Higher-Order Co-Infection on Hypergraphs

Katrina Liu
Mentor: Prof. Laura P. Schaposnik

Canyon Crest Academy

MIT PRIMES Conference 2026

Motivation

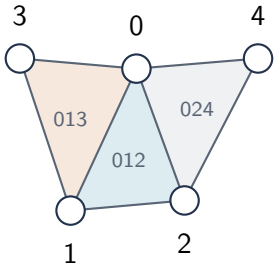
A problem in mathematical epidemiology

How does the structure of contact between people affect infection spread?

Pairwise models record *who contacts whom*.

Group transmission also depends on *who meets together*.

With two pathogens, we must also track simultaneous infection.



Overlapping contact groups.

What group information must a co-infection model retain?

Manuscript, Sections I–III. Background: Iacopini et al. (2019); Battiston et al. (2020). Group transmission is a modeling assumption.

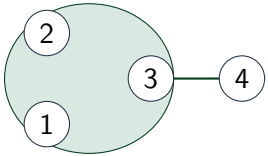
What is a hypergraph?

Definition

A **hypergraph** $H = (V, E)$ consists of vertices V and hyperedges E, where each hyperedge is a nonempty group of vertices.

$$V = \{1, 2, 3, 4\},$$

$$E = \Big\{\{1, 2, 3\}, \{3, 4\}\Big\}.$$



An ordinary graph has only two-person edges.

Interpretation: vertices are people; hyperedges are contact groups.

Manuscript, Sections I and III A. Contact groups motivate the model; transmission rules are assumptions.

Co-infection: four states per person

SIS: susceptible–infected–susceptible; recovery gives no lasting immunity.

$X_i^{(a)}(t) = 1$ if person i carries pathogen a at step t, and 0 otherwise. Here $a \in \{1, 2\}$ and there are n people.

State	Meaning	$(X_i^{(1)}, X_i^{(2)})$
S	Neither pathogen	$(0, 0)$
I_1	Pathogen 1 only	$(1, 0)$
I_2	Pathogen 2 only	$(0, 1)$
I_{12}	Both pathogens	$(1, 1)$

$$\underbrace{\rho_{12}(t)}_{\text{fraction carrying both}} = \frac{1}{n}\sum_{i=1}^{n} X_i^{(1)}(t) X_i^{(2)}(t).$$

Manuscript, Section III C. Co-infection means simultaneous carriage of both pathogens.

Research questions

1. How do group terms affect local disease-free stability?

Research questions

1. How do group terms affect local disease-free stability?
2. How much of the dynamics can pairwise models reproduce?

Research questions

1. How do group terms affect local disease-free stability?
2. How much of the dynamics can pairwise models reproduce?
3. Does the allocation of overlap reinforcement matter, even when each person's total coefficient is fixed?

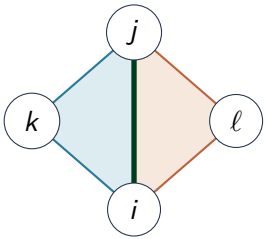
Manuscript, Sections IV–VI. We use a closure for local stability, exact structural arguments, and finite-time simulations.

Contact structure

Three-person groups: the smallest groups beyond pairs. Here $\mathcal{T} = \{\{i, j, k\}, \{i, j, \ell\}\}$.

$$A_{uv} = \begin{cases} 1, & u, v \text{ have a pairwise contact,} \\ 0, & \text{otherwise,} \end{cases} \qquad A_{uu} = 0.$$

Every triangle side is a contact. Repeated sides count once; additional pairwise contacts are allowed.



target

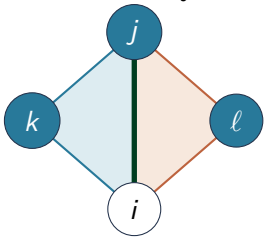
Activation threshold $r = 2$

A triangle contributes to a target only when **both other members carry the same pathogen**.

Manuscript, Sections III A–E. Distinct triples; binary pairwise adjacency A. The drawing will be reused on the next slide.

Three contributions to infection pressure

For one pathogen, $X_j = 1$ means that source j carries it; $X_j = 0$ otherwise.



target

Filled sources carry this pathogen.
Target i is susceptible.

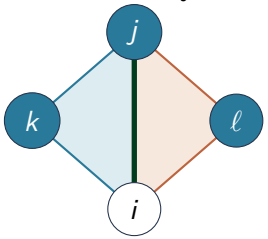
$$\lambda_i = \underbrace{\beta_1 N_i}_{\text{pairwise}} + \underbrace{\beta_h R_i}_{\text{triangle}} + \underbrace{\gamma O_i}_{\text{overlap}} .$$

λ_i: pressure during one step.
$\beta_1, \beta_h, \gamma \geq 0$: pressure coefficients.

N_i: infected pairwise neighbors.
R_i: active incident triangles.
O_i: unordered pairs of active incident triangles sharing a source besides i.

Three contributions to infection pressure

For one pathogen, $X_j = 1$ means that source j carries it; $X_j = 0$ otherwise.



target

Filled sources carry this pathogen.
Target i is susceptible.

$$\lambda_i = \underbrace{\beta_1 N_i}_{\text{pairwise}} + \underbrace{\beta_h R_i}_{\text{triangle}} + \underbrace{\gamma O_i}_{\text{overlap}} .$$

λ_i: pressure during one step.
$\beta_1, \beta_h, \gamma \geq 0$: pressure coefficients.

N_i: infected pairwise neighbors.
R_i: active incident triangles.
O_i: unordered pairs of active incident triangles sharing a source besides i.

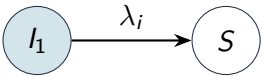
Here $N_i = 3$, $R_i = 2$, $O_i = 1$, $\boxed{\lambda_i = 3\beta_1 + 2\beta_h + \gamma.}$

Manuscript, Section III E. Threshold $r = 2$. Blue/orange regions are the two groups; the green shared side identifies their common source j. Full sums are in backup.

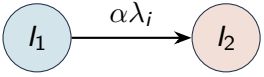
Transmission, co-infection, and recovery

For a target missing the considered pathogen,

$$p_i = \begin{cases} 1 - e^{-\lambda_i}, & \text{other pathogen absent,} \\ 1 - e^{-\alpha\lambda_i}, & \text{other pathogen present.} \end{cases}$$



p_i: per-step acquisition probability.



Acquiring pathogen 1
with the same unmodified pressure

$\alpha > 0$: multiplier of secondary infection pressure.
$\alpha = 1$: unchanged; $\alpha = 2$: doubled pressure.

$\mu \in [0, 1]$: per-step recovery probability for each carried pathogen.

Example: $\lambda_i = 0.1$, $\alpha = 2$ gives 9.5% versus 18.1% acquisition.

Updates are simultaneous and conditionally independent; no same-step reacquisition after recovery.

Manuscript, Sections III C–E. $1 - e^{-\lambda}$ can represent at least one Poisson opportunity with mean λ and exposure frozen during the step. α multiplies pressure, not probability.

Disease-free linearization

One-pathogen **mean-hazard closure**: replace products of distinct carrier indicators by products of infection probabilities x_i.

Linearization: keep first-order terms near $x = 0$

$$x(t+1) = Jx(t) + O(\|x(t)\|^2), \qquad \boxed{J = (1-\mu)I + \beta_1 A.}$$

J: Jacobian (matrix of first derivatives); I: identity matrix.
Triangle terms are quadratic; overlap terms are cubic. Neither contributes to J.

$\mathrm{spr}(B) = \max\{|\eta| : \eta \text{ is an eigenvalue of } B\}$ is the **spectral radius**.

$$\mathrm{spr}(J) = 1 - \mu + \beta_1 \mathrm{spr}(A).$$

Locally asymptotically stable if $\beta_1 \mathrm{spr}(A) < \mu$; unstable if $> \mu$. Equality is undecided by linearization.

Manuscript, Section IV A. Assume symmetric nonnegative A, $\beta_1, \beta_h, \gamma \geq 0$, $0 \leq \mu \leq 1$, distinct triangles and $r = 2$. A closure result, not an exact finite-process threshold.

Why simulate the model?

Beyond a local approximation

Simulations test finite-time outcomes from nonzero initial infections, where group and overlap terms can contribute.

Contact networks

800 people; 1,600 triangles.
4,800 pairwise contacts held fixed.
Reuse probability $p_{\text{reuse}} \in \{0, .25, .5\}$:
chance of attempting to reuse an existing side when forming a triangle.

Dynamics

$\beta_1 = .005$, $\beta_h = .1$, $\mu = .1$, $r = 2$.
$\alpha = 1$ unless stated otherwise.
Disjoint 2% seeds per pathogen; 200 steps.
Vary overlap reinforcement γ.

10 networks per reuse setting; 30 trials per network and condition. Averages include extinctions; uncertainty is assessed across networks.

Manuscript, Section V. Synthetic settings, not disease-calibrated rates. Reuse changes several structural properties. Full parameter rationale is in backup.

Effect of overlap reinforcement

Mean co-infection at step 200 ($\alpha = 1$)

Reuse p_{reuse}	$\gamma = 0$	$\gamma = 0.4$	Change (pp), 95% CI
0	3.90%	4.09%	0.19 $[-0.31, 0.67]$
0.25	12.32%	18.52%	6.20 $[4.24, 8.01]$
0.5	25.90%	40.28%	14.38 $[12.09, 16.38]$

Observation

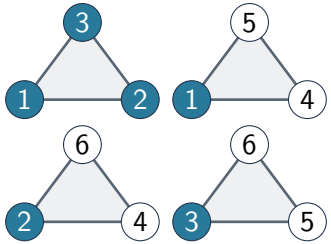
At reuse 0.25 and 0.5, stronger overlap reinforcement increased mean co-infection in these experiments.

Manuscript, Section V D. Within-network contrasts; graph-bootstrap intervals, no multiplicity adjustment. pp = percentage points. Changing reuse also changes other network properties.

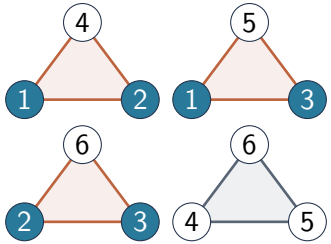
Same projection, different transitions

Each triangle below is one group; repeated labels denote the **same person**. Filled vertices $1, 2, 3$ are initially infected.

H_A



H_B



No susceptible target has an active triangle.

Each of $4, 5, 6$ has one active triangle.

Same pair multiplicities: each pair occurs once except $16, 25, 34$ (zero).

With one pathogen, $r = 2$, $\beta_1 = \gamma = 0$, $\beta_h > 0$, acquisition at each susceptible vertex is 0 in H_A, but $1 - e^{-\beta_h}$ in H_B.

Manuscript, Section IV B. Same fixed labeled initial state. The hypergraphs are isomorphic; this is not a comparison of relabeling-invariant averages. Orange triangles have an eligible target and two infected sources.

Pairwise models: what do we fit?

Three pairwise descriptions

- Binary contacts: whether a pair is connected.
- Weighted contacts: also count shared triangle memberships.
- Thresholded contacts: retain pairs whose weighted value is at least two.

Fit one transmission rate per model

On **4 training networks**, match endpoint **union prevalence**: the fraction carrying at least one pathogen.

Evaluate on **6 different networks**, without refitting.

Co-infection is not fitted.

Manuscript, Sections III F and V C. Original definitions: binary A, weighted $A + M$ where M counts triangle memberships, and threshold $A + M \geq 2$; zero diagonals. Exact formulas in backup.

Pairwise-model comparison

Binary, weighted, and thresholded projections; fit one transmission rate per model.
Fit on **4 graphs** to endpoint union prevalence (carrying either pathogen).
Evaluate on **6 different graphs**, 30 trials each, without refitting.

Held-out observable	HyCo-SIS	Pairwise models
Union prevalence at $t = 200$	67.8%	69.4–69.8%
Outbreak frequency at $t = 200$	85.6%	100%

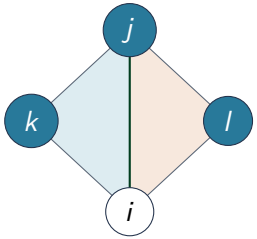
Co-infection averaged over steps 151–200:
pairwise estimates higher by 5.79–6.39 percentage points.

Matching endpoint total prevalence does not ensure agreement elsewhere.

Manuscript, Section V C. Clique, weighted, and threshold projections. Outbreak: union prevalence $\geq 5\%$ at $t = 200$. All three 95% graph-bootstrap intervals for terminal-window signed error exclude zero.

Overlap becomes a four-person interaction

For one pathogen, $X_j \in \{0, 1\}$ indicates whether source j is infected.



target

$$(X_j X_k)(X_j X_\ell) = X_j X_k X_\ell.$$

$X_j^2 = X_j$ because carriers are binary.

$$O_i(X) = \sum_{Q \ni i} W_{i,Q} \prod_{j \in Q \setminus \{i\}} X_j.$$

Q: a four-person subset of the population.
$W_{i,Q}$: number of contributing triangle pairs in Q for target i.

The same four-person group can have different coefficients at different targets.

Manuscript, Section VI A. Exact identity for distinct triangles and threshold $r = 2$.

When can one weight describe the whole group?

A **symmetric coefficient** $c_Q \geq 0$ assigns same weight to all four targets in group Q.

Exact representation criterion

For distinct triangles with threshold $r = 2$, replacing $W_{i,Q}$ by c_Q preserves every overlap pressure $O_i(X)$ in every binary state **iff** no four-set has exactly two or three triangle faces.

Triangle faces present in Q	Four target coefficients, up to order
0 or 1	$(0, 0, 0, 0)$
2	$(1, 1, 0, 0)$
3	$(3, 1, 1, 1)$
4	$(3, 3, 3, 3)$

Reason: distinct products of binary source indicators have unique multilinear coefficients. Equality in every state forces $W_{i,Q} = c_Q$.

Manuscript, Proposition VI.1. A triangle face is a three-person subset of Q present in the triangle collection.

A controlled redistribution

Hold each target's total coefficient fixed

$W_{i,Q}$: original weight; $V_{i,Q}$: redistributed weight.

$$\underbrace{s_i}_{\text{target } i\text{'s total}} = \sum_{Q \ni i} W_{i,Q} = \sum_{Q \ni i} V_{i,Q}.$$

1. Average the original target weights within each four-person group.

A controlled redistribution

Hold each target's total coefficient fixed

$W_{i,Q}$: original weight; $V_{i,Q}$: redistributed weight.

$$\underbrace{s_i}_{\text{target } i\text{'s total}} = \sum_{Q \ni i} W_{i,Q} = \sum_{Q \ni i} V_{i,Q}.$$

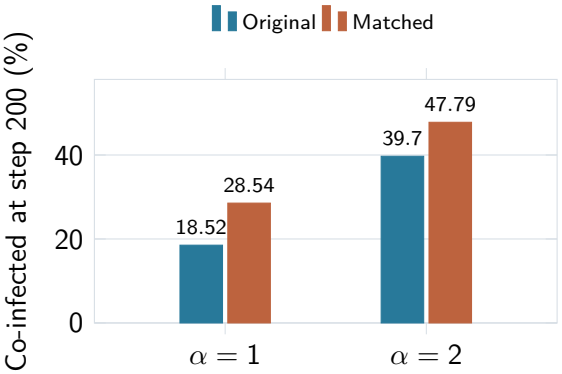
1. Average the original target weights within each four-person group.
2. Rescale the averaged weights at each target to restore its original total s_i (assign zero if $s_i = 0$).

Same networks, initial states, pairwise/triangle rules, and recovery rules.
The matched weights may still depend on the target; individual group totals may change.
Explicit normalization is in backup.

Manuscript, Section VI B. "Matched" means target-strength matched. The comparison changes allocation among source groups.

Co-infection after redistribution

800 people; 10 networks; 30 trials per network and model; 200 steps.



Matched minus original

+10.02 pp

$\alpha = 1$: unchanged susceptibility

95% interval $[6.23, 13.97]$

+8.09 pp

$\alpha = 2$: doubled secondary pressure

95% interval $[5.29, 10.98]$

pp = percentage points.

Intervals resample networks.

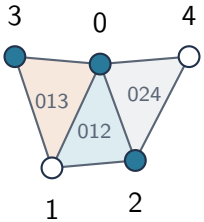
All trials included.

Synthetic disease parameters.

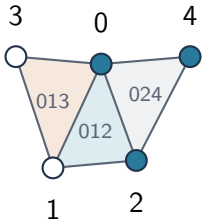
Manuscript, Section VI C, Table VII. $n = 800$, 1,600 triangles, reuse 0.25, $\gamma = .4$, $\beta_1 = .005$, $\beta_h = \mu = .1$, 2% seeds per pathogen. Intervals: paired graph bootstrap, no multiplicity adjustment. pp = percentage points.

Why redistribution can change acquisition

One pathogen on triangles $012, 013, 024$. Filled vertices are infected; target 1 is susceptible in both.



Infected: $\{0, 2, 3\}$



Infected: $\{0, 2, 4\}$

Target-1 pressure	Original	Matched
Infected $\{0, 2, 3\}$	0.615	0.415
Infected $\{0, 2, 4\}$	0.110	0.310

Key idea: Same total pressure, but more balanced. Since $f(\lambda) = 1 - e^{-\lambda}$ is concave ($f''(\lambda) = -e^{-\lambda} < 0$), redistribution changes acquisition.

Manuscript, Sec. VI E. $\beta_1 = .005$, $\beta_h = .1$, $\gamma = .4$. Source-triple weights $(1, 0) \to (1/2, 1/2)$. Same pair-edge union.

An exact one-step gain

Five people; triangles $012, 013, 024$ and their pairwise sides; one pathogen; $r = 2$.
Initially, choose **exactly three carriers uniformly** from the $\binom{5}{3} = 10$ subsets.

Special case of Proposition VI.2

Let ρ be the fraction carrying this pathogen; $\mathbb{E}$ averages over initial states and updates.
With $\beta_1, \beta_h \geq 0$ and $\gamma > 0$,

$$\mathbb{E}[\rho_{\text{matched}}(1)] - \mathbb{E}[\rho_{\text{original}}(1)] = \frac{F}{25} > 0,$$

$$F = (1 - e^{-\gamma/2}) \left(e^{-(2\beta_1 + \beta_h)} - e^{-(3\beta_1 + 2\beta_h + \gamma/2)} \right).$$

F is the summed acquisition gain at target 1 across the two preceding configurations.
Target 2 has the same gain:

$$\frac{2 \text{ targets} \times F}{10 \text{ configurations} \times 5 \text{ people}} = \frac{F}{25}.$$

Conclusions

1. In the mean-hazard closure, only pairwise transmission enters the disease-free linearization.

Conclusions

1. In the mean-hazard closure, only pairwise transmission enters the disease-free linearization.

2. Pairwise information can lose exact transition information; fitted pairwise models also differ on some tested outcomes.

Conclusions

1. In the mean-hazard closure, only pairwise transmission enters the disease-free linearization.

2. Pairwise information can lose exact transition information; fitted pairwise models also differ on some tested outcomes.

3. Preserving total overlap coefficients does not preserve simulated co-infection. A five-person proof explains a one-step mechanism.

Which source groups receive reinforcement matters.

Joint work of Katrina Liu and Laura P. Schaposnik. Manuscript, Sections IV–VI. Closure, finite-time simulation, and motif theorem claims have distinct scopes.

Acknowledgments

I would like to thank:

- My mentor, **Prof. Laura P. Schaposnik**, for her guidance throughout this project.

Acknowledgments

I would like to thank:

- My mentor, **Prof. Laura P. Schaposnik**, for her guidance throughout this project.

- The **MIT PRIMES-USA program and organizers**, for their support and the opportunity to conduct this research.

Thank you!

Selected references

K. Liu and L. P. Schaposnik. *Higher-Order Co-Infection on Hypergraphs*. Research manuscript, 2026.

I. Iacopini, G. Petri, A. Barrat, and V. Latora. *Simplicial models of social contagion*. Nature Communications **10**, 2485 (2019).

F. Battiston et al. *Networks beyond pairwise interactions: Structure and dynamics*. Physics Reports **874**, 1 (2020).

A. Salova and R. M. D'Souza. *Analyzing states beyond full synchronization on hypergraphs requires methods beyond projected networks*. arXiv:2107.13712 (2021).

E. Tan, M. Small, and S. D. Algar. *Do triangles matter? Replicating hypergraph disease dynamics with lower-order interactions*. Physical Review E **113**, 014315 (2026).

Bibliographic details follow the supplied manuscript. The talk's results and numerical values are from the joint manuscript.

Simulation settings and their purpose

Synthetic contact networks

800 people; 1,600 triangles.

Density chosen to avoid severe fragmentation.

4,800 pairwise contacts:
held fixed, giving mean degree 12.

Reuse probability $p_{\text{reuse}} = 0.25$:
probability an attempted triangle reuses an existing side pair.

Allocation comparison

Pairwise pressure	$\beta_1 = 0.005$
Triangle pressure	$\beta_h = 0.1$
Overlap pressure	$\gamma = 0.4$
Recovery probability	$\mu = 0.1$
Secondary multiplier	$\alpha = 1, 2$
Activation threshold	$r = 2$

Disjoint 2% seeds per pathogen; 200 steps; 10 networks and 30 trials per network and model. Include extinctions in averages.

Synthetic per-step settings, not fitted to a particular disease.

Pairwise and triangle counts

For one pathogen, let $X_j \in \{0, 1\}$ indicate whether person j carries it. Let $\mathcal{T}$ be the collection of distinct triangles and A the binary pairwise adjacency matrix, with $A_{ii} = 0$.

$$N_i(X) = \sum_j A_{ij} X_j, \qquad R_i(X) = \sum_{\{i,j,k\} \in \mathcal{T}} X_j X_k.$$

Every triangle side is a pairwise contact; repeated sides count once in A.

$$\lambda_i(X) = \beta_1 N_i(X) + \beta_h R_i(X) + \gamma O_i(X).$$

Here $\beta_1, \beta_h, \gamma \geq 0$. Threshold $r = 2$ requires both other members of a triangle to carry this pathogen. The next slide defines the overlap count O_i.

Manuscript, Sections III A–E. Pathogen and time indices are suppressed.

The overlap term

For a triangle $e \in \mathcal{T}$ containing target i, define its activation indicator

$$\phi_{e,i}(X) := \prod_{j \in e \setminus \{i\}} X_j.$$

Infectious overlap

$$O_i(X) = \sum_{e,f \in \mathcal{T}:\ i \in e \cap f,\ e<f} \left|(e \setminus \{i\}) \cap (f \setminus \{i\})\right| \phi_{e,i}(X)\phi_{f,i}(X).$$

- $e < f$ counts each unordered triangle pair once.

The overlap term

For a triangle $e \in \mathcal{T}$ containing target i, define its activation indicator

$$\phi_{e,i}(X) := \prod_{j \in e \setminus \{i\}} X_j.$$

Infectious overlap

$$O_i(X) = \sum_{e,f \in \mathcal{T}:\ i \in e \cap f,\ e<f} \left|(e \setminus \{i\}) \cap (f \setminus \{i\})\right| \phi_{e,i}(X)\phi_{f,i}(X).$$

- $e < f$ counts each unordered triangle pair once.
- Removing i excludes the trivial shared target.

The overlap term

For a triangle $e \in \mathcal{T}$ containing target i, define its activation indicator

$$\phi_{e,i}(X) := \prod_{j \in e \setminus \{i\}} X_j.$$

Infectious overlap

$$O_i(X) = \sum_{e,f \in \mathcal{T}:\ i \in e \cap f,\ e<f} \Big|(e \setminus \{i\}) \cap (f \setminus \{i\})\Big|\, \phi_{e,i}(X)\phi_{f,i}(X).$$

- $e < f$ counts each unordered triangle pair once.
- Removing i excludes the trivial shared target.
- For distinct triangles, the intersection factor is 0 or 1.

Manuscript, Section III E, specialized to distinct triangles. The overlap coefficient γ and the secondary-susceptibility coefficient α describe different mechanisms.

Transmission and recovery

Given all current carrier states X, let $a \in \{1, 2\}$ be the pathogen, $\bar{a} = 3 - a$ the other pathogen, and

$$u_i^{(a)}(X) := \mathbb{P}\Big(X_i^{(a)}(t+1) = 1 \mid X(t) = X\Big).$$

$$\boxed{u_i^{(a)}(X) = \begin{cases} 1 - \mu, & X_i^{(a)} = 1, \\ 1 - \exp\Big(-\alpha^{X_i^{(\bar{a})}} \lambda_i^{(a)}(X)\Big), & X_i^{(a)} = 0. \end{cases}}$$

Events are independent conditional on X; all nodes update simultaneously.

$$\rho_{12}(t) = \frac{1}{n} \sum_i X_i^{(1)}(t) X_i^{(2)}(t), \qquad \mathbb{E}[\rho_{12}(t+1) \mid X] = \frac{1}{n} \sum_i u_i^{(1)}(X) u_i^{(2)}(X).$$

Equivalent notation for the manuscript's synchronous transition rule, Sections III C–E. $\alpha = 1$ leaves secondary susceptibility unchanged; $\alpha > 1$ increases it.

Backup: what linearization means

One pathogen: $x_i(t)$ approximates person i's infection probability. Let $x = (x_1, \ldots, x_n)$ and A be the binary pairwise adjacency matrix.

Mean-hazard closure: replace distinct carrier products by products of probabilities

$$x(t+1) = F(x(t)), \quad F_i(x) = (1-\mu)x_i + (1-x_i)(1-e^{-h_i(x)}).$$

$h_i(x)$ is approximated infection pressure: pairwise terms are linear, triangle terms quadratic, and overlap terms cubic.

Linearization: retain first-order terms near the disease-free state $x = 0$.

$$x(t+1) \approx Jx(t), \qquad J_{ij} = \left.\frac{\partial F_i}{\partial x_j}\right|_{x=0}.$$

J is the **Jacobian**: its entries measure the first-order effect of one node's infection probability on another's next-step probability.

Manuscript, Section IV A. This is an approximation to stochastic dynamics; binary carrier indicators X_i are distinct from probability variables x_i.

Backup: spectral radius and local stability

Spectral radius: for a square matrix B,

$$\mathrm{spr}(B) := \max\{|\eta| : \eta \text{ is an eigenvalue of } B\}.$$

An eigenvector satisfies $Bv = \eta v$ for $v \neq 0$; its linear mode scales by η each step. Here spr avoids confusing spectral radius with prevalence ρ.

Jacobian of the mean-hazard closure at zero

I: identity matrix; A: symmetric nonnegative pairwise adjacency. Assume $\beta_1, \beta_h, \gamma \geq 0$, $0 \leq \mu \leq 1$, and distinct triangles with $r = 2$.

$$J = (1-\mu)I + \beta_1 A, \qquad \mathrm{spr}(J) = 1 - \mu + \beta_1 \mathrm{spr}(A).$$

$$\begin{aligned} \beta_1 \mathrm{spr}(A) < \mu &\quad : \text{locally asymptotically stable,} \\ \beta_1 \mathrm{spr}(A) > \mu &\quad : \text{unstable.} \end{aligned}$$

Equality is undecided by linearization. Quadratic and cubic terms have zero derivative at zero.

Pairwise projections and calibration

For $i \neq j$, let $M_{ij} := \#\{e \in \mathcal{T} : \{i, j\} \subset e\}$.

$$C = A, \qquad W = A + M, \qquad Q_{ij} = \mathbf{1}\{W_{ij} \geq 2\}.$$

All diagonals are zero. Here W denotes a projection matrix.

For a projection $P \in \{C, W, Q\}$,

$$\lambda_{i,P}^{(a)} = \widehat{\beta}_P \sum_j P_{ij} X_j^{(a)}.$$

Fit $\widehat{\beta}_P$ on four training graphs to endpoint union prevalence,

$$\rho_{\text{union}} = \frac{1}{n} \sum_i (X_i^{(1)} + X_i^{(2)} - X_i^{(1)} X_i^{(2)}).$$

Manuscript, Sections III F and V C. Rates selected from a prespecified grid, then held fixed on six evaluation graphs. Co-infection is not fitted.

Redistributing overlap coefficients

For original weights $W_{i,Q}$ on four-person groups Q, average within groups. Here s_i is the original target total and t_i the averaged total:

$$\overline{W}_Q := \frac{1}{4}\sum_{i\in Q} W_{i,Q}, \qquad s_i := \sum_{Q\ni i} W_{i,Q}, \qquad t_i := \sum_{Q\ni i} \overline{W}_Q.$$

Target-strength matched coefficients

$$\boxed{V_{i,Q} := \begin{cases} \overline{W}_Q\, s_i/t_i, & t_i > 0, \\ 0, & t_i = 0. \end{cases}}$$

$$\sum_{Q\ni i} V_{i,Q} = \frac{s_i}{t_i}\sum_{Q\ni i} \overline{W}_Q = s_i \quad (t_i > 0).$$

If $t_i = 0$, nonnegativity gives $s_i = 0$. Individual group totals may change.

Manuscript, Section VI B. The matched control remains target-dependent and does not isolate directionality alone. All other

Equal mean pressure does not fix acquisition

Let carriers X_j be independent, each infected with probability π. Here $\mathbb{E}$ means expectation. For any four-person set $Q \ni i$,

$$\mathbb{E}\left[\prod_{j \in Q \setminus \{i\}} X_j\right] = \pi^3.$$

Since $\sum_{Q \ni i} W_{i,Q} = \sum_{Q \ni i} V_{i,Q} = s_i$,

$$\boxed{\mathbb{E}[\gamma O_i^W] = \gamma \pi^3 s_i = \mathbb{E}[\gamma O_i^V].}$$

Equal mean pressure does not fix acquisition

Let carriers X_j be independent, each infected with probability π. Here $\mathbb{E}$ means expectation. For any four-person set $Q \ni i$,

$$\mathbb{E}\left[\prod_{j\in Q\setminus\{i\}} X_j\right] = \pi^3.$$

Since $\sum_{Q\ni i} W_{i,Q} = \sum_{Q\ni i} V_{i,Q} = s_i$,

$$\boxed{\mathbb{E}[\gamma O_i^W] = \gamma\pi^3 s_i = \mathbb{E}[\gamma O_i^V].}$$

The same holds for eligible (currently susceptible) targets:

$$\mathbb{E}[(1-X_i)\gamma O_i^W] = \gamma(1-\pi)\pi^3 s_i = \mathbb{E}[(1-X_i)\gamma O_i^V].$$

But the mean pressure does not determine $\mathbb{E}[1-e^{-\lambda_i}]$:

$$f(\lambda) = 1-e^{-\lambda}, \qquad f''(\lambda) = -e^{-\lambda} < 0.$$

Manuscript, Section VI B. One pathogen, independent equal-probability carriers. Independence need not persist under evolution; equal means alone imply no ordering of acquisition.

An exact one-step result

Put $b = \beta_1$, $h = \beta_h$, $g = \gamma$, $\Lambda_1 = 3b + 2h$, and $\Lambda_2 = 2b + h$.

Proposition VI.2

On triangles $012, 013, 024$ and their pair-edge union, use one pathogen, $r = 2$, and identical recovery probabilities. Assume the initial law is unchanged by relabeling. Let K_0 count initial carriers and $q_3 = \mathbb{P}(K_0 = 3)$. Write ρ for the single-pathogen infected fraction and $\mathbb{E}$ for expectation. Then

$$\boxed{\mathbb{E}[\rho_{\text{matched}}(1)] - \mathbb{E}[\rho_{\text{original}}(1)] = \frac{q_3 F}{25},}$$

$$F = (1 - e^{-g/2})\big(e^{-\Lambda_2} - e^{-\Lambda_1 - g/2}\big) > 0 \quad (g > 0,\ b, h \geq 0).$$

Proof idea: only three-carrier states contribute. Two targets each contribute F across ten equally likely configurations:

$$\Lambda \qquad 2F$$

Backup: interpreting the simulation parameters

Setting	Meaning and documented rationale
800 vertices, 1,600 triangles	Synthetic population; triangle density avoids fragmentation seen with 200 triangles.
4,800 pairwise contacts	Same edge count across reuse settings; mean degree $2(4800)/800 = 12$.
$\beta_1 = .005$, $\beta_h = .1$	Pressure coefficients, not probabilities. A pressure .1 alone gives probability $1 - e^{-.1} \approx 9.52\%$.
$\mu = .1$	A carried pathogen recovers with probability 10% each step.
2% seeds; 200 steps	16 initial carriers per pathogen, disjoint; a finite, dimensionless observation horizon.
10 graphs; 30 trials each	Replication across contact structures and stochastic trajectories.

Same pairwise information, different transitions

On vertices $\{1, 2, 3, 4, 5, 6\}$, let the two triangle collections be

$$\mathcal{T}_A = \{123, 145, 246, 356\}, \qquad \mathcal{T}_B = \{124, 135, 236, 456\}.$$

Here 123 means the group $\{1, 2, 3\}$. Each pair belongs to the same number of triangles in both collections: once, except 16, 25, 34 (zero).

Same pairwise information, different transitions

On vertices $\{1,2,3,4,5,6\}$, let the two triangle collections be

$$\mathcal{T}_A = \{123, 145, 246, 356\}, \qquad \mathcal{T}_B = \{124, 135, 236, 456\}.$$

Here 123 means the group $\{1,2,3\}$. Each pair belongs to the same number of triangles in both collections: once, except 16, 25, 34 (zero).

One-pathogen counterexample

Initially infected: $\{1,2,3\}$. Set $r = 2$, $\beta_1 = \gamma = 0$, $\beta_h > 0$.

For each susceptible vertex $4,5,6$	$H_A = (V, \mathcal{T}_A)$	$H_B = (V, \mathcal{T}_B)$
Active triangles	0	1
Acquisition probability	0	$1 - e^{-\beta_h}$

Same **fixed labeled initial state**. The two hypergraphs are isomorphic; this does not compare relabeling-invariant averages.

Manuscript, Section IV B. Weighted pairwise information counts triangle memberships. This example isolates triangle

The transmission rule: three contributions

Why triples? Three-person groups are the smallest groups beyond pairs.
r **= activation threshold:** here $r = 2$, so *both other members* must carry the same pathogen for a triangle to be active.

$$\boxed{\lambda_i = \beta_1 N_i + \beta_h R_i + \gamma O_i.}$$

Count	What is counted around target i	Pressure per contribution
N_i	Infected pairwise neighbors	$\beta_1 \geq 0$
R_i	Active triangles	$\beta_h \geq 0$
O_i	Pairs of active triangles sharing another source	$\gamma \geq 0$

λ_i is total pressure during one step, for one pathogen. Triangle sides are also pairwise contacts; repeated sides count once in N_i.

Manuscript, Section III E. Distinct triangles. The additive group and overlap contributions are modeling assumptions.