.

3.3. Synergy-aware prioritization with redundancy and safety penalties

The third component defines a set-level objective. A multi-compound intervention should cover multiple disease modules, avoid redundant modulation of the same mechanism, and remain within a tolerable safety envelope. The following score combines these considerations.

$$\Psi(S) = \sum_{m \in M} \lambda(m) \cdot R(m,S) - \mu \cdot \sum_{m<n} \kappa(m,n) \cdot R(m,S) \cdot R(n,S) - \nu \cdot H(S)$$

Here λ(m) is the importance of module m in the disease context, R(m,S) is the beneficial reversal defined above, μ is a redundancy-penalty weight, κ(m,n) measures mechanistic overlap between modules m and n, and ν weights the safety burden. The second term penalizes pairs of modules that are both strongly reversed when their target sets or pathways overlap. This penalty is not a claim that redundancy is always undesirable; it encodes the design principle that a combination should add mechanism coverage rather than duplicate the same effect. The overlap parameter κ(m,n) can be derived from shared targets, pathway annotations, or community structure [16].

The safety burden H(S) aggregates predicted off-target engagement weighted by adverse-effect relevance.

$$H(S) = \sum_{c \in S} x(c) \cdot \sum_{t \notin T_d is} w(c,t) \cdot a(t)$$

In this expression, T_dis is the disease-relevant target set, a(t) is an adverse-effect or tissue-risk weight for target t, and the outer sum accumulates burden across compounds. The complement t∉ T_dis is intentional: engagement of disease-relevant targets is evaluated through module reversal, while engagement of unrelated targets contributes to safety burden. The framework does not assume that all off-target binding is harmful; a(t) expresses prior evidence or expert judgment about the likelihood and severity of adverse consequences. In applications, H(S) should be interpreted as a comparative penalty, not as a clinical risk estimate.

3.4. Constrained optimization and uncertainty propagation

Candidate prioritization is formulated as a constrained optimization problem. The goal is to select a compound set of bounded size that maximizes the mechanism-aware score after accounting for evidence uncertainty.

$$S* = argmax_\{S \subseteq C, |S| \leq K\} [\Psi(S) - \rho \cdot \Omega(S)]$$

Here S is the top-ranked set, K is the maximum number of compounds, Ω(S) is an uncertainty penalty, and ρ controls the trade-off between expected score and evidence stability. Additional constraints can be imposed, such as minimum target coverage, maximum predicted safety burden, or exclusion of compounds with contraindicated interactions. The equation does not define a trained predictor; it defines a ranking rule whose inputs must be supplied by curated evidence. The choice of K is a design decision because larger sets can increase mechanism coverage while also increasing safety burden and diminishing returns from target saturation.

A first-order uncertainty score can be obtained by propagating edge-weight variance through the score.

$$\Omega(S) = \sqrt{(\sum_\{c \in S\}\sum_\{t \in T\} (\partial\Psi/\partial w(c,t))\text{^}2 \cdot \sigma\text{^}2(c,t))}$$

In this expression, $\partial\Psi/\partial$w(c,t) is the local sensitivity of the score to a compound-target edge weight, and σ^2(c,t) is the variance or evidence-derived uncertainty of that weight. The square root yields a standard-deviation-like quantity under an independence approximation. The approximation is not exact because edges interact through module terms and redundancy penalties, but it provides a transparent way to penalize candidates whose high scores depend on weakly supported annotations. More elaborate propagation, such as Monte Carlo sampling or interval arithmetic, can be used when edge dependence is substantial. The important point is that uncertainty is part of the ranking rule rather than a post hoc annotation.

Decision logic can then distinguish three classes of candidates: robustly prioritized sets with high Ψ(S) and low Ω(S); exploratory sets with high Ψ(S) but high Ω(S); and deprioritized sets with low Ψ(S) or high H(S). This classification is more informative than a single ordinal list because it indicates whether a recommendation is driven by strong mechanism evidence or by speculative edges. False-discovery control can be applied when many candidate sets are screened [19], and causal claims should be reserved for mechanisms supported by perturbation data rather than graph proximity alone [20]. Active-module and co-expression methods can inform module construction and edge confidence [17], [18], but they do not remove the need for external validation.

4. Analytical Assessment and Design Trade-offs

4.1. Mechanism coverage versus redundancy

The framework's main advantage over target-centric ranking is that it rewards coverage of distinct disease modules while penalizing duplicated mechanisms. A single compound with high affinity for one target may score well in a target-based model but leave several disease modules unaddressed. A multi-compound set can distribute engagement across modules, but it can also over-modulate a shared pathway. The redundancy term in the objective makes this trade-off explicit. If κ(m,n) is high, two modules that appear distinct at the annotation level but share most targets will not be treated as independent sources of benefit. This is important in complex diseases where pathway crosstalk can make nominal module diversity misleading.

The choice of μ determines how strongly redundancy is penalized. A very low μ produces broad sets that may be mechanistically redundant; a very high μ favors narrow, nonoverlapping mechanisms and may discard useful combinations that exploit partial overlap for efficacy. The parameter should therefore be interpreted as a design preference constrained by biological context. In inflammatory disease, partial overlap in cytokine signaling may be desirable; in oncology, redundant survival pathways may be a liability. The framework does not resolve this preference automatically, but it makes the consequence of the choice visible in ranking.

4.2. Graph topology and evidence quality

Graph-based prioritization is sensitive to topology and annotation bias. Hubs and well-studied proteins receive more edges, and disease modules enriched for characterized genes may dominate the score. The logistic edge-weight function reduces, but does not eliminate, this problem because more evidence can still produce higher weights. A useful safeguard is to stratify rankings by evidence class or to report scores under alternative edge-weight settings.

If a candidate set remains highly ranked when weak edges are down-weighted, its prioritization is less dependent on annotation density.

The framework also inherits errors from module construction. Large modules can mix disease-relevant and bystander processes; small modules can fragment a single mechanism into several units that are then incorrectly treated as independent. Sensitivity analysis over module resolution is therefore essential. The ranking should be recomputed across plausible module definitions, and candidates that are stable across resolutions should be distinguished from those that depend on one partition. This is consistent with the broader lesson from network biology that topology is informative but not self-validating [1], [5].

4.3. Synergy, dose, and context

The framework represents synergy as complementarity in target and module coverage, not as a substitute for experimental synergy quantification. This distinction matters because a set can be mechanism-complementary yet pharmacologically antagonistic at a given dose ratio. The saturating target-engagement term captures diminishing returns from repeated target modulation, but it does not capture Loewe additivity, Bliss independence, or context-specific signaling adaptation [8], [9], [10]. Experimental dose-response designs remain necessary to determine whether the predicted complementarity translates into a synergistic or additive phenotype.

Dose context enters through x(c), which normalizes exposure across compounds. If x(c) is set to a clinically achievable or experimentally tested concentration, the ranking reflects plausible exposure rather than theoretical target occupancy. If x(c) is unknown, the framework can rank under multiple exposure assumptions and report the stability of the ranking. This is preferable to assigning arbitrary equipotent doses, which can make combinations appear favorable only because one compound is implicitly administered at a saturating level. The framework therefore treats dose as a scenario variable rather than a fixed property of a compound.

4.4. Computational and inferential limitations

The optimization problem is combinatorial in the number of candidate compounds. Exhaustive search is feasible only for small candidate pools; larger pools require branch-and-bound, integer programming, heuristic search, or greedy approximation with post hoc refinement. The framework does not prescribe a single solver because the appropriate choice depends on constraints, graph size, and whether exact optimality is required. What matters is that any approximation reports its optimality gap or search stability, so that a ranked set is not mistaken for a certified optimum.

The inferential limitation is more fundamental. A graph edge is not a causal effect, and a high-ranked combination is not evidence of efficacy. The framework organizes mechanistic hypotheses and makes assumptions explicit, but it cannot replace perturbation experiments, dose-response assays, or clinical evaluation. Its output should be read as a prioritized set of testable interventions with stated evidence dependencies. This limitation is not a defect of the graph approach; it is a boundary condition that separates computational prioritization from biological validation.

5. Discussion and Translation

The proposed framework is most useful when a disease mechanism is distributed across several modules and when candidate compounds have heterogeneous evidence. In that setting,

a single-target ranking discards information about coverage and redundancy, while a purely expression-based ranking may reward phenotypic similarity without distinguishing mechanism. The graph formulation provides a common language for compounds, targets, pathways, and modules, and the objective function makes trade-offs among coverage, redundancy, and safety explicit. The framework can be applied to combination repurposing, where approved compounds with known safety profiles are evaluated for new multi-module indications, and to preclinical lead combination design, where the goal is to select a small set for experimental testing.

Translation requires versioned evidence and transparent parameter reporting. The same compound-target edge may come from a curated interaction database, a predicted binding model, or a transcriptional signature; these sources should not be silently merged. The edge-weight equation provides a place to encode evidence quality, but the chosen α, β, and γ values must be reported and subjected to sensitivity analysis. Similarly, module definitions, safety weights, and the maximum set size K should be documented because they shape the ranking as much as the graph itself. This level of reporting is necessary for reproducibility and for comparing prioritized sets across studies.

The framework also clarifies the relationship between mechanistic reasoning and observed evidence. Mechanistic edges can generate hypotheses about why a combination might work, while observed synergy or perturbation data provide evidence that it does work in a particular context. The two should not be conflated. A candidate set with strong mechanistic support but no experimental synergy is a hypothesis; a set with experimental synergy but unexplained mechanism is an empirical observation awaiting mechanistic interpretation. The graph-based score is designed to integrate both kinds of information without treating either as sufficient.

6. Conclusion

This manuscript presented a graph-based systems-biology framework for prioritizing multi-compound interventions in complex disease mechanisms. The framework represents compounds, targets, pathways, and disease modules in a heterogeneous graph, fuses heterogeneous evidence into bounded edge weights, propagates compound-set effects through saturating target engagement, and ranks candidate sets using mechanism coverage, redundancy penalties, safety burden, and uncertainty. Synergy is treated as nonredundant mechanistic complementarity, which connects network-level reasoning to the dose-response questions that experimental combination studies must answer. The analytical assessment identified the framework's main strengths and boundary conditions: it can make coverage and redundancy explicit, but it depends on module resolution, evidence quality, dose assumptions, and solver behavior. Its output is a prioritized set of testable hypotheses, not a clinical recommendation. Used with transparent evidence provenance and sensitivity analysis, the framework offers a disciplined way to move from complex disease mechanisms to multi-compound intervention hypotheses.

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). Systems biology: A brief overview. Science, 295(5560), 1662–1664. https://doi.org/10.1126/science.1069492

3. 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

4. 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

5. 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: A comprehensive review. Pharmacology & Therapeutics, 138(3), 333–408. https://doi.org/10.1016/j.pharmthera.2013.01.016

6. Guney, E., Menche, J., Vidal, M., & Barabási, A.-L. (2016). Network-based in silico drug efficacy screening. Nature Communications, 7, 10331. https://doi.org/10.1038/ncomms10331

7. Cheng, F., Kovács, I. A., & Barabási, A.-L. (2019). Network-based prediction of drug combinations. Nature Communications, 10, 1197. https://doi.org/10.1038/s41467-019-09186-x

8. Chou, T.-C., & Talalay, P. (1984). Quantitative analysis of dose-effect relationships: The combined effects of multiple drugs or enzyme inhibitors. Advances in Enzyme Regulation, 22, 27–55. https://doi.org/10.1016/0065-2571(84)90007-4

9. Loewe, S. (1953). The problem of synergism and antagonism of combined drugs. Arzneimittel-Forschung, 3(6), 285–290.

10. Bliss, C. I. (1939). The toxicity of poisons applied jointly. Annals of Applied Biology, 26(3), 585–615. https://doi.org/10.1111/j.1744-7348.1939.tb06990.x

11. Lamb, J., Crawford, E. D., Peck, D., Modell, J. W., Blat, I. C., Wrobel, M. J., ... Golub, T. R. (2006). The Connectivity Map: Using gene-expression signatures to connect small molecules, genes, and disease. Science, 313(5795), 1929–1935. https://doi.org/10.1126/science.1132939

12. Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., ... Mesirov, J. P. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences, 102(43), 15545–15550. https://doi.org/10.1073/pnas.0506580102

13. Szklarczyk, D., Gable, A. L., Nastou, K. C., Lyon, D., Kirsch, R., Pyysalo, S., ... von Mering, C. (2021). The STRING database in 2021: Customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets. Nucleic Acids Research, 49(D1), D605–D612. https://doi.org/10.1093/nar/gkaa1074

14. Wishart, D. S., Feunang, Y. D., Guo, A. C., Lo, E. J., Marcu, A., Grant, J. R., ... Wilson, M. (2018). DrugBank 5.0: A major update to the DrugBank database for 2018. Nucleic Acids Research, 46(D1), D1074–D1082. https://doi.org/10.1093/nar/gkx1037

15. 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

16. Newman, M. E. J. (2006). Modularity and community structure in networks. Proceedings of the National Academy of Sciences, 103(23), 8577–8582. https://doi.org/10.1073/pnas.0601602103

17. Ideker, T., Ozier, O., Schwikowski, B., & Siegel, A. F. (2002). Discovering regulatory and signalling circuits in molecular interaction networks. Bioinformatics, 18(Suppl. 1), S233–S240. https://doi.org/10.1093/bioinformatics/18.suppl_1.S233

18. Zhang, B., & Horvath, S. (2005). A general framework for weighted gene co-expression network analysis. Statistical Applications in Genetics and Molecular Biology, 4, Article 17. https://doi.org/10.2202/1544-6115.1128

19. Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. Journal of the Royal Statistical Society: Series B, 57(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

20. Pearl, J. (2009). Causality: Models, reasoning, and inference (2nd ed.). Cambridge University Press.