Asking the Right QUESTIONS: Network Spillover Reveals Temporal Signal in Gene Co-expression Networks

Carly A. Bobak

Research Computing and Data in Information, Technology & Consulting, & Biomedical Data Science, Dartmouth College
Hanover, NH 03784, USA
†*E-mail: carly.a.bobak@dartmouth.edu*

Jaini Shah

Geisel School of Medicine, Dartmouth College
Lebanon, NH 03755, USA
E-mail: jaini.h.shah@dartmouth.edu

A. James O'Malley

The Dartmouth Institute for Health Policy and Clinical Practice & Department of Biomedical Data Science, Geisel School of Medicine, Dartmouth College
Lebanon, NH 03755, USA
E-mail: james.omalley@dartmouth.edu

Gene co-expression networks group genes into modules that approximate biological programs, but standard analyses summarize each module by its own activity and ignore what the module's network neighbors contribute. We adapt network spillover, a construct developed to quantify indirect (exogenous peer or spillover) effects that falls under the causal inference concept of interference in social networks, to gene co-expression networks. Treating modules as units and the co-expression graph as the conduit of influence, we decompose a module's relationship to an outcome into a direct term, its own activity, and a spillover term, the exposure it receives from other modules across between-module edges, and we test whether cross-module spillover improves the predictive accuracy of an outcome beyond that provided from its own direct activity. Applied to a longitudinal cynomolgus macaque model of *Mycobacterium tuberculosis* infection, with lung FDG avidity as a measure of disease severity, cross-module spillover of a given module into myeloid modules during the peak-response window predicts severity beyond the modules' own activity, is stable under resampling and cross-validation, and yields a platelet module whose own activity is uninformative while its spillover is not. The unsupervised network additionally recovers a coherent map of blood cell lineages and immune programs. The associations are correlational. We provide the framework as an open-access R package, QUESTIONS, and discuss extensions to time-varying and single-sample networks.

Keywords: Gene co-expression networks; network spillover; interference; disease severity; tuberculosis.

© 2026 The Authors. Open Access chapter published by World Scientific Publishing Company and distributed under the terms of the Creative Commons Attribution Non-Commercial (CC BY-NC) 4.0 License.

1. Introduction

Biological function arises not from genes in isolation but from their coordinated activity.[1] Gene co-expression networks have become a standard tool for representing this coordination, grouping genes into modules that approximate pathways and cell-type programs and relating those modules to phenotypes.[2,3] When expression is profiled longitudinally, across development, circadian cycles, disease progression, or response to a perturbation, the coordination itself changes over time: modules gain and lose importance, and the relationships among genes rewire as the biological context shifts.[4,5] Capturing how network structure evolves, rather than collapsing a time course into a single static picture, is widely recognized as necessary for understanding dynamic regulatory processes, since a network built from all time points pooled together obscures precisely the transient structure that distinguishes one phase of a response from the next.[4]

Most co-expression analyses stop at the level of individual modules, relating each module's own activity to phenotype but rarely asking whether the information carried between modules is itself informative. Most temporal transcriptomic analysis still operates one gene at a time, tracking expression trajectories while ignoring the coordinated structure that co-expression is meant to capture.[5] Where network structure is modeled over time, the dominant strategies pursue a different goal than ours. One line of work estimates a separate network at each time point, for example through time-varying graphical models, and asks how edges appear or disappear;[6,7] a second, network propagation, diffuses a signal such as a set of seed genes across a fixed molecular network to amplify and prioritize associated genes, and has become a mainstay of disease-gene discovery;[8,9] a third, differential and single-sample co-expression analysis, quantifies how much the wiring around a gene or module changes between conditions or in an individual patient, and can relate that rewiring to clinical outcomes.[10,11] These approaches establish that network structure carries phenotypic information beyond mean expression. What none of them provides is a way to ask, for a given module, whether the activity of its network neighbors carries information about an outcome over and above the information in the module's own activity, which is the question we take up here.

The above question has a direct precedent outside biology. In the analysis of interference in social networks, spillover (or peer-exposure) methods quantify how a unit's outcome is shaped not only by its own treatment but by the treatments of the units it is connected to.[12] A concrete illustration comes from a stepped-wedge trial of a video-game intervention designed to increase physicians' Advance Care Planning (ACP) conversations with hospitalized patients.[13,14] Physicians were connected through a shared-patient network constructed from insurance billing records. Because physicians who had not yet received the intervention still collaborated with those who had, the standard trial analysis risked conflating the intervention's direct effect with the effect of proximity to already-treated peers. Modeling each physician's network exposure separately from their own treatment status showed that peer exposure predicted ACP billing in its own right, and that omitting it biased, and in some specifications reversed, the estimated direct effect. Other areas of social science or medicine in which peer-effects have been modeled include Medicare fraud[15] or extreme spending in home health care,[15] polypharmacy in older people,[16] the adoption of medical technologies such as

robotic surgery,[17] hospital adoption of implantable cardiac defibrillators,[18] and physician willingness to participate in research studies.[19] The same logic underlying each of these examples applies here: a module's neighbors in a co-expression network may carry information about a phenotype that the module's own activity does not.

In recent work extending stepped-wedge cluster-randomized trials to account for interference, we represented a physician collaboration network as a weighted adjacency matrix, row-normalized it to a transition matrix $\mathbf{W}$ whose entry $(\mathbf{W})_{ij}$ is the share of influence acting upon actor i from actor j (by construction of the network, $W_{ii} = 0$), and defined each actor's network exposure as the neighbor-weighted sum of other actors' treatment, $\boldsymbol{e} = \mathbf{W}\,\boldsymbol{d}$.[13,14] Including this exposure term alongside the direct-treatment term in an outcome model separates an actor's own effect from the peer-to-peer influence reaching it through the network. Ignoring such spillover biases the estimated direct effect when interference is present.[13] In the following, our construction is deliberately generic: it requires only a weighted network in which the units are nodes, a per-unit quantity to propagate, and an outcome.

We adapt the above statistical approach used to account for spillover effects in human and institutional/organizational studies in which the actors have the potential to influence one another such that one actors outcome might be affected by another actor's treatment ("intervening on the untreated") to gene co-expression networks. We treat co-expression modules as the units and a gene network as the conduit of influence, applying the same construction to a distinct biological question. This decomposes each module's relationship to an outcome into a direct term (the module's own activity) and a spillover term (the exposure it receives from other modules through between-module edges). We then ask whether cross-module spillover predicts the outcome beyond direct activity. We apply the method to a longitudinal cynomolgus macaque model of *Mycobacterium tuberculosis* infection in which whole-blood transcriptomes are paired with lung FDG avidity, an imaging measure of disease severity.[20] We provide the framework as an R package, QUESTIONS.

2. The QUESTIONS framework

2.1. *Direct and cross-module spillover on a co-expression network*

Let $\mathbf{A}$ be the symmetric weighted adjacency matrix of a gene co-expression network on p genes, partitioned into modules by a community assignment $m(\cdot)$. A biological network carries signal between modules, not only within them. We therefore treat each module as a unit of analysis whose outcome has two quantities: its own state (the direct term) and a neighbor-weighted summary of the states of the modules it connects to (the exposure, or spillover, term). The weights come from an (optionally) row-normalized network.[12–14,21] We row-normalize the network so that a module's spillover reflects the proportion of its connections that are active rather than their absolute number. This suits correlation-based edge weights and prevents large modules from registering greater spillover merely as an artifact of their size. Retaining the unnormalized adjacency is a defensible alternative under which influence is treated as cumulative rather than proportional, at the cost of a spillover coefficient whose magnitude is entangled with module connectivity.

To make the spillover term carry only cross-module influence, we first row-normalize the

full adjacency matrix, $P_{gh} = A_{gh} / \sum_{h'} A_{gh'}$, so that each row of $\mathbf{P}$ is a gene's distribution of influence over all of its neighbors, and then retain only the between-module entries,

$$W_{gh} = P_{gh} \mathbb{1}\{\, m(g) \neq m(h) \,\}. \tag{1}$$

We do not renormalize after this step, so the row sum of $\mathbf{W}$ is the fraction of gene g's influence that leaves its own module. This has a useful consequence: a gene wired predominantly to its own module contributes little cross-module exposure, because most of its normalized weight was assigned to within-module neighbors that are then removed. Restricting to between-module weights prevents the spillover term from re-expressing a module's own internal activity, so that direct and spillover terms measure distinct quantities by construction.

For a unit of observation i (a sample, subject, or, in the case study below, an animal), let $\boldsymbol{d}_i \in \mathbb{R}^p$ be a gene-level effect vector, for example each gene's departure from a reference state (defined below). The gene-level spillover reaching each gene is the neighbor-weighted departure $\boldsymbol{e}_i = \mathbf{W}\,\boldsymbol{d}_i$, exactly paralleling the social-network exposure $\boldsymbol{e} = \mathbf{W}\boldsymbol{d}$.[13] We then summarize both quantities per module.

Let $G_M \subseteq \{1, \ldots, p\}$ index the genes of module M, and let $\mathbf{X}_i \in \mathbb{R}^{|G_M|}$ denote unit i's standardized expression of those genes. The module eigengene[22] is the first principal component, $\mathrm{eig}_{iM} = \mathbf{v}_M^\top \mathbf{X}_i$, where $\mathbf{v}_M$ is the leading loading vector of the standardized module-expression matrix, sign-aligned so that $\mathbf{v}_M$ correlates positively with mean module expression. We define a departure operator relative to unit i's own baseline samples $\mathcal{B}_i$,

$$\Delta(z_i) \;=\; z_i - \frac{1}{|\mathcal{B}_i|} \sum_{b \in \mathcal{B}_i} z_b, \tag{2}$$

which subtracts a unit's mean over its baseline samples. The direct term for module M is the eigengene expressed as a departure, $\mathrm{direct}_{iM} = \Delta(\mathrm{eig}_{iM})$, and the gene-level effect vector propagated above is the per-gene departure, $(\boldsymbol{d}_i)_g = \Delta(x_{ig})$ for gene g. The spillover term is the mean of the gene-level spillover over the module's genes,

$$\mathrm{spill}_{iM} \;=\; \frac{1}{|G_M|} \sum_{g \in G_M} (\boldsymbol{e}_i)_g. \tag{3}$$

In words, the direct term is how much module M itself departed from its reference, and the spillover term is how much departure reached M from the rest of the network.

2.2. *Outcome model*

Given an outcome y_i, the framework fits, for each module M,

$$y_i = \beta_0 + \beta_{\mathrm{dir}}\, \mathrm{direct}_{iM} + \beta_{\mathrm{spill}}\, \mathrm{spill}_{iM} + \varepsilon_i, \tag{4}$$

with the direct and spillover terms standardized so that β_{spill} is interpretable per standard deviation sized change of the spillover predictor. The estimand of interest is β_{spill}: whether a module's cross-module spillover predicts the outcome beyond the module's own activity. The link function is chosen to suit the outcome (identity, logistic, log, or otherwise). Because the underlying network is an undirected co-expression graph, in which edges may arise through a selective process such as preferential attachment[23] rather than random assignment, β_{spill} is an association rather than a causal effect.

2.3. *Inference for small samples*

Given the small sample sizes typical of omics studies, we pair the outcome model with distribution-free and resampling inference to interpret β_{spill} from the full (spillover) model relative to the reduced (direct-only) model. The partial spillover effect is tested by a Freedman–Lane permutation scheme, which permutes the residuals of the direct-only model and refits, yielding a distribution-free test of β_{spill} that conditions correctly on the direct term;[24,25] when many module × condition combinations are tested, permutation p-values are adjusted by Benjamini–Hochberg over a pre-specified candidate family.[26] Coefficient stability is summarized by bootstrapping units to test for sign-consistency and to estimate confidence intervals of β_{spill};[27] added predictive value is summarized by repeated k-fold cross-validation, as the gain in out-of-sample R^2 from adding the spillover term to a direct-only model; and leave-one-unit-out refitting is used as an influence check.

2.4. *Implementation*

The framework is implemented in the R package QUESTIONS (Quantifying Expression Spillover Trajectories in Omic Network Series), which constructs the between-module weight matrix, computes each module's direct and spillover effects, and runs the outcome model with the validation battery above. It is freely available at `https://github.com/CarlyBobak/QUESTIONS`.

3. Case study: tuberculosis disease severity

3.1. *Data and preprocessing*

We applied the framework to GSE84152, a longitudinal study of *Mycobacterium tuberculosis* infection in cynomolgus macaques in which whole-blood transcriptomes were profiled on an Illumina HumanHT-12 v4 array (human gene symbols) at serial time points from two pre-infection baselines through six months post-infection, with per-animal lung FDG avidity measured by PET-CT as a surrogate for disease severity.[20] The outcome is total lung FDG avidity; because it is strongly right-skewed we modeled it on the $\log(1+x)$ scale, standard for total-glycolytic PET summaries and robust to a few high-burden animals. Expression was corrected for batch structure across processing runs with the empirical-Bayes ComBat method.[28] Each per-animal, per-gene departure was computed relative to that animal's own pre-infection baseline, so the propagated quantity is a within-animal response and between-animal baseline differences are removed. The analysis included $n = 37$ animals. Nine time points were profiled per animal (two pre-infection baselines and seven post-infection: days 7, 10, 20, 30, 42, and 56, and month 3). Pooling the per-animal, per-time-point profiles that passed quality control yields the 337 windowed samples used to build the network.

3.2. *Consensus network construction*

To focus the network on genes that respond to infection rather than on constitutive covariation, we restricted the analysis to infection-responsive genes identified by a moderated F-test for any change across any time points, with within-animal correlation accommodated by

treating animal as a blocking factor via `duplicateCorrelation` and using an empirical-Bayes variance estimator;[29–31] genes passing a false-discovery threshold of 0.01 were retained (3,829 genes). Over these genes we built a single consensus network, aggregated across 337 windowed samples so that it reflects coordination sustained across the time course rather than the structure of any one time point. Gene pairs were joined by a signed mutual k-nearest-neighbor rule ($k = 50$): an edge is kept only when each gene lies in the other's k-neighborhood.[32] The reciprocal criterion prunes one-sided ties, yielding a sparse, modular graph that preserves the between-module bridges the spillover analysis requires.

3.3. *Module identification*

Communities were detected with Infomap, a flow-based method that partitions the graph by compressing the description of a random walk over it;[33] each community defines a module, and each module is summarized by its eigengene as in Section 2. The resulting network is strongly modular (modularity 0.761, within-module edge fraction 0.83).

3.4. *Pathway enrichment and module annotation*

Modules were annotated by over-representation analysis against curated gene-set collections, using the responsive-gene network as background so that enrichment reflects structure within the analyzed set.[34,35] Because a single immune-centric collection under-annotates non-immune modules and a naive top-hit-per-collection rule is dominated by large generic terms, we combined a cell-type call from the Blood Transcription Modules[36] with a specificity-scored (term frequency–inverse document frequency, TF–IDF) pooling of hits across each of GO biological process, Reactome, WikiPathways, and MSigDB Hallmark.[37] Enrichment is considered corroborating evidence of the network's biological coherence, not as independent statistical validation of the spillover results, since the same modules carry both.

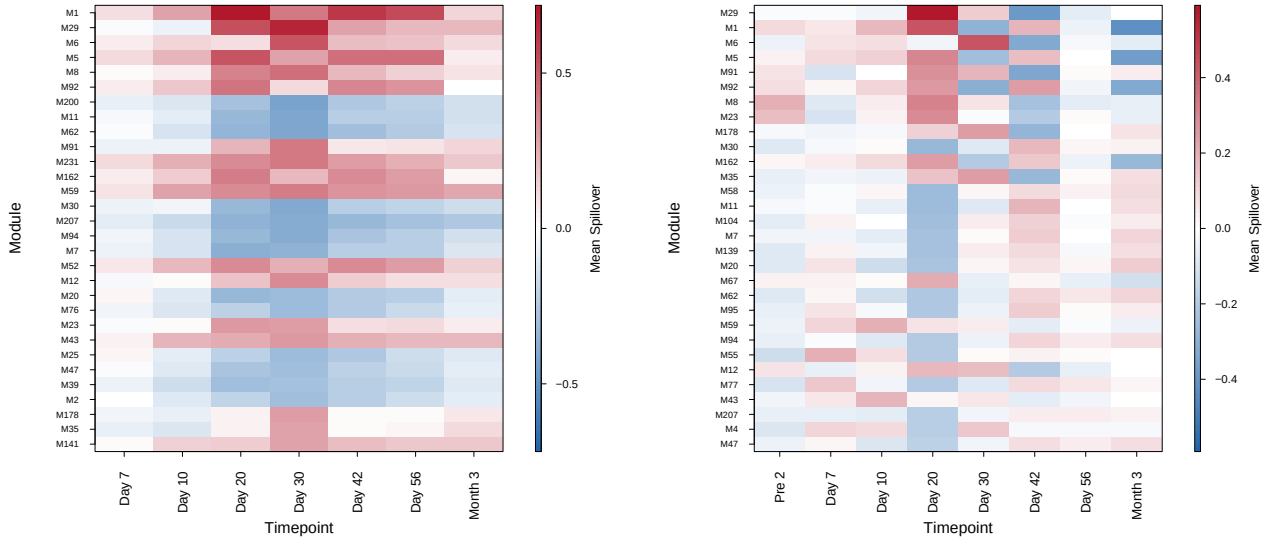
3.5. *Candidate selection and testing*

Applying the outcome model of Section 2, candidate module $\times$ time-point combinations were restricted, before examining the outcome, to the modules with the most pronounced descriptive spillover dynamics (Figure 1). This outcome-independent pre-selection (the "candidate family") fixes the set of tests before any outcome is examined. The 15 modules with the most pronounced descriptive spillover dynamics, each evaluated at the seven post-baseline time points, give the 105 tests over which Benjamini–Hochberg correction is applied. The direct and spillover terms were standardized, the model fit with an identity link on the $\log(1 + x)$ outcome across $n = 37$ animals, and β_{spill} evaluated with the permutation, bootstrap, cross-validation, and leave-one-out procedures described in Section 2.

4. Results

4.1. *Cross-module spillover predicts severity beyond direct activity*

Across the pre-specified family of 105 module $\times$ time-point tests, spillover clears Benjamini–Hochberg adjustment for five combinations, all in myeloid or platelet modules and all in



(a) Spillover of baseline departure.

(b) Spillover of the timepoint increment.

Fig. 1: **Cross-module spillover dynamics over the infection time course, and the two questions spillover can ask.** Each cell is the mean cross-module spillover received by a module (row) at a time point (column): the between-module, neighbor-weighted departure defined in Section 2, averaged over animals and over the module's genes. Warm values indicate a module receiving positive cross-module influence, cool values the reverse. The panels differ only in how each gene's change is defined before it is spread across the network: relative to the pre-infection baseline in (a), or as the change from the preceding time point in (b). In (a) spillover reads as the cumulative cross-module influence that has accrued relative to the healthy state. In (b) spillover reads as the influence being exchanged at each interval. The two views pose different temporal questions of the same network, cumulative burden versus instantaneous exchange, and clarify that spillover is a lens applied to a chosen departure rather than a fixed quantity. Modules with the most pronounced spillover dynamics, selected independently of the outcome, defined the candidate set carried to the outcome model.

the peak-response window (Table 1). The strongest are cross-module spillover into the neutrophil/inflammatory module M8, at both day 42 and month 3, and into the platelet module M6 at day 56, each at adjusted $p = 0.017$; spillover into the monocyte module M91 follows at day 30 and day 42 (adjusted $p = 0.026$ and 0.052). All five carry a large positive standardized spillover coefficient (for example M8 at day 42, $\beta_{\text{spill}} = 8.14$, bootstrap 95% CI $[5.42, 11.67]$; M91 at day 30, 7.72, $[3.01, 11.60]$; M6 at day 56, 2.45, $[0.47, 3.98]$), are sign-stable in 100% of bootstrap resamples, are robust to leave-one-animal-out, and improve out-of-sample prediction (cross-validated ΔR^2 between +0.23 and +0.39). The spillover p-values are corrected across the full family whereas the direct-term p-values are reported unadjusted, so this comparison holds spillover to the stricter standard. The signal concentrates in the day-42 to month-3 window, coinciding with the peak transcriptional response reported for this cohort.[20]

Table 1: Cross-module spillover predicts lung FDG avidity beyond a module's own activity. Outcome is $\log(1 + \text{total lung FDG avidity})$; β_{spill} is the standardized spillover coefficient (per SD), with a bootstrap 95% confidence interval over animals. p_{perm} is a Freedman–Lane permutation p-value for the partial spillover effect; p_{BH} is Benjamini–Hochberg adjusted p-values across all 105 module × timepoint tests. ΔR^2_{CV} is the repeated 5-fold cross-validated gain in out-of-sample R^2 from adding spillover to a direct-only model. p_{direct} is the direct-term (module eigengene) p-value in the same model, reported unadjusted, so the spillover effect is held to a stricter multiple-testing standard than the direct term. All five combinations are sign-stable in 100% of bootstrap resamples and robust to leave-one-animal-out.

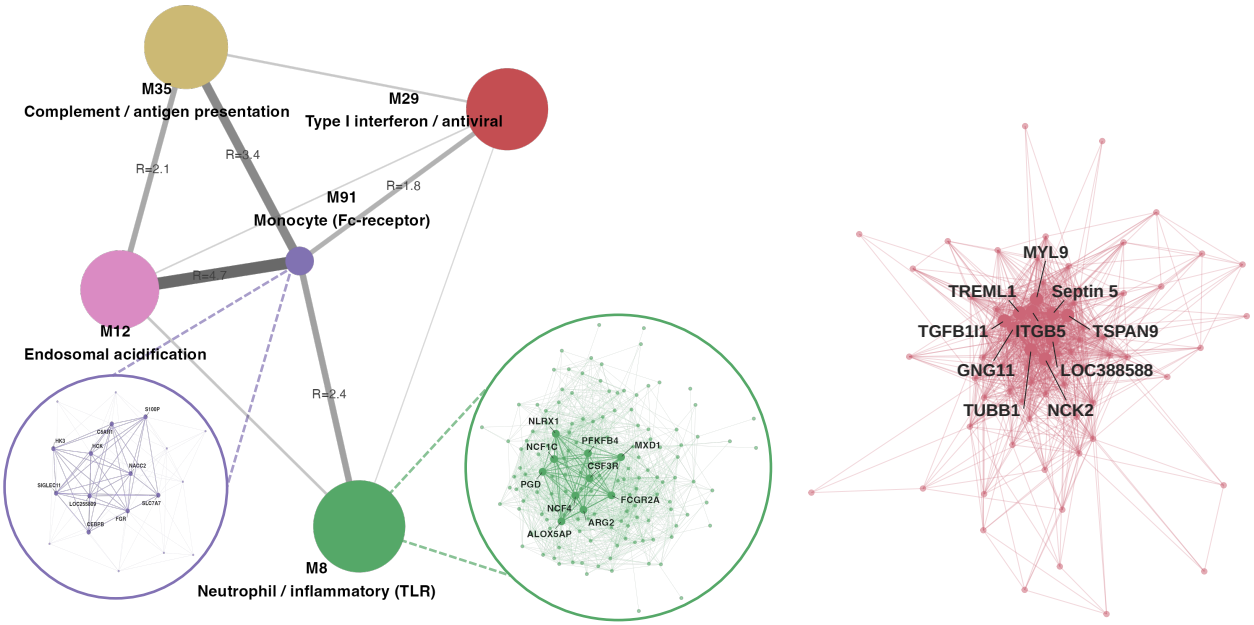
Module	Time	β_{spill}	95% CI	p_{perm}	p_{BH}	ΔR^2_{CV}	p_{direct}
M8 (neutrophil/inflammatory)	D42	8.14	[5.42, 11.67]	5×10^{-4}	0.017	+0.389	2.9×10^{-4}
M8 (neutrophil/inflammatory)	M3	7.27	[3.82, 11.18]	5×10^{-4}	0.017	+0.296	3.1×10^{-3}
M6 (platelet/coagulation)	D56	2.45	[0.47, 3.98]	5×10^{-4}	0.017	+0.227	0.118
M91 (monocyte, Fc-receptor)	D30	7.72	[3.01, 11.60]	1×10^{-3}	0.026	+0.230	5.8×10^{-3}
M91 (monocyte, Fc-receptor)	D42	6.90	[3.18, 11.25]	2×10^{-3}	0.052	+0.230	4.1×10^{-3}

4.2. *Direct activity and spillover capture distinct information*

The platelet module M6 best illustrates what the spillover value. Its cross-module spillover is among the strongest predictors of severity (day 56, adjusted $p = 0.017$; Figure 2b), yet its own activity is uninformative (direct-term $p = 0.118$). Thus, this framework allows a module whose activity carries no outcome signal to still receive outcome-relevant spillover through the network. The interferon module M29 shows the opposite pattern. Its spillover never clears correction (best adjusted $p = 0.62$), while its own activity is strongly associated with severity (direct-term $p = 2.8 \times 10^{-5}$ at day 42). M6 carries outcome information through its neighbors and M29 through its own activity, which shows that the two terms capture different information rather than restating each other. A binary high/low FDG outcome recovered essentially no spillover signal, consistent with dichotomization discarding graded severity.

4.3. *The unsupervised network recovers coherent blood biology*

The network recovers a coherent map of the major hematopoietic lineages and immune programs, erythroid, translation, neutrophil, interferon, complement, T, B, NK, monocyte, plasma, and platelet modules, with no cell-type input; the full network is shown in Figure 3). The outcome-associated modules sit where TB immunopathology would place them: M8 and M91 in the myeloid-inflammatory compartment active during the peak-response window (Figure 2a), and M6 in platelet/coagulation biology (Figure 2b). The myeloid core is densely intercoupled while M6 is peripheral, coupling to the myeloid modules at near-zero degree-adjusted strength, so the M6 spillover signal does not travel an identifiable myeloid route. Spillover reaches a module over all of its between-module edges regardless of the confidence tier of the neighbor. M6's signal therefore arrives through its individually weak, non-myeloid connections rather than any strong link to the myeloid core, which is why no single coupling in Figure 2 accounts for it.



(a) Myeloid-core meta-network. (b) Platelet module M6.

Fig. 2: **Network structure behind the spillover result.** (a) Meta-network of the myeloid core: each node is a module, sized by gene count, and each edge a degree-adjusted coupling strength R between modules, with only above-chance couplings drawn and edge width scaled to R. The monocyte module M91 sits at the hub, coupling most strongly to the endosomal module M12 ($R = 4.7$) and linking the complement (M35), interferon (M29), and neutrophil (M8) modules. Gene-level zoom insets show the internal hub structure of M8 and M12. Spillover into M8 and M91 during the peak-response window predicts disease severity. (b) The platelet/coagulation module M6 at gene level, with platelet-specific hubs (MYL9, TREML1, ITGB5, TUBB1, GNG11). In contrast to the densely inter-coupled myeloid core, M6 is peripheral: its degree-adjusted coupling to the myeloid modules is near zero. Its spillover is associated with severity even though its own activity is not (Section 4)

5. Discussion

We have shown that network spillover, a construct developed in the analysis of interference in social networks, transfers to gene co-expression networks and captures temporal biological signals that a module's own activity does not. The framework separates a module's own activity from the activity that reaches it across between-module edges, and tests whether the latter predicts an outcome beyond the former. In a longitudinal model of tuberculosis this decomposition identified cross-module spillover into myeloid modules as predictive of disease severity, stably and out of sample, and produced a clean case, the platelet module, in which the module's own activity was uninformative while its spillover was not.

The platelet module M6 invites an analogy to interference in randomized trials, where a control unit that acquires exposure through its treated peers can appear affected despite

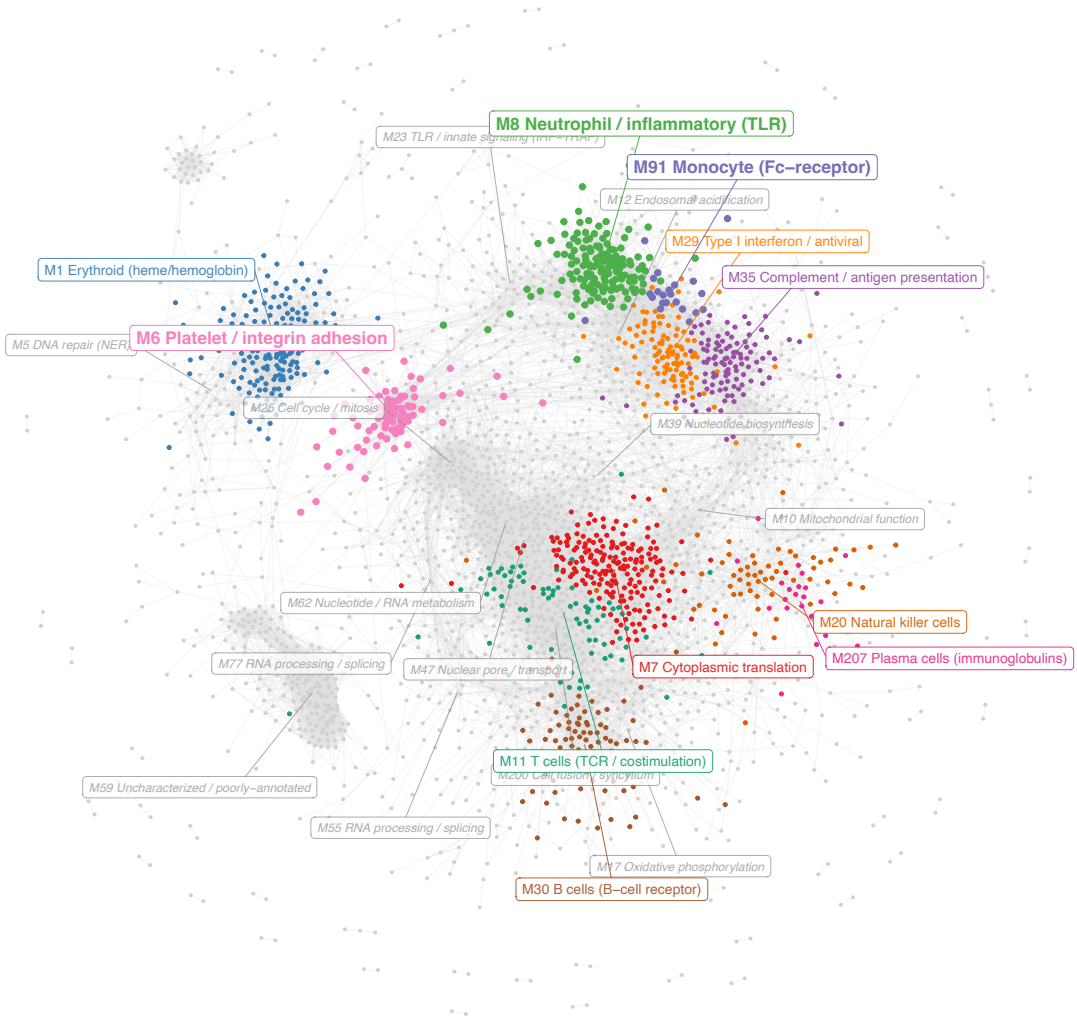


Fig. 3: **The full infection-responsive co-expression network.** All non-isolated genes, laid out by a force-directed algorithm and colored by module (high- and medium-confidence modules only; tentative modules in gray italic, unannotated modules unlabeled). Isolated singletons are omitted. The network recovers the major hematopoietic lineages and immune programs as distinct communities with no cell-type input; the outcome-associated modules M6, M8, and M91 are emphasized.

no direct assignment. Here, M6's own activity is uninformative about severity, yet outcome-relevant signal reaches it across the network. The analogy is partial. In a trial, contamination is a nuisance that biases the treatment-effect estimate and is adjusted away. Here, the neighbor-borne signal is the quantity of interest.

This paper fills a gap left by existing approaches to network structure in biology. Network propagation diffuses a seed signal over a fixed network to amplify and prioritize genes,[8,9] and differential and single-sample co-expression methods quantify how wiring changes between conditions or individuals and relates that change to outcomes;[10] both establishing that network

structure carries phenotypic information. What the spillover decomposition adds is the explicit separation of a module's own signal from its neighbors' contribution within a single outcome model, so that the neighbor contribution can be tested on its own terms. The two descriptive views of spillover in Figure 1, one relative to a pre-infection baseline and one as a timepoint increment, show that the same construction answers different temporal questions depending on the departure that is propagated.

Although the framework is the focus, the biological hits are plausible and literature-consistent, which supports the method's validity. The modules whose spillover tracked severity are the myeloid neutrophil and monocyte programs, and the network recovered the canonical type I interferon signature as a distinct module, all of which are central to the blood transcriptional response to *M. tuberculosis* and to the source cohort's own findings on FDG-defined severity.[20] The platelet module is likewise coherent with a growing literature on platelet involvement in tuberculosis immunopathology.[38] We emphasize that these associations are correlational and that a full immunological interpretation is beyond the scope of this paper.

Several limitations bound these conclusions. The most important is that every result is conditional on the inferred network. Spillover is defined over the between-module edges, so a mis-estimated network, wrong edges, missed edges, or an unstable community partition, propagates directly into the spillover term and into every downstream test. Co-expression networks are sensitive to construction choices such as the similarity measure, sparsification, and community-detection algorithm, and our use of a single fixed consensus network conditions the entire analysis on one such choice. A different but equally defensible network could shift which modules and bridges carry the signal. Second, and common to longitudinal omics, the sample size is small ($n = 37$ animals from a single cohort with no external replication), which is why inference rests on permutation, bootstrap, cross-validation, and leave-one-out rather than on parametric p-values, and why candidate modules were pre-specified independently of the outcome. The surviving associations should nonetheless be read as strong internal evidence awaiting replication rather than as established effects. Third, the enrichment that anchors the biological reading corroborates the network's coherence but does not independently validate the spillover results, since the same modules carry both. Fourth, the direction of influence is not identified. The network is undirected co-expression, and although temporal ordering offers a directionality proxy, spillover between modules does not establish that one module drives another, so the results are associational. Fifth, the module summaries (eigengenes) and the mean-over-genes spillover summary discards within-module heterogeneity, and results can depend on the resolution at which communities are defined. Finally, the platform is an older microarray with deprecated probe annotations, which limits annotation of some modules, and the outcome is a single imaging surrogate of severity rather than a clinical endpoint.

The clearest extensions concern the network itself. Our analysis propagates per-animal departures over one network fixed across the time course, which assumes the wiring is stable even as expression changes; a natural next step is to let the network vary over time, so that spillover is computed on timepoint-specific or window-specific networks and can capture rewiring rather than only changing node states. A complementary direction operates at the level of individuals. Single-sample network inference, in particular LIONESS, estimates a

network for each sample by measuring how the aggregate network shifts when that sample is left out,[39] so that spillover could be computed on per-sample networks and related to per-sample outcomes, turning the present fixed-network analysis into one that also captures individual network heterogeneity. Because such per-sample networks yield an individual-level spillover profile, this is also the natural route to patient stratification and precision-medicine applications of the framework. Both directions would let the framework ask not only whether a module's neighbors carry outcome information, but whether the connections themselves reorganize over time or across individuals in outcome-relevant ways. Further extensions include incorporating lagged spillover terms to model influence that acts across, rather than within, time points, and applying the same decomposition to other network types and omics layers, since the construction requires only a weighted network over units, a per-unit quantity to propagate, and an outcome.

6. Conclusion

Network spillover offers a principled way to ask a question that standard co-expression analysis leaves unaddressed: whether the activity of a module's network neighbors carries information about an outcome beyond the module's own activity. By adapting the direct-versus-spillover decomposition from the analysis of interference in social networks, and pairing it with small-sample inference, the QUESTIONS framework makes that question testable on temporal omics data. In a longitudinal model of tuberculosis it identified cross-module spillover into myeloid modules as predictive of disease severity and recovered coherent blood biology without supervision, including a platelet module whose spillover predicted severity while its own activity did not. The approach is general, requiring only a weighted network over units, a per-unit quantity to propagate, and an outcome, and it extends naturally to time-varying and single-sample networks. As longitudinal and multi-omic studies grow, network spillover effect modeling and associated statistical tests provide a compact, interpretable lens on how influence travels through biological networks.

Use of large language models

The network-spillover methodology adapted here was developed in the social-network setting without the use of large language models.[13,14] In preparing this adaptation to gene co-expression networks, the authors used an LLM-based assistant in a supporting role: to help draft and revise manuscript text, to assist in writing and debugging R-package code, and to help prepare figures and documentation. All methodological, modeling, and scope decisions were made by the authors, who ran every analysis, verified all code and numerical results against the underlying data, and take full responsibility for the content. The assistant did not generate data or produce results autonomously.

References

1. L. H. Hartwell, J. J. Hopfield, S. Leibler and A. W. Murray, From molecular to modular cell biology, *Nature* **402**, C47 (December 1999).

2. P. Langfelder and S. Horvath, WGCNA: An R Package for Weighted Correlation Network Analysis, *BMC Bioinformatics* **9**, p. 559 (2008).
3. S. van Dam, U. Võsa, A. van der Graaf, L. Franke and J. P. de Magalhães, Gene Co-expression Analysis for Functional Classification and Gene–Disease Predictions, *Briefings in Bioinformatics* **19**, 575 (2018).
4. K. Greenham and C. R. McClung, Time to Build on Good Design: Resolving the Temporal Dynamics of Gene Regulatory Networks, *Proceedings of the National Academy of Sciences* **115**, 6325 (2018).
5. M. Toolabinejad and others, A Comprehensive Survey of Gene Co-expression Network Analysis: Methods, Tools, Challenges, and Future Directions, *Network Modeling Analysis in Health Informatics and Bioinformatics* **14** (2025).
6. M. Kuismin and M. J. Sillanpää, tvsfglasso: Time-varying Scale-free Graphical Lasso for Network Estimation from Time-series Data, *PLOS Computational Biology* **21**, p. e1013710 (2025).
7. S. Ozminkowski, L. Hou, D. R. Jacobs and H. Jiang, Dynamic Co-Expression Network Estimation via Multivariate Mixed-Effects Models (2026).
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. Visonà, E. Bouzigon, F. Demenais and G. Schweikert, Network Propagation for GWAS Analysis: A Practical Guide to Leveraging Molecular Networks for Disease Gene Discovery, *Briefings in Bioinformatics* **25**, p. bbae014 (2024).
10. L. Wei, Y. Xin, M. Pu and Y. Zhang, Patient-specific Analysis of Co-expression to Measure Biological Network Rewiring in Individuals, *Life Science Alliance* **7**, p. e202302253 (2024).
11. D. M. Gysi, T. de Miranda Fragoso, F. Zebardast, W. Bertoli, V. Busskamp, E. Almaas and K. Nowick, Whole Transcriptomic Network Analysis Using Co-expression Differential Network Analysis (CoDiNA), *PLOS ONE* **15**, p. e0240523 (2020).
12. M. G. Hudgens and M. E. Halloran, Toward Causal Inference With Interference, *Journal of the American Statistical Association* **103**, 832 (2008).
13. A. J. O'Malley, C. A. Bobak and A. E. Barnato, Causal Inference for Intervention Spillover in a Stepped-Wedge Cluster-randomized Trial: Lessons from a Physician Network, *Social Networks* **86**, 431 (2026).
14. C. A. Bobak, M. A. Murphy, A. E. Barnato and A. J. O'Malley, Gaming the System: Evaluating Spillover in a Video Game Intervention for Advance Care Planning using Physician Social Networks, *Health Services and Outcomes Research Methodology* (2026).
15. A. J. O'Malley, T. A. Bubolz and J. S. Skinner, The diffusion of health care fraud: A bipartite network analysis, *Social Science & Medicine* **327**, p. 115927 (June 2023).
16. X. Ran, N. E. Morden, E. Meara, E. L. Moen, D. N. Rockmore and A. J. O'Malley, Exploiting relationship directionality to enhance statistical modeling of peer-influence across social networks, *Statistics in Medicine* **43**, 4073 (September 2024).
17. G. Chen and A. J. O'Malley, Developing and Comparing Four Families of Bayesian Network Autocorrelation Models for Binary Outcomes: Estimating Peer Effects Involving Adoption of Medical Technologies, *Biometrical Journal. Biometrische Zeitschrift* **67**, p. e70030 (February 2025).
18. A. J. O'Malley, E. L. Moen, J. P. W. Bynum, A. M. Austin and J. S. Skinner, Modeling peer effect modification by network strength: The diffusion of implantable cardioverter defibrillators in the US hospital network, *Statistics in Medicine* **39**, 1125 (April 2020).
19. C. A. Bobak, D. Mohan, M. A. Murphy, A. E. Barnato and A. J. O'Malley, Constructing within and between hospital physician social networks for modeling physician research participation, *BMC Medical Research Methodology* **23**, p. 253 (October 2023).
20. H. P. Gideon, J. A. Skinner, N. Baldwin, J. L. Flynn and P. L. Lin, Early Whole Blood Transcrip-

tional Signatures Are Associated with Severity of Lung Inflammation in Cynomolgus Macaques with Mycobacterium tuberculosis Infection, *The Journal of Immunology* **197**, 4817 (2016).
21. P. M. Aronow and C. Samii, Estimating Average Causal Effects Under General Interference, with Application to a Social Network Experiment, *The Annals of Applied Statistics* **11**, 1912 (2017).
22. P. Langfelder and S. Horvath, Eigengene Networks for Studying the Relationships Between Co-expression Modules, *BMC Systems Biology* **1**, p. 54 (2007).
23. A.-L. Barabasi and R. Albert, Emergence of Scaling in Random Networks, *Science* **286**, 509 (October 1999).
24. D. Freedman and D. Lane, A Nonstochastic Interpretation of Reported Significance Levels, *Journal of Business & Economic Statistics* **1**, 292 (1983).
25. A. M. Winkler, G. R. Ridgway, M. A. Webster, S. M. Smith and T. E. Nichols, Permutation Inference for the General Linear Model, *NeuroImage* **92**, 381 (2014).
26. Y. Benjamini and Y. Hochberg, Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing, *Journal of the Royal Statistical Society: Series B* **57**, 289 (1995).
27. B. Efron, Bootstrap Methods: Another Look at the Jackknife, *The Annals of Statistics* **7**, 1 (1979).
28. W. E. Johnson, C. Li and A. Rabinovic, Adjusting Batch Effects in Microarray Expression Data Using Empirical Bayes Methods, *Biostatistics* **8**, 118 (2007).
29. M. E. Ritchie, B. Phipson, D. Wu, Y. Hu, C. W. Law, W. Shi and G. K. Smyth, limma Powers Differential Expression Analyses for RNA-sequencing and Microarray Studies, *Nucleic Acids Research* **43**, p. e47 (2015).
30. G. K. Smyth, Linear Models and Empirical Bayes Methods for Assessing Differential Expression in Microarray Experiments, *Statistical Applications in Genetics and Molecular Biology* **3**, p. Article 3 (2004).
31. G. K. Smyth, J. Michaud and H. S. Scott, Use of Within-array Replicate Spots for Assessing Differential Expression in Microarray Experiments, *Bioinformatics* **21**, 2067 (2005).
32. M. R. Brito, E. L. Chávez, A. J. Quiroz and J. E. Yukich, Connectivity of the Mutual k-nearest-neighbor Graph in Clustering and Outlier Detection, *Statistics & Probability Letters* **35**, 33 (1997).
33. M. Rosvall and C. T. Bergstrom, Maps of Random Walks on Complex Networks Reveal Community Structure, *Proceedings of the National Academy of Sciences* **105**, 1118 (2008).
34. U. Raudvere, L. Kolberg, I. Kuzmin, T. Arak, P. Adler, H. Peterson and J. Vilo, g:Profiler: A Web Server for Functional Enrichment Analysis and Conversions of Gene Lists (2019 update), *Nucleic Acids Research* **47**, W191 (2019).
35. T. Wu, E. Hu, S. Xu, M. Chen, P. Guo, Z. Dai, T. Feng, L. Zhou, W. Tang, L. Zhan, X. Fu, S. Liu, X. Bo and G. Yu, clusterProfiler 4.0: A Universal Enrichment Tool for Interpreting Omics Data, *The Innovation* **2**, p. 100141 (2021).
36. S. Li, N. Rouphael, S. Duraisingham and others, Molecular Signatures of Antibody Responses Derived from a Systems Biology Study of Five Human Vaccines, *Nature Immunology* **15**, 195 (2014).
37. A. Liberzon, C. Birger, H. Thorvaldsdóttir, M. Ghandi, J. P. Mesirov and P. Tamayo, The Molecular Signatures Database Hallmark Gene Set Collection, *Cell Systems* **1**, 417 (2015).
38. D. E. Kirwan, D. L. W. Chong and J. S. Friedland, Platelet Activation and the Immune Response to Tuberculosis, *Frontiers in Immunology* **12** (May 2021).
39. M. L. Kuijjer, M. G. Tung, G. Yuan, J. Quackenbush and K. Glass, Estimating Sample-Specific Regulatory Networks, *iScience* **14**, 226 (2019).