Network Medicine Approaches for Revealing Shared Molecular Mechanisms Between Chronic Diseases and Natural Compound Therapies

Conor Parkes
Department of Computer Science, University of North Texas, Denton, TX, USA.
helloconor@unt.edu

Ekaraj Hari
Department of Electrical Engineering and Computer Science, University of Missouri, Columbia, MO, USA.
ekaraj1972@missouri.edu

Naomh O'Callaghan
Department of Computer Science and Engineering, University of Nevada, Reno, Reno, NV, USA.
nocallaghan@unr.edu

Abstract

Chronic diseases such as type 2 diabetes, cardiovascular disease, and neurodegeneration frequently co-occur, yet their shared molecular determinants remain poorly resolved by single-disease reductionism. Natural compounds offer a multimodal therapeutic logic that is difficult to reconcile with target-centric pharmacology. This manuscript develops a network medicine framework for identifying the molecular mechanisms that couple chronic disease modules to natural-compound target sets. The contribution is a formal analytical pipeline that integrates disease-gene module construction, compound-target network projection, module-level proximity and convergence analysis, community-level decomposition of polyphenol and alkaloid target sets, and an interaction-context layer that distinguishes intrinsic target topology from pathway-mediated crosstalk. We define a convergent module score for quantifying shared etiological architecture across disease pairs and a compound-mediated convergence index for ranking natural products by their capacity to engage disease module intersections rather than isolated hubs. The analysis specifies the conditions under which topological proximity reflects mechanistic overlap, including module boundary sensitivity, hub bias, and the distinction between physical and functional interactions. We also show why natural compounds require community-aware rather than single-target evaluation and how synergistic combinations can be identified from overlapping but non-identical module neighborhoods. The framework provides a reproducible conceptual basis for experimental prioritization and clarifies when network-derived claims about shared mechanisms are defensible versus overinterpreted. Its significance lies in connecting comorbidity biology with polypharmacological natural product research through explicitly stated, testable, and falsifiable network criteria.

Keywords

network medicine, chronic disease comorbidity, natural compounds, disease modules, polypharmacology, synergistic interactions

1. Introduction

Chronic diseases are rarely isolated. Clinical epidemiology consistently shows that type 2 diabetes, ischemic heart disease, chronic kidney disease, and neurodegenerative conditions cluster within individuals and families, and that this clustering cannot be fully explained by shared behavioral risk factors. The molecular basis of such comorbidity is distributed: it involves overlapping genetic susceptibility, shared inflammatory and metabolic signaling, and convergent damage responses across tissues. Single-gene and single-pathway models capture only fragments of this architecture. Network medicine emerged as a response to this explanatory gap, proposing that disease phenotypes arise from perturbations within localized neighborhoods of the human interactome, termed disease modules, and that comorbidity reflects measurable overlap or topological proximity between such modules [1-3]. This view does not deny the importance of individual molecules; it reframes them as components whose pathological significance depends on their position within a broader interaction context.

Natural compounds present a complementary puzzle. A large fraction of approved drugs derive from or are inspired by natural products, and traditional medical systems have long used multi-herb formulations whose effects are difficult to attribute to a single constituent [4,5]. Experimental pharmacology has identified numerous bioactive natural compounds with anti-inflammatory, antioxidant, metabolic, and neuroprotective activities, but the dominant evaluation paradigm remains target-by-target screening. This is epistemically mismatched to compounds that typically modulate multiple proteins with modest individual affinities. Network pharmacology and related approaches have therefore proposed that natural compounds should be evaluated by their collective target profile and its relationship to disease modules rather than by potency at a single receptor [6,7]. Yet much of this literature remains descriptive, reporting target lists and pathway enrichments without a principled account of when a target set actually engages a disease module, when apparent overlap is attributable to hub bias, and when combination synergy is mechanistically plausible rather than merely additive.

The gap addressed here is methodological and conceptual. Existing work has separately developed disease-module detection, network proximity between disease and drug targets, and community-based analysis of herbal combinations [8-10]. What remains underdeveloped is a connected analytical framework that (i) defines what counts as a shared mechanism between two chronic diseases, (ii) specifies how natural-compound target sets should be decomposed so that multimodal action is represented rather than collapsed, and (iii) articulates the boundary conditions under which network-derived convergence claims are and are not licensed. The required reference at position [9] exemplifies the community-based logic relevant to the third point, but the present manuscript is not an exposition of that model; it is a broader synthesis that situates community decomposition within a disease-comorbidity framework.

The central contribution is a formal network medicine pipeline for revealing shared molecular mechanisms between chronic diseases and natural-compound therapies. The framework makes four specific advances. First, it defines a convergent module score that distinguishes genuine etiological overlap from incidental topological proximity. Second, it introduces a compound-mediated convergence index that rewards engagement of disease module intersections and penalizes hub-dominated targeting. Third, it specifies a community-level decomposition of natural-compound target sets so that synergy can be assessed as interaction between partially overlapping module neighborhoods. Fourth, it states falsifiable

conditions and sensitivity analyses that determine when the resulting mechanistic interpretations are warranted. The manuscript is analytical rather than empirical: it develops the formal relationships, defends the modeling choices against alternatives, and derives testable consequences that can guide subsequent experimental work.

2. Conceptual Foundations and Formal Representation

2.1 Disease Modules and Comorbidity

A disease module is defined as a connected subgraph of a molecular interaction network whose perturbation is sufficient to produce or sustain a disease phenotype [2,3]. Operationally, module identification begins with a seed set of disease-associated genes and expands to include interaction partners that are statistically enriched for association with the disease. Let G = (V, E) denote the human interactome, with V the set of proteins and E the set of physical or functional interactions. For disease d, let S_d $\subset$ V be the seed set derived from curated association data, and let M_d $\subset$ V be the inferred module. Module inference can be expressed as a seed-and-expand optimization in which a module score balances association evidence against topological compactness:

Here a_d(v) is a normalized association score of node v with disease d, cut(M, V\M) counts the interaction edges crossing the module boundary, and $\lambda \geq 0$ controls the penalty for topological fragmentation. The first term rewards inclusion of disease-relevant nodes, while the second prevents the module from dissolving into disconnected high-scoring singletons. The parameter λ is not a fitted biological constant; it is a modeling choice that determines module granularity, and its effect must be reported as a sensitivity dimension. Smaller λ yields larger, more inclusive modules; larger λ yields compact cores. Because downstream proximity and overlap estimates depend on module boundaries, any claim of shared mechanism must be evaluated across a range of λ values rather than at a single arbitrary setting.

Comorbidity between diseases d and e is then assessed through module overlap and network proximity. Direct overlap is the Jaccard-like ratio of shared module nodes, but overlap alone is insufficient because two modules may be adjacent without sharing nodes, which is mechanistically meaningful. Network proximity between modules is defined as the average shortest-path distance from nodes in M_d to nodes in M_e, normalized against a degree-preserving null distribution [3]. A separation score z_{de} quantifies how many standard deviations the observed proximity deviates from random expectation:

$$z_\{de\} = (\langle d(M_d, M_e) \rangle - \mu_\{de\}) / \sigma_\{de\}$$

where $\langle$ d(M_d, M_e)$\rangle$ is the observed mean module distance, and μ_{de} and σ_{de} are the mean and standard deviation of distances between degree-matched random node sets. Negative z_{de} indicates that the two disease modules are closer than expected by chance. The degree-preserving null is essential because high-degree hubs reduce distances generically; without it, nearly all large modules appear proximate. This normalization is a safeguard against the most common inferential error in disease-proximity studies.

2.2 Natural-Compound Target Sets

A natural compound c is represented by a target set T_c $\subset$ V, assembled from experimentally reported binding, inhibition, or modulation events. Because such data are heterogeneous in confidence, each target receives a reliability weight w_c(v) $\in$ [0,1]. The

weighted target set is not treated as a flat list. Instead, it is decomposed into topological communities, reflecting the observation that multimodal compounds tend to engage several distinct functional neighborhoods rather than a random scatter of proteins. Community decomposition partitions T_c into subsets C_{c,1}, ..., C_{c,k} such that within-community connectivity exceeds between-community connectivity. This step is where the community-based logic of the required reference [9] becomes relevant: rather than scoring a compound by its single strongest target, the framework evaluates how its constituent communities map onto disease modules.

The mapping between a compound community C_{c,j} and a disease module M_d is quantified by a weighted proximity score. Let π(v, M_d) be the shortest-path distance from target v to the nearest node in M_d. The compound-module engagement is:

$$\mathrm{E}(c,d) = \sum_{j=1}^{k} \beta_j \cdot \exp\left(-\gamma \cdot \min_{v \in C_{c,j}} \pi(v, M_d)\right)$$

The term min over v captures the fact that a compound community engages a disease module if any of its members lies near the module; requiring all members to be proximal would wrongly penalize compounds with one high-affinity anchor and several weak modulators. The exponential decay with $\gamma > 0$ encodes the diminishing mechanistic relevance of distant targets, and β_j weights communities by their internal coherence and target reliability. The choice of γ controls how sharply engagement falls with distance; it should be selected relative to the characteristic path length of the interactome and reported explicitly. This formulation is deliberately transparent: it does not claim that proximity equals causation, only that proximity is a necessary precondition for direct mechanistic coupling, which can then be tested experimentally.

3. Methodological Framework for Shared Mechanisms

3.1 Convergent Module Score

The first analytical objective is to determine whether two chronic diseases share a molecular mechanism rather than merely coexisting clinically. Shared mechanism is defined as the existence of a set of molecules that (i) belongs to or lies immediately adjacent to both disease modules and (ii) forms a connected subgraph whose perturbation is plausibly sufficient to affect both phenotypes. This motivates a convergent module score. Let I_{de} = M_d ∩ M_e be the direct intersection and let A_{de} be the set of nodes adjacent to both modules. The convergent core K_{de} is the largest connected component of the subgraph induced by I_{de} ∪ A_{de}. The score combines core size, connectivity, and association evidence:

$$CMS_\{de\} = (|K_\{de\}| / |M_d \cup M_e|) \cdot (\sum_\{v \in K_\{de\}\} min(a_d(v), a_e(v)) / |K_\{de\}|)$$

The first factor measures the fraction of the combined module space occupied by the convergent core, penalizing small incidental overlaps. The second factor measures the average joint association of core nodes with both diseases, using the minimum so that a node strongly associated with one disease but irrelevant to the other does not inflate the score. CMS_{de} is bounded in [0,1] and is interpretable as the joint fraction of etiological architecture shared by the two diseases. A high score indicates that the diseases converge on a connected, jointly associated molecular neighborhood. A low score despite clinical

comorbidity suggests that the comorbidity arises from parallel rather than shared mechanisms, which is itself an informative finding.

3.2 Compound-Mediated Convergence Index

The second objective is to rank natural compounds by their capacity to engage the convergent core rather than isolated disease modules. A compound that targets only M_d may treat disease d but will not address the shared mechanism. A compound that targets K_{de} is a candidate for simultaneous modulation of both diseases. The compound-mediated convergence index is defined as:

$$CMI(c) = \mathrm{E}\frac{(c, K_{de})}{\big(1 + H(c)\big)}$$

where E(c, K_{de}) is the engagement score from Equation 3 evaluated against the convergent core, and H(c) is a hub-bias penalty. The penalty is necessary because compounds whose targets are high-degree hubs will appear to engage almost any module. H(c) is defined as the mean degree percentile of T_c relative to the full interactome, normalized to [0,1]. Dividing by 1 + H(c) ensures that the index rewards specific engagement of the convergent core over generic hub binding. This is a deliberate departure from raw proximity scores, which systematically favor promiscuous hub-targeting compounds. The denominator also prevents division by zero and keeps the index finite for compounds with no hub targets.

The CMI is a ranking device, not a therapeutic efficacy claim. Its value lies in prioritizing compounds for experimental testing of dual-disease effects. A compound with high CMI but low individual potency may still be a rational starting point for combination or formulation studies, whereas a compound with high potency at a single hub and low CMI is likely to produce broad but nonspecific effects.

3.3 Community-Level Synergy Analysis

Natural-compound therapies frequently involve combinations, and the framework must therefore address synergy. Two compounds c and c' are predicted to act synergistically on the convergent core if their target communities engage complementary rather than redundant portions of K_{de}. Let the coverage of compound c over K_{de} be the set of core nodes within distance τ of any target in T_c, denoted Cov_c(K_{de}). Redundancy is the overlap of coverage sets; complementarity is the union minus the intersection. A synergy score is:

$$Syn(c, c') = |Cov_c \cup Cov_\{c'\}| \,/\, |K_\{de\}| - \delta \cdot |Cov_c \cap Cov_\{c'\}|$$

The first term rewards broad joint coverage of the convergent core. The second term penalizes redundant coverage, with $\delta \geq 0$ controlling how strongly redundancy is discounted. When $\delta = 0$, the score reduces to union coverage, which would favor combinations of two broad-spectrum compounds; positive δ favors combinations that reach different parts of the core. This reflects a mechanistic hypothesis: synergy is more likely when two compounds perturb distinct nodes or subcommunities within the same etiological neighborhood than when both saturate the same nodes. The score is explicitly a hypothesis-generating quantity. It does not establish synergy, but it identifies combinations whose target topology makes synergy plausible and therefore worth experimental testing.

The distance threshold τ is a critical parameter. It should be chosen to reflect the scale at which local perturbations propagate in the interactome, and results should be reported across a range of τ. A combination that is synergistic only at implausibly large τ is not credible, because at large distances nearly all targets are "near" the core.

4. Analytical Assessment and Boundary Conditions

4.1 Why Network Proximity Is Not Mechanism

The framework's credibility depends on stating clearly what network proximity can and cannot establish. Proximity in a physical interactome indicates that two proteins can influence each other through a bounded number of interactions; it does not establish that they do so in the relevant tissue, at the relevant expression level, or with the relevant directionality. A negative z_{de} for two disease modules is consistent with shared mechanism, but it is also consistent with both modules being embedded in a densely connected inflammatory or metabolic neighborhood that is generic to many diseases. The convergent module score addresses this partly by requiring joint association evidence, but it cannot eliminate the possibility that the shared core is a general stress-response module rather than a disease-specific one. This is why the framework treats CMS_{de} as a prioritization statistic and not as proof of shared etiology. Experimental validation, ideally through perturbation of core nodes in disease-relevant models, remains necessary.

4.2 Hub Bias and Degree Normalization

The most pervasive failure mode in network pharmacology is hub bias. High-degree proteins participate in many interactions and therefore lie near almost any module. Compounds targeting hubs appear to have broad therapeutic potential, and diseases whose modules include hubs appear to be proximate to many other diseases. Degree-preserving null models are the standard corrective, but they are not sufficient when the query set itself is enriched for hubs. The CMI's explicit hub penalty is designed for this situation, but it introduces its own trade-off: a compound that genuinely acts through a hub may be unfairly downranked. The appropriate response is to report both the raw engagement and the penalized index, so that readers can see whether a compound's ranking depends on hub targeting. Neither number alone is decisive.

4.3 Module Boundary Sensitivity

Module inference is sensitive to seed selection, association thresholds, and the expansion parameter λ. A convergent core identified at one granularity may disappear at another. This is not a defect to be hidden; it is a property to be quantified. The framework requires that CMS_{de} and CMI(c) be computed across a grid of λ and association thresholds, and that conclusions be restricted to findings that are stable across that grid. Stability can be summarized by the rank correlation of compounds across parameter settings or by the persistence of core nodes. A compound that ranks highly only at a single arbitrary parameter setting should not be prioritized. This requirement imposes a computational cost but prevents the more serious cost of chasing parameter-specific artifacts.

4.4 Physical Versus Functional Interactions

Interactome edges may represent physical binding, catalytic relationships, regulatory interactions, or curated pathway co-membership. Mixing these edge types without distinction can create spurious proximity. Two proteins that never bind may be "close" because they share a pathway annotation. The framework therefore recommends stratifying analyses by edge type where data permit, and at minimum reporting whether conclusions depend on functional rather than physical edges. Natural-compound target sets are especially vulnerable here, because many reported targets derive from pathway-level inference rather than direct

binding assays. Weighting targets by evidence type, as in w_c(v), partially addresses this, but the limitation must be stated.

5. Technical Discussion and Implications

5.1 Relation to Existing Approaches

The framework differs from single-target drug repurposing in its unit of analysis. Repurposing typically asks whether a known drug's primary target lies near a disease module; the present approach asks whether a compound's community-decomposed target set engages the intersection of two disease modules. It differs from descriptive network pharmacology in its insistence on null models, hub penalties, and parameter sensitivity. It differs from purely community-based herbal analysis in its explicit linkage to disease comorbidity rather than to a single indication. The required reference [9] is relevant precisely at the community-decomposition step, where the question is how a multi-component herbal or compound set should be partitioned before scoring. The present framework does not claim that community decomposition alone resolves synergy; it embeds decomposition within a disease-module framework and adds the convergent-core and hub-penalty logic that comorbidity analysis requires.

5.2 Experimental Implications

The framework generates specific, testable predictions. A high CMS_{de} predicts that perturbing core nodes should affect phenotypes of both diseases in model systems. A high CMI predicts that the compound will modulate both disease modules, which can be tested by profiling its effects on module activity rather than on single markers. A high Syn(c, c') predicts that the combination will outperform either compound alone at doses where individual effects are submaximal. These predictions are falsifiable, and their failure is informative: if a high-CMI compound affects only one disease module, the proximity assumption underlying its ranking is wrong for that compound, and the target set or weighting requires revision.

The framework also has implications for how natural-compound data are curated. Because community decomposition and hub penalties depend on target reliability weights, investment in standardized, evidence-typed target annotation would substantially improve the framework's discriminative power. Pooling binding assays, pathway inferences, and expression correlations into a single unweighted list undermines every downstream score. This is a data-governance implication that follows directly from the analytical structure, not an external policy recommendation.

5.3 Limitations

The framework inherits the limitations of the underlying interactome. Tissue-specific expression is not represented in a static network, so a mechanism that is plausible in liver may be irrelevant in neurons. Directionality of regulation is largely absent, meaning that inhibition and activation are treated equivalently in proximity calculations. Temporal dynamics of disease progression are not modeled; a module that is central in early disease may be peripheral in late disease. These limitations do not invalidate the framework, but they bound its claims. The appropriate conclusion from a high convergence score is that a shared mechanism is topologically and associatively plausible, not that it has been demonstrated. The framework is a prioritization tool that narrows the space of hypotheses for experimental work.

A further limitation concerns the interpretation of the synergy score. Equation 5 rewards complementary coverage, but complementarity at the network level does not guarantee complementarity at the level of physiological effect. Two compounds may cover different core nodes yet produce antagonistic downstream effects. The score is therefore a topological prior, and its predictions require dose-response and combination-index testing to be meaningful.

6. Conclusion

This manuscript has developed a network medicine framework for revealing shared molecular mechanisms between chronic diseases and natural-compound therapies. The framework treats disease comorbidity as a problem of module convergence rather than clinical coincidence, defines a convergent module score that requires both topological proximity and joint disease association, and introduces a compound-mediated convergence index that rewards specific engagement of the convergent core while penalizing generic hub targeting. It further specifies a community-level synergy score that identifies combinations whose target topology makes complementary action plausible. Throughout, the analysis has emphasized boundary conditions: proximity is not mechanism, hub bias must be explicitly penalized, module boundaries must be tested for sensitivity, and physical interactions must be distinguished from functional annotations. These conditions are not qualifications appended to the framework; they are constitutive of its inferential discipline. The practical significance is a reproducible prioritization logic that connects comorbidity biology to natural-product pharmacology without overstating what network data can establish. The scientific significance is a clearer statement of when claims about shared molecular mechanisms are warranted and what evidence would overturn them. Future work can instantiate the framework on specific disease pairs and compound sets, but the analytical structure presented here is independent of any particular dataset and is intended to make such instantiations comparable and falsifiable.

References

1. Barabási, A.-L., Gulbahce, N., & Loscalzo, J. (2011). Network medicine: A network-based approach to human disease. Nature Reviews Genetics, 12(1), 56-68.

2. Loscalzo, J., & Barabási, A.-L. (2011). Systems biology and the future of medicine. Wiley Interdisciplinary Reviews: Systems Biology and Medicine, 3(3), 253-261.

3. Menche, J., Sharma, A., Kitsak, M., Ghiassian, S. D., Vidal, M., Loscalzo, J., & Barabási, A.-L. (2015). Uncovering disease-disease relationships through the incomplete interactome. Science, 347(6224), 1257601.

4. Newman, D. J., & Cragg, G. M. (2020). Natural products as sources of new drugs over the nearly four decades from 01/1981 to 09/2019. Journal of Natural Products, 83(3), 770-803.

5. Atanasov, A. G., Zotchev, S. B., Dirsch, V. M., & Supuran, C. T. (2021). Natural products in drug discovery: Advances and opportunities. Nature Reviews Drug Discovery, 20(3), 200-216.

6. Hopkins, A. L. (2008). Network pharmacology: The next paradigm in drug discovery. Nature Chemical Biology, 4(11), 682-690.

7. Li, S., & Zhang, B. (2013). Traditional Chinese medicine network pharmacology: Theory, methodology and application. Chinese Journal of Natural Medicines, 11(2), 110-120.

8. Goh, K.-I., Cusick, M. E., Valle, D., Childs, B., Vidal, M., & Barabási, A.-L. (2007). The human disease network. Proceedings of the National Academy of Sciences, 104(21), 8685-8690.

9. 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

10. Guney, E., Menche, J., Vidal, M., & Barabási, A.-L. (2016). Network-based in silico drug efficacy screening. Nature Communications, 7, 10331.

11. Cheng, F., Kovács, I. A., & Barabási, A.-L. (2019). Network-based prediction of drug combinations. Nature Communications, 10, 1197.

12. Yildirim, M. A., Goh, K.-I., Cusick, M. E., Barabási, A.-L., & Vidal, M. (2007). Drug-target network. Nature Biotechnology, 25(10), 1119-1126.

13. Vidal, M., Cusick, M. E., & Barabási, A.-L. (2011). Interactome networks and human disease. Cell, 144(6), 986-998.

14. Zitnik, M., Agrawal, M., & Leskovec, J. (2018). Modeling polypharmacy side effects with graph convolutional networks. Bioinformatics, 34(13), i457-i466.

15. Hsin, K.-Y., Ghosh, S., & Kitano, H. (2013). Combining machine learning systems and multiple docking simulation packages to improve docking prediction reliability for network pharmacology. PLoS ONE, 8(12), e83922.

16. Zhang, B., & Horvath, S. (2005). A general framework for weighted gene co-expression network analysis. Statistical Applications in Genetics and Molecular Biology, 4, Article17.

17. 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.

18. Ideker, T., & Sharan, R. (2008). Protein networks in disease. Genome Research, 18(4), 644-652.

19. Furlong, L. I. (2013). Human diseases through the lens of network biology. Trends in Genetics, 29(3), 150-159.

20. Pujol, A., Mosca, R., Farrés, J., & Aloy, P. (2010). Unveiling the role of network and systems biology in drug discovery. Trends in Pharmacological Sciences, 31(3), 115-123.

21. Zhou, X., Menche, J., Barabási, A.-L., & Sharma, A. (2014). Human symptoms-disease network. Nature Communications, 5, 4212.

22. Lee, D.-S., Park, J., Kay, K. A., Christakis, N. A., Oltvai, Z. N., & Barabási, A.-L. (2008). The implications of human metabolic network topology for disease comorbidity. Proceedings of the National Academy of Sciences, 105(29), 9880-9885.

23. Barabási, A.-L., & Oltvai, Z. N. (2004). Network biology: Understanding the cell's functional organization. Nature Reviews Genetics, 5(2), 101-113.

24. Kitano, H. (2002). Systems biology: A brief overview. Science, 295(5560), 1662-1664.

25. Schadt, E. E. (2009). Molecular networks as sensors and drivers of common human diseases. Nature, 461(7261), 218-223.