PLOS COMPLEX SYSTEMS

RESEARCH ARTICLE

A spatial hypergraph model to smoothly interpolate between pairwise graphs and hypergraphs to study higher-order structures

Omar Eldaghar[1]*, **Yu Zhu**[2], **David Gleich**[2]

1 Department of Mathematics, Purdue University, West Lafayette, Indiana, United States of America,
2 Department of Computer Science, Purdue University, West Lafayette, Indiana, United States of America

* oeldagha@purdue.edu

OPEN ACCESS

Citation: Eldaghar O, Zhu Y, Gleich D (2025) A spatial hypergraph model to smoothly interpolate between pairwise graphs and hypergraphs to study higher-order structures. PLOS Complex Syst 2(9): e0000066. https://doi.org/10.1371/journal.pcsy.0000066

Editor: Akrati Saxena, Leiden University: Universiteit Leiden, NETHERLANDS, KINGDOM OF THE

Received: February 28, 2025

Accepted: August 1, 2025

Published: September 12, 2025

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: https://doi.org/10.1371/journal.pcsy.0000066

Data availability statement: All code to generate data, plots, and run experiments can

Abstract

We introduce a spatial graph and hypergraph model that smoothly interpolates between a graph with purely pairwise edges and a graph where all connections are in large hyperedges. The key component is a spatial clustering resolution parameter that varies between assigning all the vertices in a spatial region to individual clusters, resulting in the pairwise case, to assigning all the vertices in a spatial region to a single cluster, which results in the large hyperedge case. A key component of this model is that the spatial structure is invariant to the choice of hyperedges. Consequently, this model enables us to study clustering coefficients, graph diffusion, and epidemic spread and how their behavior changes as a function of the higher-order structure in the network with a fixed spatial substrate. We hope that our model will find future uses to distill or explain other behaviors in higher-order networks.

Author summary

Higher-order structure in networks encompasses group-level interactions beyond simple pairwise links. These group structures can profoundly shape dynamics like epidemics and synchronization, often in counterintuitive ways. Studying these effects is challenging because even basic measures like the clustering coefficient have multiple, non-equivalent higher-order generalizations. We introduce a flexible hypergraph model that smoothly interpolates between purely pairwise and higher-order interactions while preserving network connectivity. The model incorporates geometric or feature-based node information from sources such as spatial data or embeddings, enabling realistic network constructions. We demonstrate its utility through case studies on clustering, higher-order PageRank diffusions, and epidemic spreading. Our model provides a simple

be obtained from our public github repo at https://github.com/oeldaghar/spatial-hypergraph-epidemics. The data is too large to fit in this repo and is instead stored in Zenodo with a link to this stored in the github repo.

Funding: Eldaghar, Zhu, and Gleich all acknowledge funding and support from DOE award DE- SC0023162. Gleich is also partially supported by NSF IIS-2007481 and IARPA AGILE.

Competing interests: The authors have declared that no competing interests exist.

and flexible method to better delineate the distinct roles of pairwise and higher-order structures in complex networks.

Introduction

Parametric graph and hypergraph models have played an important role in network science and complex systems research—from the Watts and Strogatz model of local clustering and graph diameter [1], to Kleinberg's study of small world routing [2], to the famous mixing parameter in the Lancichinetti, Fortunato, and Radicchi (LFR) stochastic block model [3]. More generally, parametric variation in complex systems helps to identify different regimes of behavior and often phase transitions among them. Two examples of this include connectivity in a simple uniform random graph (Erdős–Rényi–Gilbert model) [4–7] and synchronization in the Kuramoto oscillation [8,9]. The key feature of these models is that they allow us to study a system as only one aspect varies.

While these parametric models have yielded foundational insights on pairwise data, many complex systems exhibit interactions beyond simple pairwise relationships. When these are present, they are often critical for identifying important structure in the networks [10]. Beyond simple structure, higher-order dynamics can introduce complex effects such as bistability or hysteresis in epidemic spreading [11,12]. In addition, dynamics such as synchronization can be sensitive to the choice of data representation in the higher-order setting [13]. Even classical pairwise notions such as homophily can break down in the higher-order setting [14]. For more about the state of higher-order studies, see the surveys [15] and [12]. An important aspect of higher-order networks is that ideas and concepts that are well-defined and unique in the pairwise setting often admit multiple, distinct, and different generalizations when extended to higher-order frameworks. This complexity motivates the development of parametric models that interpolate between pairwise and higher-order interactions, enabling controlled exploration of these richer dynamics and settings.

In this paper, we introduce a spatial model of graphs and hypergraphs that enables parametric variation from a purely pairwise edge behavior to the case where all nodes are involved in only hyperedges. Crucially, this can be done while retaining the same spatial graph substrate. This enables us to study changes due to the hyperedge structure alone, which represents a unique capability among hypergraph models.

The proposed model is simple and scalable. We randomly assign points to each vertex and also assign a number of neighbors, which is usually sampled from a random distribution. In the pairwise case, they directly connect to this number of nearest neighbors. However, in the general case, we run a clustering algorithm on the spatial connections among the points within this spatial region. The idea is that if points represent some latent similarity space, then nearby points will reflect similar features. Consequently, we use the spatial clusters in these regions to induce hyperedges. By varying the spatial cohesion, we can adjust the presence of pairwise edges compared with hyperedges. We discuss the model formally in Sect 1. One challenge was defining the spatial cohesion parameter to preserve scaling as we vary the dimension of the space from which points are drawn. This resulted in developing a key parameter α that smoothly interpolates from the pairwise to hypergraph case for multiple dimensions and number of neighbors assigned. A different feature of this model is that much of the connectivity among nodes is invariant to the choice of edges versus hyperedges, a result we formalize in Theorem 1.

We then explore how this model enables us to study the impact of higher-order structures in hypergraphs. Our first study is on clustering coefficients (Sect 3). There are many different types of clustering coefficients in hypergraphs. We study a few of the simplest and most common including both unweighted and weighted clustering coefficients of the clique expansion as well as bipartite clustering coefficients of the node-hyperedge incidence matrix. In what is a small surprise, scaling from medium-sized to large-sized hyperedges reduces the global clustering coefficient of the projected graph. This occurs because adding new projected hyperedges can cause the number of length-2 paths, or wedges, to grow much faster than one might expect. This study also shows how the behavior of different clustering coefficients varies quite substantially as we interpolate from pairwise graphs to hypergraphs, which suggests that results about clustering coefficients in hypergraphs may not be robust to changing the type of clustering coefficient used.

The next study is in terms of diffusion (Sect 4). We use a seeded or personalized PageRank diffusion in hypergraphs [17]. In this case, the behavior of the diffusion is governed by the spatial substrate underlying the network. Consequently, we see little difference in the behavior of the diffusion as we move from the pairwise case to the hypergraph case. The choice to include this study is meant to show that the model behaves as expected when higher-order structure may not impact the underlying physics.

The final study is on epidemic spread (Sect 5). In this case, there is tremendous uncertainty about the impact of higher-order structure and, indeed, a variety of mixed results in the literature. For example, the addition of higher-order structures and group-level spreading can greatly alter the stability of epidemic thresholds by inducing a region of bistability [11] not seen in traditional pairwise models. Moreover, heterogeneity can also play competing roles in pairwise and higher-order structures [18] to accelerate or inhibit spreading. In this case, we wish to study epidemic spread in a model that attempts to mimic an airborne virus in the presence of ventilation. In this scenario, large hyperedge interactions require spaces with additional ventilation, which corresponds to a dilution effect of infectious aerosols. Of note, we find that the impact of higher-order structure varies with the epidemic parameters in non-intuitive ways in this scenario (Sect 5.5).

This paper extends a previous introduction of these ideas from the same authors [16]. Key differences in this greatly expanded version include (i) a discussion of the model beyond two-dimensional spatial graphs, (ii) a theoretical characterization of the connectivity of the model, and (iii) studies of the hypergraph model in terms of clustering coefficients as well as (iv) graph diffusion. Finally, the epidemic study includes a more detailed analysis of the specific hypergraph mechanisms underlying the differences observed in [16]. We discuss additional related work in the space of random geometric graphs, random geometric hypergraphs, and random geometric simplicial complexes in Sect 2. The ability to easily interpolate between a pairwise graph and a hypergraph appears to be a unique feature of this model.

1. Model description

The model we propose is simple, fast, and flexible. It begins with a set of points in a space $\boldsymbol{X}$ along with a distance metric D. In all of our studies, these points are sampled from the d-dimensional unit-cube $[0,1]^d$ uniformly at random, although in principle a different space or distribution can be used. For each point v, we give it a radius of influence, d_v, expressed as a number of nearest neighbors, which we call the *degree*. Note that this is a mild abuse of notation, because in the final graph construction, the degree of the node is typically larger than d_v, although d_v is a lower bound on the degree. We typically sample values of d_v from a log-normal distribution. A summary of our notation can be found in Table 1.

Table 1. A summary of our notation.

Symbol	Interpretation
$\boldsymbol{X}$	the set of all points
$D(\cdot,\cdot)$	the distance metric
$\boldsymbol{X}_v$	the coordinate of a specific point v
d_v	the degree associated with point v
$\boldsymbol{d}$	the set of all degrees for all vertices
$N(v)$	the set of d_v nearest neighbors of v under distance D

https://doi.org/10.1371/journal.pcsy.0000066.t001

Let $\boldsymbol{X}_v$ be the coordinates of a point v and d_v be the associated degree. In a standard spatial nearest-neighbor graph, we would connect node v to the nearest d_v neighbors by adding edges for each neighbor. Our model is based on this setup. However, we wish to *cluster* the points within v's radius of influence. Let $N(v)$ be the set of d_v nearest neighbors for point v, and let r_v be the distance to the d_vth nearest neighbor, formally, $r_v = \max_{u \in N(v)} D(v,u)$. The goal is to cluster the set of points in $N(v)$ and form edges or hyperedges based on these *clusters* rather than individual points. (To be completely clear, we do not consider v in the set of points we cluster.) We illustrate this process in Fig 1. Formally, for each cluster of points in $N(v)$ we create a hyperedge consisting of all points in the cluster combined with the original point v. The inspiration for this idea is that we would have a *group interaction* among v and points that are all themselves *close* within its region of influence.

The key feature of the model is that we can control the behavior of the graph by controlling the behavior of the clustering function. Suppose that each point in $N(v)$ is clustered into a *separate cluster*. Then we simply recover the pairwise nearest neighbor graph among the points. Alternatively, suppose that $N(v)$ is clustered into a *single cluster*. Then we recover a geometric hypergraph construction where all edges are hyperedges (unless $d_v = 1$). Thus, by varying the cluster sizes, we can control the extent of hyperedge effects. This idea is illustrated in Fig 2.

The model could accommodate any clustering function we desire and, in addition, support attributes on nodes as well. However, for concreteness, we use the DBSCAN clustering algorithm [19]. The DBSCAN method depends on two parameters to control the clustering: ε and min_pts. The choice of ε is a distance and governs when two points are considered in the neighborhood of each other, and, in turn, whether or not they might be placed in the same cluster. While we can change ε to achieve the interpolation from pairwise to hypergraph we

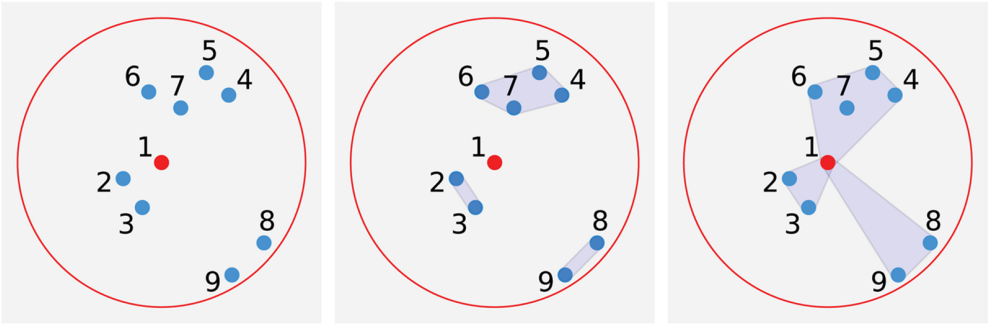

Fig 1. How hyperedges are formed in our model. Eight nearest neighbors are computed for node 1 (leftmost plot) which are then clustered into 3 clusters (middle plot). Finally, each of the clusters serves as a separate group interaction for node 1 and they become hyperedges.

https://doi.org/10.1371/journal.pcsy.0000066.g001

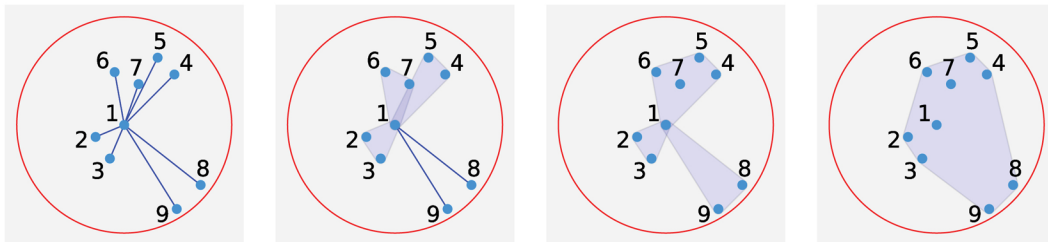

Fig 2. The extent of clustering controls interpolation. As we vary the number of clusters produced among each node's nearest neighbors, we are able to interpolate between purely pairwise (leftmost plot) and purely higher-order structure (rightmost plot).

https://doi.org/10.1371/journal.pcsy.0000066.g002

want, the effect depends greatly on the local context of each node. Meanwhile, min_pts serves as a density check for labeling points as noise. We fix min_pts = 1 here and ignore the distinction. Consequently, we wish to develop a simple and interpretable parameter to control the interpolation.

We introduce a parameter α for this goal. When we set $\alpha = 0$, we want a pairwise graph – corresponding to the scenario at the left of Fig 2. When we set $\alpha = 2$, we want a graph where all edges are complete hyperedges within the region of influence – corresponding to the scenario at the right of Fig 2. For $\alpha = 1$, we'd like a point in the middle where there are around $\sqrt{d_v}$ clusters. Crucially, we want to *smoothly* interpolate among these conditions. Consequently, we want to build a function

$$\varepsilon_\alpha(r_v, d_v, d)$$

that depends on the number of neighbors d_v, the distance enclosed by the region of influence r_v, along with the ambient dimension d and produces a value of ε for DBSCAN to achieve this goal. Fig 3 illustrates the reason why the function needs to scale with both d_v and r_v (and implicitly, the ambient dimension d).

We map α to distance parameter ε in DBSCAN via

$$\varepsilon_\alpha(r_v, d_v, d) = \begin{cases} \alpha^{1/d} \frac{r_v}{d_v^{1/2d}}, & \alpha \in [0,1] \\ \frac{r_v}{d_v^{1/2d}} + (\alpha - 1)\left(r_v - \frac{r_v}{d_v^{1/2d}}\right), & \alpha \in (1,2]. \end{cases} \tag{1}$$

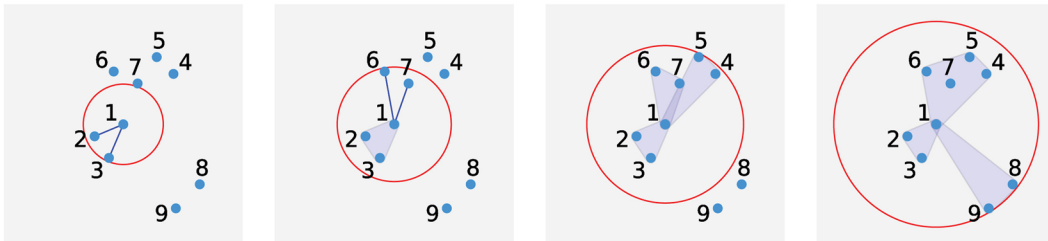

Fig 3. Example showing local hyperedge formation. Hyperedges formed around the node $v = 1$ as the number of nearest neighbors, d_v, increases. We want the neighborhood radius parameter ε of DBSCAN to scale with d_v and r_v (the maximum distance among the d_v neighbors). To establish a concrete and controllable model, we design a function ε_α that interpolates between individual clusters around each point and a single cluster. This function needs to scale with the parameters of the neighborhood to achieve its aims and this example indicates our scaling.

https://doi.org/10.1371/journal.pcsy.0000066.g003

Recall that we want about $\sqrt{d_v}$ clusters when $\alpha = 1$. So we want to scale our value of ε_1 to do so. The intuition for our choice of ε_1 is that if $B^d_{r_v}$ denotes a d–dimensional ball with radius r_v, then if we split the volume of this ball, $Vol(B^d_{r_v})$, into $\sqrt{d_v}$ equal pieces (ignoring the issue of sphere packing), it would yield pieces of volume $\sqrt{d_v}\varepsilon_1^d \approx r_v^d$ so that $\varepsilon_1 \approx \frac{r_v}{d_v^{1/2d}}$. The choice of ε_α in Eq (1) then gives an estimate of linear scaling for $\alpha \in [0,1]$ to account for changes in the dimension d. For $\alpha \in (1,2]$, we interpolate from a radius of ε_1 to r_v which corresponds to moving from approximately $\sqrt{d_v}$ hyperedges to a single hyperedge.

Our complete spatial hypergraph construction procedure is described in Algorithm 1. This shows how we build up a list of hypergraph edges by considering the results of the clustering. We also have our computational codes for both our graph model and applications available at https://github.com/oeldaghar/spatial-hypergraph-epidemics that implement this routine. To give an intuition for the resulting graph, a few small samples are shown in Fig 4.

Algorithm 1: Spatial hypergraph model.

```
function SpatialHypergraph(X, d, f(·,α)) where d gives the degree for
each node, and f(·,α) is a clustering algorithm with parameter(s)
α such that α = 0 will cluster into individual pieces, α = 2 will
cluster into a single group, and α = 1 will cluster into about
√points groups
   H ← [ ]                          ▷ Initialize an empty list of hyperedges
   for v = 1 : n do                 ▷ for each point in the the set X
      N(v) ← d_v nearest neighbors of X_v excluding X_v
      Y ← {X_u for u ∈ N(v)}        ▷ Build a subset of points to cluster
      C_v ← f(Y, α)                 ▷ run the clustering with parameter α
      for C in C_v do               ▷ for each cluster in the output
         APPEND(C, v)  ▷ add vertex v to the cluster before we add it
as a hyperedge
         APPEND(H, C)               ▷ add a new hyperedge to the graph
   return H
end function
```

1.1. The choice of clustering distance

This choice of clustering distance ε_α differs from that in our previous paper [16]. In particular, the scaling in [16] did not scale with dimension and had a different midpoint ε_1 value. We illustrate the importance of setting this parameter correctly in S1 Fig.

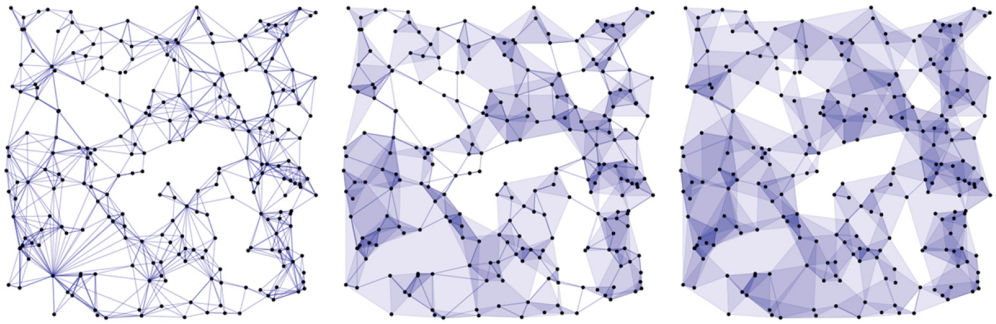

Fig 4. Visualization of generated hypergraphs from our model. Hypergraphs generated on $n = 250$ nodes in $d = 2$ dimensions for $\alpha = 0, 1, 2$ (left to right) for the same spatial embedding $\boldsymbol{X}$ and specified degrees $\boldsymbol{d}$.

https://doi.org/10.1371/journal.pcsy.0000066.g004

We generate data from our model where $n = 10000$ points are sampled from $[0,1]^d$, the degree distribution is sampled from a log-normal with parameters $\mu = \log(3)$ and $\sigma = 1$ (which gives an overall average degree of around 3.5). We fix the geometric information $\boldsymbol{X}$ and node degrees d_v for 25 simulations as we vary the parameter α. Put another way, we do not regenerate the spatial information for each distinct value of α, and reuse the same information for an entire sweep through the choices of α.

Our choice of ε_α allows for a smooth interpolation in the total number of hyperedges as shown in Fig 5. As we increase the radius in the DBSCAN algorithm, we expect the total number of hyperedges to decrease. Ideally, we'd like this decrease to be smooth. The *jump* near $\alpha = 2$ occurs because of a discontinuity in the behavior of the clustering algorithm as we go from 2 clusters to 1 cluster at each vertex, which cannot be smooth. A comparison with alternative choices is depicted in S1 Fig and discussed in S1 Text.

1.2. Invariance of connected components

We next show that the connectivity of the overall graph or hypergraph is invariant to the choice of α and depends only on the set of points $\boldsymbol{X}$ and the assigned degrees $\boldsymbol{d}$. This is intuitively straightforward because the edges formed only depend on the points and the specific number of neighbors chosen by d_v for each vertex, but the following argument makes this intuition rigorous.

Theorem 1. *α-Invariance of Connected Components*
Let $G_\alpha = H(\boldsymbol{X}, \boldsymbol{d}, \alpha)$ denote the spatial graph generation model outlined above with parameters $\boldsymbol{X}, \boldsymbol{d}, \alpha$. For fixed values of $\boldsymbol{X}, \boldsymbol{d}$, let $C_{\boldsymbol{X},\boldsymbol{d}}(\alpha)$ denote the connected components in G_α. Then $C_{\boldsymbol{X},\boldsymbol{d}}(\alpha)$ is independent of α.

Proof: Let $\boldsymbol{X}, \boldsymbol{d}$ be fixed and consider $\alpha > 0$. It suffices to study the connectivity of the graph formed by replacing each hyperedge with a spanning star. Replacing all hyperedges $h \in E(G_\alpha)$ with the spanning star centered on the generating vertex for that hyperedge, yields the same graph for all values of $\alpha \in [0, 2]$. Since replacing hyperedges with spanning stars does not alter connected components, $C_{\boldsymbol{X},\boldsymbol{d}}(\alpha)$ is independent of α. □

A key impact of this result is that it allows us to inherit the usual results about connectedness such as critical thresholds and giant components when $\boldsymbol{X}$ and $\boldsymbol{d}$ are constructed to match such statements. For instance, prior work [20] has derived a critical connectivity threshold for

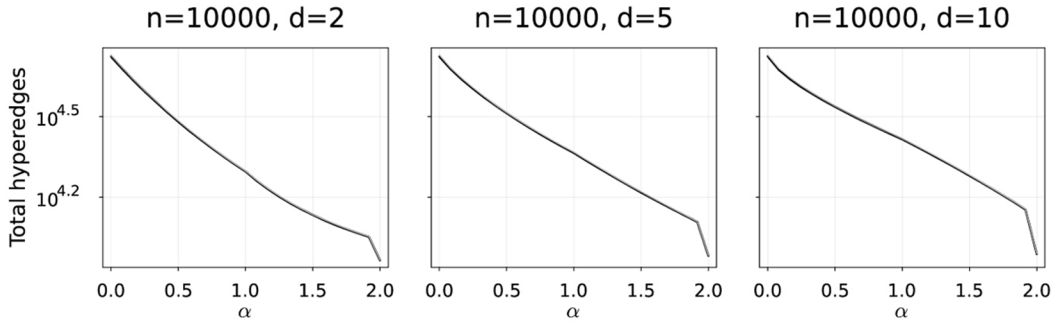

Fig 5. The total number of hyperedges formed is smooth. Total number of hyperedges formed using Eq (1) as we vary the dimension of the space $\boldsymbol{X}$ for $d = 2, 5, 10$. This illustrates a smooth interpolation between pairwise effects and pure hyperedge effects as α varies from 0 to 2.

https://doi.org/10.1371/journal.pcsy.0000066.g005

the pairwise random geometric graph model. Because the connectivity of the model is equivalent to the underlying pairwise graph due to Theorem 1, these same thresholds apply to our model as well.

1.3. Graph statistics

We continue by empirically studying simple graph statistics. We study what happens as we vary the dimension d and value of α in Fig 6. We report the average number of hyperedges of a given size, the total number of hyperedges, and the total number of triangles in the pairwise projected graph as a function of α. Recall that the pairwise projected graph, or *clique expansion*, results in a clique to represent each hyperedge of the original graph. We use the same experimental setup as in Fig 5 and Sect 1.1.

In Fig 6, as we increase α and consequentially ε_α, the total number of hyperedges decreases while the total number of triangles in the pairwise projection increases monotonically. This result is expected because the projected graph has more cliques, which adds more triangles. The key point of this figure is that we get larger hyperedges with smaller values of α as the dimension d increases. In terms of the impacts on triangles, this results in a steeper initial increase in triangles, although overall fewer triangles as the hyperedges get larger.

2. Related work

As mentioned in the introduction, there are a variety of similar geometric graph or spatial hypergraph models, although none of them enable the same type of seamless mapping from pairwise to higher-order structures that we achieve. In this context, our model extends both pairwise geometric random graph models as well as geometric hypergraph models. We briefly review these constructions and give pointers for more information. We have a longer survey on random hypergraph models in preparation [21].

Direct inspiration. Our proposed model draws inspiration directly from the geometric protean model [22] and a simplified extension [23]. These describe a similar latent space model for graphs where nodes have fixed degrees.

Geometric pairwise models. Pairwise random geometric models are created by pairing geometric information $\boldsymbol{X}_v$ for each node v with a distance function D and some rule for connecting nodes that depends on D. A common construction is just to connect points below a fixed distance. Given two nodes u,v and their coordinates $\boldsymbol{X}_u, \boldsymbol{X}_v$, an edge is added if $d(\boldsymbol{X}_u, \boldsymbol{X}_v) \leq c$ for some threshold c. The threshold c may vary with nodes as well. Another common variant is k-nearest neighbor (kNN) graphs where each node connects to its k closest nodes. This is akin to using different radii, r_v, for each node. Yet another is to relax to a soft geometric model using a kernel function f by connecting nodes with probability $f(d(\boldsymbol{X}_u, \boldsymbol{X}_v))$ that typically decays as the distance increases.

See the survey [24] for an overview of spatial graph models. Spatial models are specific instances of latent space models where edge connection depends on latent node features. A key focus of research on spatial graph models involves connectivity thresholds [20]. This has important implications for routing in *ad-hoc* networks of agents where the connection radius is implied by a radio transmission.

Geometric hypergraph models. There is a completely different notion of a geometric hypergraph described in the survey [25]. This alternative model directly creates the bipartite incidence matrix of the hypergraph. The idea is to create two different types of points among

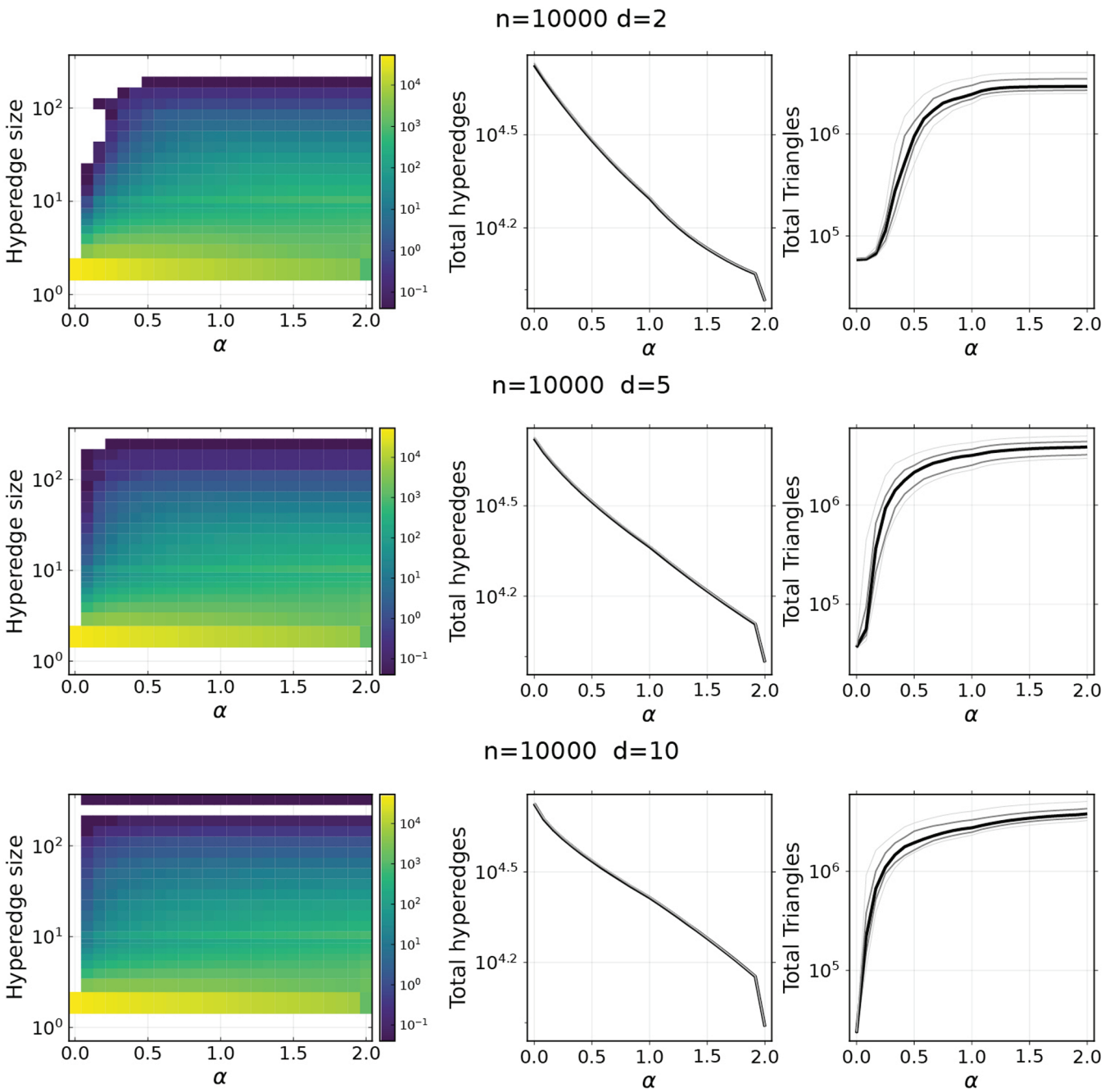

Fig 6. Simple hypergraph statistics as we vary parameters. Simple graph statistics as we vary α (Eq (1)) and the dimension d. The α denotes the interpolation parameter while we show (leftmost column) the average number of hyperedges of a given size, (middle column) the total number of hyperedge present (which is repeated from Fig 5), and (rightmost column) the total number of triangles in the projected graph for $d = 2, 5, 10$. Grey to black to gray bands in the middle and right columns indicate 10th, 25th, 50th, 75th, and 90th percentiles. In the heatmaps, we show the average number of edges in each bin over the 25 trials. The entries are log-scaled, so 4 corresponds to 10^4 edges.

https://doi.org/10.1371/journal.pcsy.0000066.g006

the samples of $\boldsymbol{X}$. One type of point represents nodes and the other type represents hyperedges. Then we directly build the incidence matrix of the hypergraph by using any of the spatial connection methods for a geometric graph. Connectivity properties of a model similar to that in [25] were analyzed asymptotically [26].

Another class of methods is based on random simplices [27–29]. In such methods, topological tools such as the Čech complex or the Vietoris–Rips complex are used to form hyperedges based on spatial information. Another method [30] makes use of the same topological tools but places priors on point configurations in order to induce desired structures in the generated simplices. While of note, these methods are more restrictive than those of hypergraphs. A relaxation from simplices to a geometric hypergraph model of varying sizes was made using latent space modeling and sampling [31]. In particular, the latent space model uses a shared sequence of radii for all nodes $r_2 < \cdots < r_K$. A hyperedge of size k is placed among vertices whenever the respective balls with radius r_k intersect.

Our proposed model is distinct from both of these ideas. We allow the hyperedges to vary based on a clustering algorithm instead of a random point selection. Also, we directly generate hypergraphs instead of going through simplicial complexes.

3. Clustering coefficients on hypergraphs

There are two common clustering coefficients for a pairwise graph. The global clustering coefficient of a graph G is defined as

$$C(G) = \frac{6 \times T}{P_2},$$

where T denotes the total number of triangles and P_2 denotes the total number of length-two paths. These length-two paths are often called *wedges*. The average local clustering coefficient for a graph is given by

$$\bar{C}(G) = \frac{1}{|V|} \sum_{v \in V} C_v = \frac{1}{|V|} \sum_{v \in V} \frac{T_v}{\binom{k_v}{2}},$$

where V is the vertex set, T_v is the number of triangles that contain the node v, and k_v is the number of neighbors of node v.

As mentioned in the introduction, when a pairwise concept is generalized to a higher-order structure, it often has multiple generalizations. This is the case for clustering coefficients, and a number of different generalizations of clustering coefficients for hypergraphs have been proposed (see [32,33]). We will compare how a few of these behave on our model as we vary α and the spatial dimension d.

Let $\mathcal{H}$ denote a hypergraph with a set of vertices V and a set of hyperedges H. Perhaps the simplest such generalization is to project the hypergraph onto a graph via *clique expansion* and then compute the pairwise clustering coefficient for the projected graph. In clique expansion, each hyperedge is replaced by a clique over all of the nodes within the hyperedge. We then arrive at two distinct clustering coefficients in the hypergraph based on the global and local clustering coefficients in the projected graph.

The projected graph is unweighted. However, it is built with a union of cliques. We can also consider the projected multigraph. This is a multigraph interpretation of the weighted projected graph from `networkx` for example [34]. In this case, we allow the graph to have multiple edges as we take the union of all the projected cliques. (A weighted version of this same multigraph will arise in our study of epidemics as well.) We use this multigraph to define a weighted local clustering coefficient by counting the number of triangles – including repetitions due to multiedges – divided by the number of wedges centered at a node v. In this case, it is possible for a node to have *more triangles* than wedges. A scenario where this occurs is if the edge that closes the triangle is repeated whereas the edges defining the wedge are not.

For this reason, we clip the maximum value of the local weighted clustering coefficient at 1. The overall value is then

$$\bar{C}_M(G) = \frac{1}{|V|}\sum_v \min\left(\frac{T_v^M}{\binom{k_v^M}{2}}, 1\right),$$

here T_v^M is the number of triangles and k_v^M is the number of edge end-points that start at node v – both including multiplicities due to multiedges. We note that there are other notions of a weighted clustering coefficient as well [35]. While we are not aware of any place $\bar{C}_M(G)$ has been proposed, we suspect that it has been.

Another intuitive idea is that triangles are the shortest cycles without repeated edges. This leads to a different generalization of clustering coefficients to bipartite graphs [36]. We can apply this idea in the bipartite representation (star expansion) of a hypergraph. This clustering coefficient amounts to computing the quantity

$$CC_4 = \frac{4 \times C_4}{L_3},$$

where C_4 denotes the number of 4-cycles and L_3 denotes the number of 3-paths in a bipartite graph. The quantity CC_4 is knonw as the Robins-Alexander clustering coefficient [36]. It is also sometimes known as the metamorphosis coefficient [37].

In summary, we will study the four quantities:

C	global clustering coefficient in projected graph (clique expansion)
$\bar{C}$	average local clustering coefficient in projected graph (clique expansion)
$\bar{C}_M$	multigraph local
CC_4	the bipartite clustering coeff. in hypergraph incidence matrix (star expansion)

3.1. Experimental results

We study the clustering coefficients in the same experimental regime from Sects 1.1 and 1.3. The results of computing the four different clustering coefficients are shown in Fig 7. The bands indicate the minimum and maximum values across 25 trials, while both the embedding $\boldsymbol{X}$ and degrees $\boldsymbol{k}$ are fixed across values of α for each independent trial.

The first thing we note is that there is no regularity in the behavior as a function of α. Both the global and local clustering coefficients (C and $\bar{C}$) initially increase before decreasing for large α. We find this result puzzling, as we found that the total number of triangles in the projected graph grows with α in previous figures (Fig 6). We explain this finding in Sect 3.2 next.

Our next observation is the critical impact of the spatial dimension d on the results. All of the different clustering coefficients show changes in the behavior with regard to this parameter. For instance, $\bar{C}$ is much lower when $d = 10$ compared with $d = 2$. Consider also the results with $\bar{C}_M$ as a function of d. Here, we observe that $\bar{C}_M$ is *smaller* when $\alpha = 1$ and $d = 2$ whereas $\bar{C}_M$ is larger. Finally, for CC_4 we see an overall decrease in clustering as dimension increases.

In comparison with the three clustering coefficients on the pairwise graph, the behavior of the bipartite clustering coefficient CC_4 in the star expansion is more closely aligned with our expectations that as α increases one would expect clustering coefficients to increase.

Overall, these results suggest that clustering coefficients in hypergraphs are far more subtle and complex than in pairwise graphs. This supports the idea that there could be a multitude of reasonable generalizations depending on exactly which features are desirable to capture in

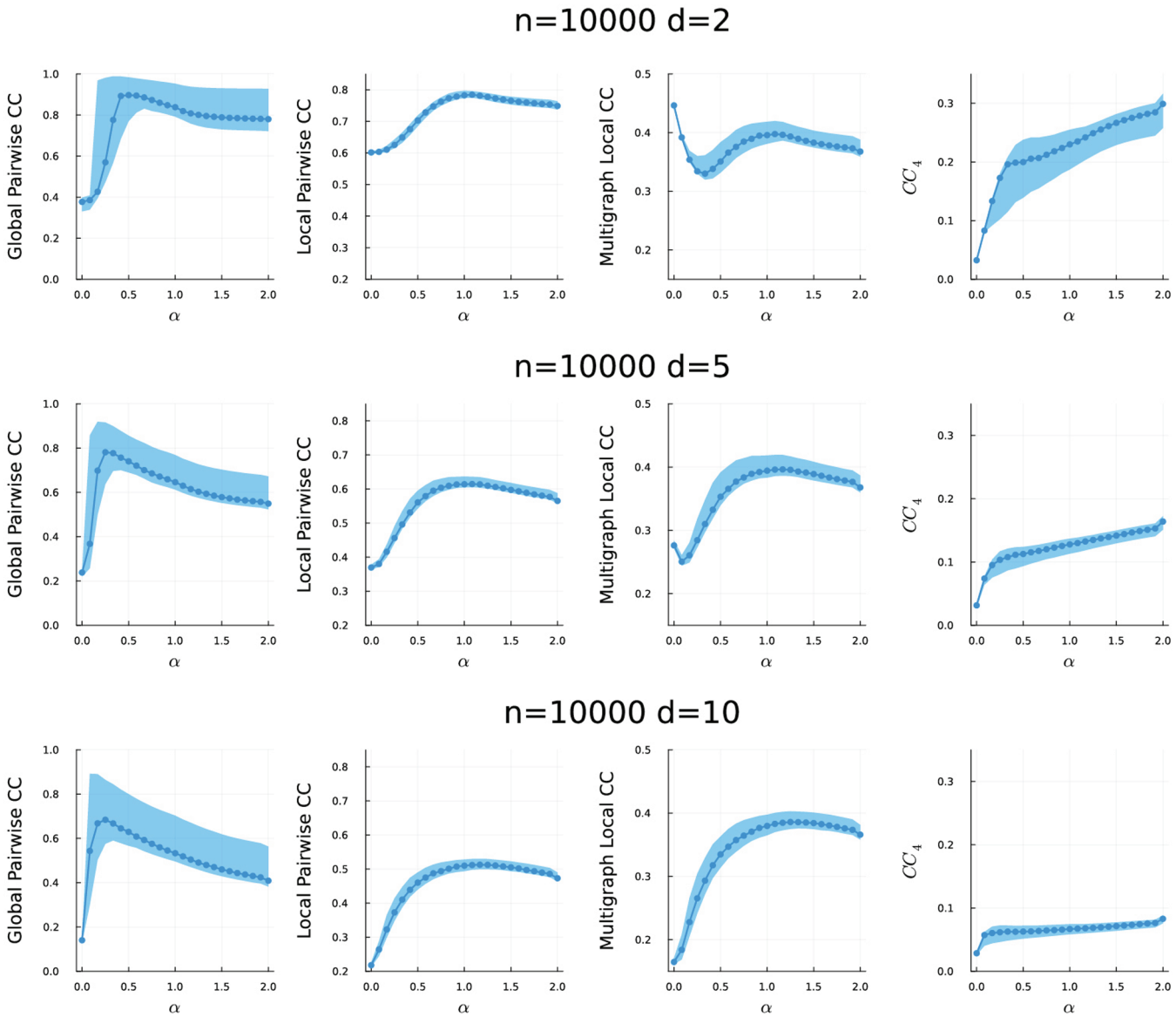

Fig 7. The evolution of various notions of clustering as we vary parameters. Various clustering coefficients (columns) as we vary ε_α (x-axes) and the dimension of the node embedding (rows). From left to right, the columns show: (1) pairwise global clustering coefficient, (2) pairwise local clustering coefficient, (3) a multigraph clustering coefficient, and (4) the Robins-Alexander clustering coefficient from the bipartite representation. Bands indicate the maximum and minimum values over 25 trials. Spatial information and degrees are shared across trials.

https://doi.org/10.1371/journal.pcsy.0000066.g007

the generalization. It also suggests that there probably is not going to be a universal clustering coefficient in all scenarios.

3.2. Why global and local projected clustering coefficients decrease

We now return to the scenario that initially left us puzzled. Both the local and global clustering coefficient of the pairwise graph C and $\bar{C}$ decrease for large values of α.

Fig 8 shows a detailed example of how the local clustering of a single node can decrease despite an increase in triangles. The left column (panels a, c, e) shows how the nodes u and

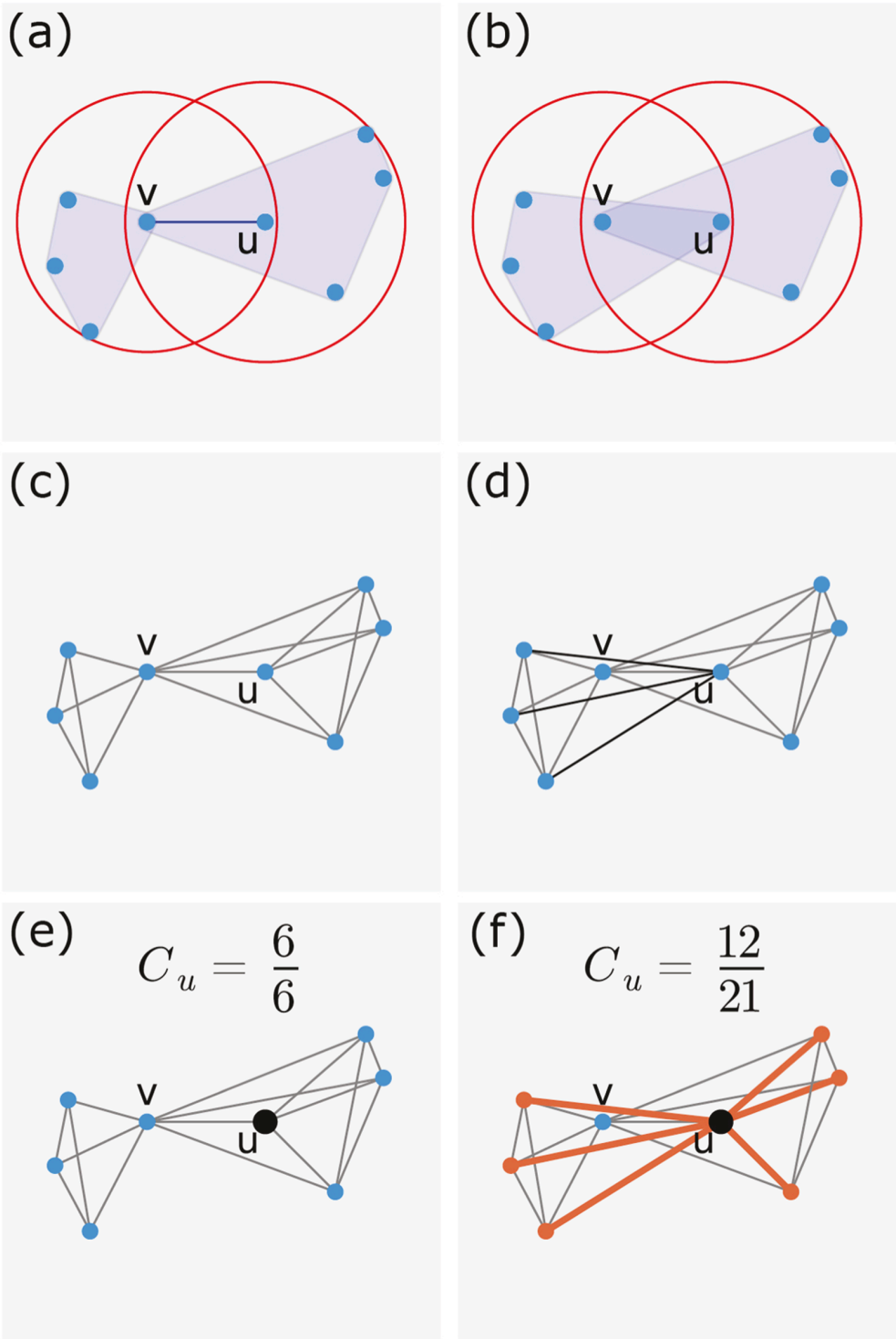

Fig 8. Example comparing the local clustering coefficient for two pairwise projections from instances of our model. The first row shows the hyperedges generated by our model. The second row shows the pairwise projections from row 1, highlighting the different edges in black in the right column. The third row highlights all edges that participate in a wedge centered about the node u. There are none for the left column but 9 wedges in the right column. C_u is displayed as a ratio of triangles to wedges centered on the node u.

https://doi.org/10.1371/journal.pcsy.0000066.g008

v are generating hyperedges for $\alpha = 1.9$ while the right column (panels b, d, f) shows this process for $\alpha = 2$ for our model. Note the that pairwise edge between v and u in panel (a) is formed when running DBSCAN on the node v as the node u is an outlier or boundary point.

The second row (panels c-d) shows the unweighted pairwise projections of each of these hypergraphs. Panel (d) has the new pairwise edges relative to panel (c) highlighted in black. These new pairwise neighbors add both triangles and wedges to the node u. In the last row (panels e and f), we show C_u and highlight edges in unclosed wedges centered on the node u. We get an unclosed wedge for any combination of left vertices and right vertices in panel (f). So there are $3 \times 3 = 9$ new wedges centered on the node u. Moreover, in this example the clustering coefficients $\bar{C}, C, CC_4$ also decrease. This shows how adding triangles with a clique expansion in the projected graph can, ironically, introduce even more wedges in the graph.

4. Diffusion models in spatial hypergraphs do not show higher-order effects

Next, we use our model to study diffusion. In comparison to clustering coefficients, we do not expect to see large changes in the behavior of diffusion for our graph model. This is because the same spatial substrate underlies the graph as we vary α to interpolate between pairwise and higher-order effects. Consequently, the physics of planar diffusion dominates the choice of higher-order or pairwise model. That said, we still see small differences between the two models in our study.

To perform the study, we generated a random set of $n = 500$ points in the [0,1] region and a fixed set of degrees for each node sampled from a log-normal distribution with mean $\log(3)$ and variance 1. (So the mean degree should be around 3.5). Then we created instances of the graph as we varied α. On each graph, we solved a seeded PageRank problem as an instance of a higher-order diffusion. We used the model from [17], although we'd expect similar results from other hypergraph generalizations of PageRank such as [38,39]. The algorithm we pick is a strongly local and sparsity promoting PageRank method that we choose because it might encourage slightly more differences among the models than global seeded PageRank solutions. The specific parameters we used were a 2-norm loss, $\kappa = 0.0001$ (this causes the solution to grow away from the seed), $\gamma = 1$ (this reflects the "width" around the seed), and $\rho = 0.5$ (an approximation tolerance). Then we pick a seed in a corner of the graph and zoom in on the region identified by the seeded PageRank vector. For more details on these parameters, see S2 Text.

The results when using $\alpha = 0, 1, 2$ are shown in Fig 9. These show that the behavior of the diffusion propagates radially away from the seed vector. There are some small differences in the propagation, especially around the boundary where the solution entries are small and the sparsity truncation changes which entries are truncated slightly. On the other hand, we see broad agreement in terms of element magnitude among the cases. We compare the values more closely in the scatterplot which shows that there are differences in the values, but the relative ordering is largely preserved with the pairwise case.

Consequently, and as expected, this model shows little difference on a diffusion computation.

5. Pairwise vs higher-order epidemics

In our final case study, we illustrate the utility of our spatial hypergraph model for understanding epidemics on hypergraphs.

While higher-order contagion is relatively new compared to pairwise epidemics, there are a number of significant differences between pairwise and higher-order spreading [11,18,40,41].

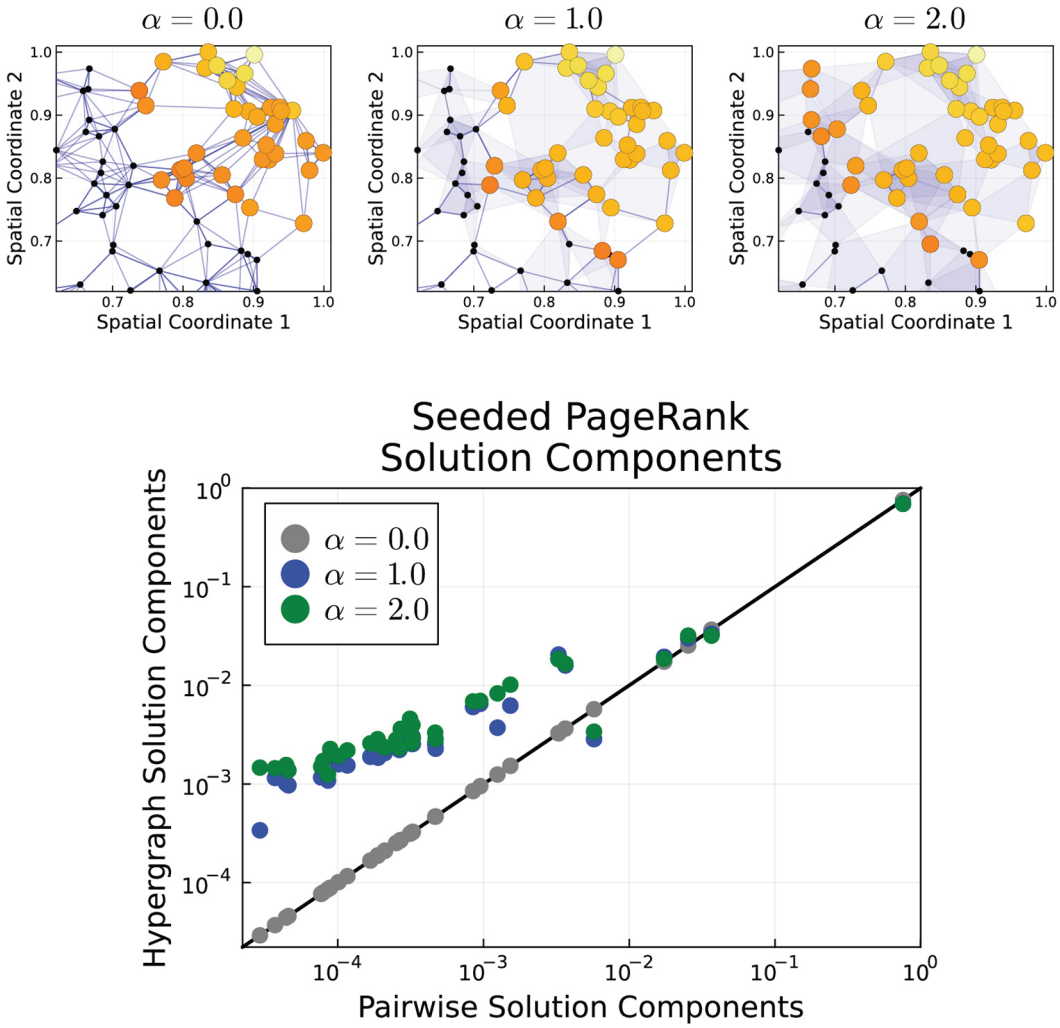

Fig 9. Minor differences in seeded and sparse PageRank solutions for our spatial hypergraph model. When using the same seed node, seeded and sparse PageRank solutions for our spatial hypergraph model only show minor differences as we vary the amount of higher-order structure. The first row shows the seeded PageRank solutions using the top-rightmost node for $\alpha = 0, 1, 2$. The colors are chosen based on a log scale and black nodes did not meet the sparsity criteria for inclusion in the solution. The second row shows a scatter plot of how the solutions for $\alpha = 0, 1, 2$ compare to the pairwise case ($\alpha = 0$) for the same seed node. This shows a scatter plot of the components of the seeded PageRank components where each node is a single dot. The black line would indicate exact agreement for all components of the seeded PageRank solution vector. Here, we see some small differences, but they largely reflect difference in the precise values computed rather than the relative ordering of behavior.

https://doi.org/10.1371/journal.pcsy.0000066.g009

In particular, there have been a number of conflicting results on the relevance of higher-order effects in spreading. On the one hand, epidemic spread is *inherently* a pairwise behavior in which a real or virtual pathogen spreads from one individual to another in an infection event. On the other hand, pathogens spread via airborne routes have obvious group-relevant interactions [42]. Theoretical and empirical studies on these findings have been mixed. As shown by [40], without strong hyperedge-dependent infection effects, hyperedge transmission models reduce to weighted pairwise transmission models. Studies of human mobility and SARS-CoV-2 showed that super-spreading and the associated group interactions were key routes of

transmission [43,44]. Additional theoretical studies show bistable [11] parameter regions in epidemics with simplicial complexes.

We use our model to study the impact of group-level or higher-order spreading in epidemics. Our vision is to model an airborne pathogen where group-level behavior may be important. We make several simplifying assumptions. A single infectious node in a group is enough to infect any other node within that group. This is in contrast to collective contagion in which some fraction of nodes must be infectious to enable group-level spreading. Moreover, we scale the probability of a node becoming infected with both the number of infected nodes within that group as well as the size of the group in different ways. This is because the more infectious contacts a node has within a group, the more infectious aerosols would be produced. However, a joint contact among a large group requires more space. Moreover, ventilation standards in the US [45] state air dilution rates that scale with the number of people. This dilution will be a key feature in our models.

5.1. The epidemic model

We used a discrete time Susceptible-Infected-Recovered-Susceptible (SIRS) compartment model with an additional exogenous infection term. At each time, t, a node can be in exactly one of three states: susceptible, infected, or recovered. Recovered nodes are temporarily immune from infection and they lose immunity with probability δ. Infected nodes transition to recovered with probability γ. Susceptible nodes become infected due to contacts with infectious individuals through an edge or hyperedge or an exogenous infection with probability η. We include this exogenous infection term because we model a small population embedded within a larger population that can drive infections through other means. In terms of the interaction with the hypergraph, we view each hyperedge as a separate interaction the node experiences during a time period. Consequently, each edge or hyperedge represents a possible infection route. Let $v \in V$ and let $h \in H$ be a hyperedge containing v that represents a group interaction. We set the infection probability for node v from hyperedge h at time t to be

$$\beta_h^{(t)} = \Pr(v \text{ infected by } h | v \text{ susceptible}) = \frac{1 - (1-\beta)^{i_h^{(t)}}}{g(|h|)}, \tag{2}$$

where β denotes a baseline pairwise infection probability, $i_h^{(t)}$ denotes the total number of infected nodes in the hyperedge h at time t, and g is function that represents the impact of ventilation that depends on the total number of nodes in h. The term $1 - (1-\beta)^{i_h^{(t)}}$ represents the pairwise infection probability within the hyperedge h for a susceptible node. We further assume independence among distinct hyperedges. So two nodes can interact among several groups and each of those interactions can independently transmit infection.

We make three simple choices for $g(m)$ to represent various potential ventilation scenarios. We use $g(m) = 1$, $g(m) = \sqrt{m}$ and $g(m) = m$ to model no ventilation, moderate ventilation, and high ventilation scenarios. The case of $g(m) = 1$ corresponds to no ventilation for group interactions and hence infections are transmitted in a pairwise fashion within each hyperedge. The case of $g(m) = m$ corresponds to a linear dilution due to improved airflow. The US ventilation standards [45] provide ventilation rates per person, which should provide increased infectious aerosol dilution in large groups. We also study a low-ventilation scenario where $g(m) = \sqrt{m}$, which may correspond to some intermediate regime such as a partially failed ventilation system, a room over capacity, or some type of transition space such as hallway or boarding platform.

5.2. Related work

Many methods related to epidemic spreading often make use of individual-level stochastic models or mean-field approximations of the continuous-time process [40,46,47]. While these approaches share similarities with discrete event simulations, there are some notable differences in the pairwise case regarding fine-scaled information and the impact of homogeneity assumptions on total infections [48–50]. For this reason, we make use of a discrete event simulator to more accurately model fine-scaled epidemic behavior. Thus where other efforts use a rate of infection in continuous time (and the ensuing non-linear term), we directly use probabilities for each discrete time step.

The biggest difference in our approach is how we treat infections within hyperedges. Studies such as [11] designate simplices as distinct group interactions with special rules for when infection can be transmitted that depends on the number of infected nodes. They may also embed all pairwise edges induced by a simplex and treat group spreading as a separate mechanism for spreading from pairwise spread. In particular, scaling the infection rate with the size of the hyperedge is uncommon. For instance, [46] uses an infection rate that scales with $i_h^{(t)}$ and β only but not the size of the hyperedge. The only exception we are aware of is [40], which uses a partitioned model that allows a different function for each value of m, but their analysis is based on a continuous-time formulation and shows that a mean-field approximation can be reduced to a weighted pairwise model. The research in [51] also uses a projected graph to explore stability points in higher-order dynamics. For a more detailed review on higher-order epidemics, see the survey papers [12] and [52].

5.3. Results from epidemics on our spatial hypergraph model

Throughout our remaining experiments, we use $n = 50,000$ node graphs in $d = 2$ with the same log-normal degree distribution with parameters $\log(3)$, 1, to give a mean initial degree of 3.5. We use the same exogenous infection rate for all experiments, $\eta = 5/1,000,000$, which corresponds to one exogenous infection every 4 time-steps. We also use the same recovery parameter $\gamma = 0.1$ for all experiments, which corresponds to an expected infection time of 10 time-steps. We will vary the infectivity parameter β and immunity parameter δ in the experiments. We run the simulation for 2000 time steps.

We simulate epidemics using a discrete event simulation to produce trajectories of the number of infected nodes. Detailed pseudo-code for the SIRS model on a general hypergraph is outlined in S2 Text. A few sample epidemic trajectories from a pairwise graph are shown in Fig 10 (left) where we vary the number of initially infected nodes. These all show convergence to a steady state over the time history of the epidemic. While higher-order epidemics can exhibit bistability, we are not in a bistable regime. This is illustrated in S2 Text and S2 Fig.

As we report on the results for other epidemic parameters, we found that the average number of infected nodes over the last 1000 timesteps was a reliable quantity. We show a histogram of this value over a number of distinct simulations in Fig 10 (right). This shows that the *maximum* difference between any simulation was around 100 infections.

5.4. Threshold behavior in the epidemics as the infection probability varies

We next study the same experiment but with a focus on the actual quantity of the average trailing infected nodes. Fig 11 shows the average trailing infections as we vary the amount of higher-order structure in our model (α) and the infection probability (β). As expected, each figure shows a clear epidemic threshold. As β increases, we go from a small steady state to

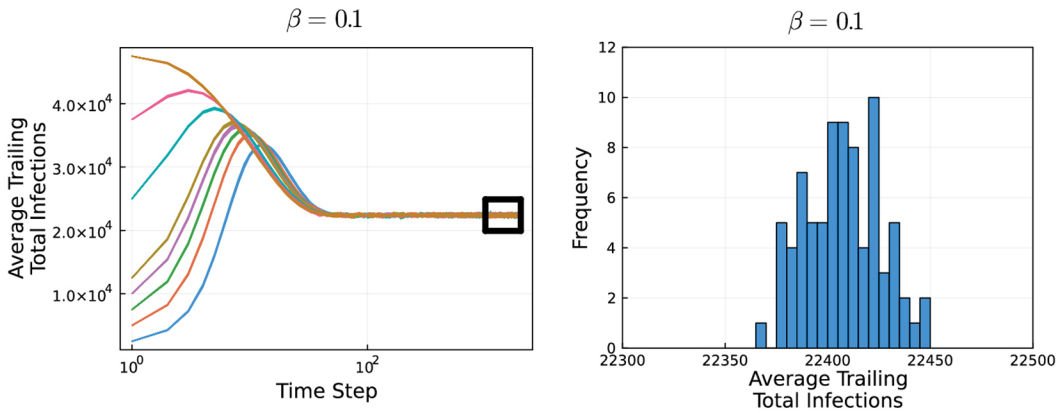

Fig 10. An example of infected node trajectories. (Left) An example of the infected node trajectories on a pairwise graph ($\alpha = 0$) with $n = 50000$, $d = 2$. Despite large changes to the fraction of initially infected nodes, these show convergence to a steady state. The black box indicates the last 1000 times steps, over which we compute the average number of infected nodes. (Right) This shows a histogram over the trailing infected nodes for a number of different samples showing that the average trailing number of infected nodes is a reliable quantity. Other epidemic parameters are fixed at $\gamma = \delta = 0.05$.

https://doi.org/10.1371/journal.pcsy.0000066.g010

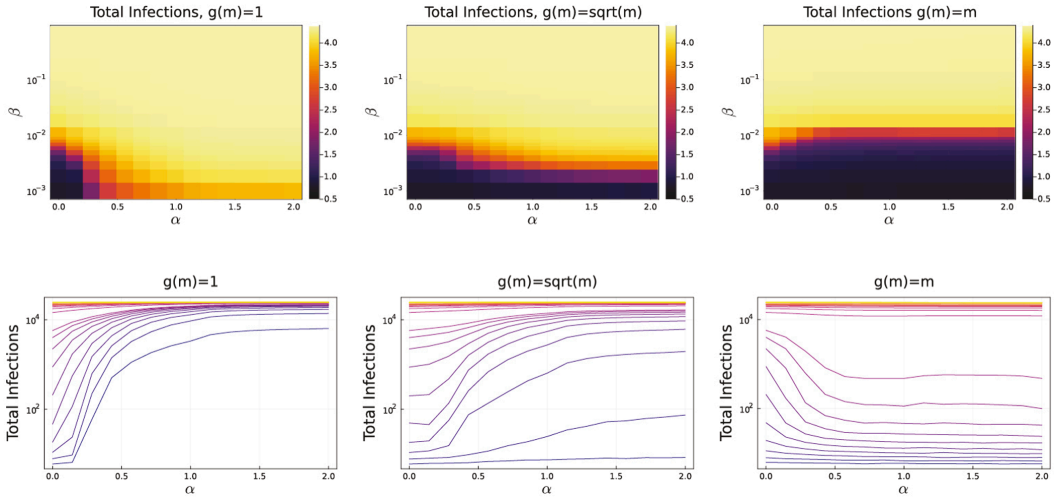

Fig 11. Ventilation can have non-intuitive effects on average trailing infections. The top row shows infections as a heatmap in α (higher-order structure) and β (infection probability) space while fixing $\gamma = \delta = 0.05$. The entries are log-scaled, so 4 means 10^4 average trailing infected nodes. The bottom row shows the same data as a set of lines where larger values of β correspond with more red / yellow colors. The columns correspond to the emulated ventilation functions $g(m) = 1, g(m) = \sqrt{m}, g(m) = m$. Note that, in the absence of strong ventilation (left column), increasing α increases total infections. In contrast, under linear ventilation (right column), increasing α decreases total infections

https://doi.org/10.1371/journal.pcsy.0000066.g011

an outbreak with about half the graph. A second observation is that, as we emulate a greater amount of ventilation ($g(m) = \sqrt{m}$ or $g(m) = m$), we need larger infection probabilities as we transition the graph from pairwise to higher-order structure.

Perhaps the most interesting observation is that the impact of increasing α is coupled with ventilation. When $g(m) = 1$ then each hyperedge represents a quadratic number of possible infection pathways. This corresponds to increasing the total number of edges, as we will see shortly (Fig 15). Consequently, we *expect* this to show that highly infected populations occur at lower infection probabilities, see e.g., [40]. That this same impact occurs for $g(m) = \sqrt{m}$

is also expected by the same reasoning. This simply doesn't change the probabilities enough to mask the overall increase in effective edges. In contrast, setting $g(m) = m$ should show roughly constant behavior as a function of α by the same reasoning. We do not see this behavior. In the case of linear ventilation (right column) increasing α causes total infections to fall—especially right around the threshold value of β. This indicates some coupling between ventilation and higher-order structure, which we will explore in the next section.

5.5. The impact of hyperedges varies with the epidemic parameters

Indeed, the relationship between our ventilation term $g(m)$ and the amount of higher-order structure is not straightforward. In our initial conference paper [16], we found that average trailing infections *increases* with α whereas in the previous section we found that average trailing infections *decreases* with α. In a small surprise, the impact is extremely sensitive to the epidemic parameters. In the case of linear ventilation, interpolating to higher-order structure can case total infections to increase or decrease in a non-linear fashion. We illustrate this while separately changing two different epidemic parameters, the infection probability β and the waning immunity term δ. Fig 12 shows average trailing total infections as we vary β while fixing $\delta = 0.01$ under linear ventilation. While the number of infections differs among those plots, they show dramatically different shapes depending on how much higher-order structure is present. Increasing α can cause total infections to decrease, increase, or produce non-linear mixed effects. Similar effects are present when varying δ instead of β in Fig 13. Note that column 3 of Fig 12 is the same as column 1 of Fig 13.

In pairwise epidemics, the dominant eigenvalue is related to the epidemic threshold in the mean-field and often used as a proxy for epidemic strength [53,54]. The total edge volume (total edges in pairwise graphs) is related to epidemic thresholds in randomized networks [55]. We compute both of these quantities in the pairwise projections to illustrate that the effect we are seeing cannot be explained by simple pairwise tools. In order to do so, we compute a weighted clique expansion of generated hypergraphs where hyperedges are weighted using the ventilation term $g(m)$. A single hyperedge of size m is mapped to a clique on m nodes with edge weights $1/g(m)$. We sum up the weights from all hyperedges on the same pair of nodes. We then compute the dominant eigenvalue λ_1 and sum of weighted degrees. These are the leftmost columns of Figs 14 and 15 respectively. In this case, increases in total infections are not due to changes in either the dominant projected eigenvalue of projected edge volumes.

In terms of mechanisms underlying this effect, we note that the overall changes are modest with respect to the population size. However, they are reliable and repeatable. We were

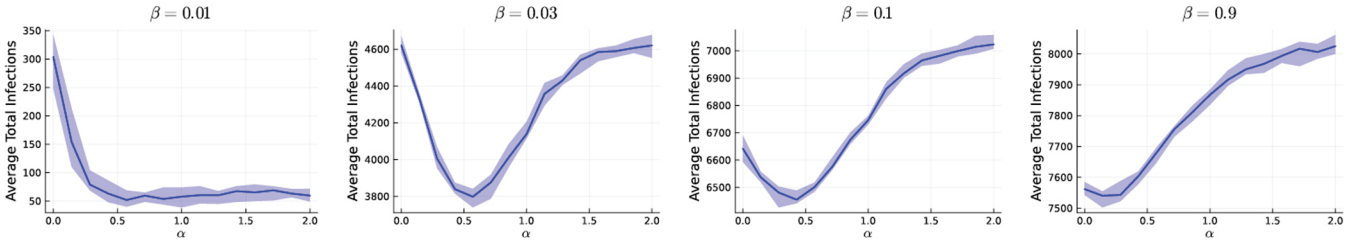

Fig 12. Average trailing infections is sensitive to epidemics parameters under our model. From the left to right, we show what happens as we increase the value of β for epidemic simulations for the waning immunity probability $\delta = 0.01$ under linear ventilation $g(m) = m$ with $\gamma = 0.05$. While average trailing infections differ among the plots, the impact of interpolating from a pairwise to higher-order is sensitive to epidemic parameters.

https://doi.org/10.1371/journal.pcsy.0000066.g012

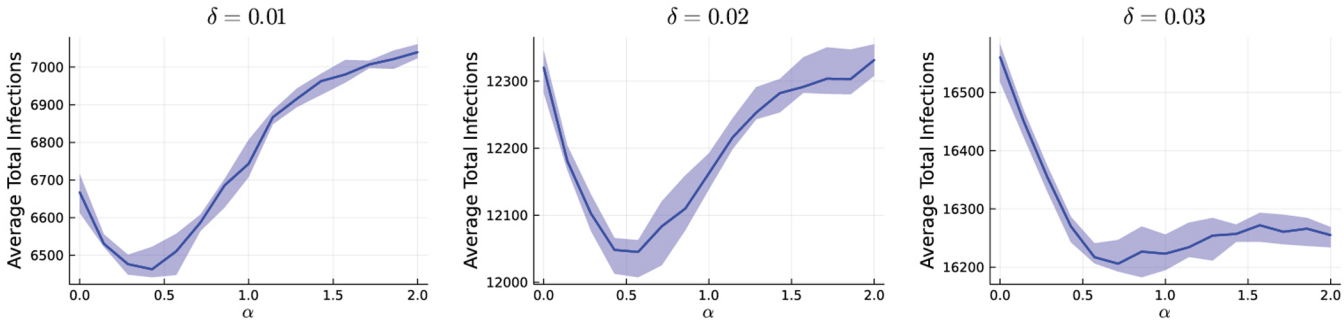

Fig 13. Changing a single epidemic parameter can alter shape of epidemics. From left to right, we increase the parameter δ while fixing other epidemic parameters ($\gamma = 0.05$ and $\beta = 0.1$) and recording average trailing infected nodes. All figures use linear ventilation, $g(m) = m$. The impact of higher-order structure (large α) can cause both growth and decay in the epidemic impact. Note that the leftmost figure is the same as column 3 of Fig 12.

https://doi.org/10.1371/journal.pcsy.0000066.g013

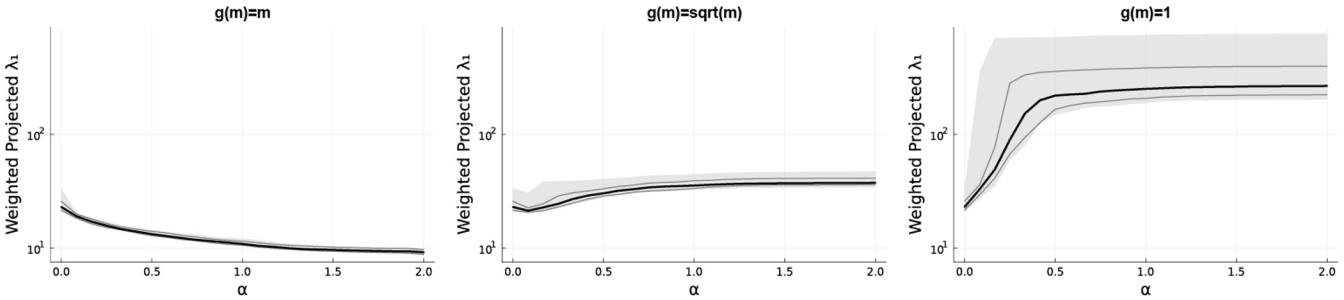

Fig 14. The dominant eigenvalue of weighted clique expansions from our model. Using the ventilation function $g(m)$, a single hyperedge of size m is projected to a clique with edge weights of $1/g(m)$. Gray and black lines indicate the 10th, 50th, and 90th percentiles while the gray bands denote the max and min over 25 trials. Linear ventilation causes the pairwise dominant eigenvalue to decrease while other choices for $g(m)$ generally cause an increase. This would correspond to a weaker epidemic, which is not what is always found in the experiments from Figs 12 and 13.

https://doi.org/10.1371/journal.pcsy.0000066.g014

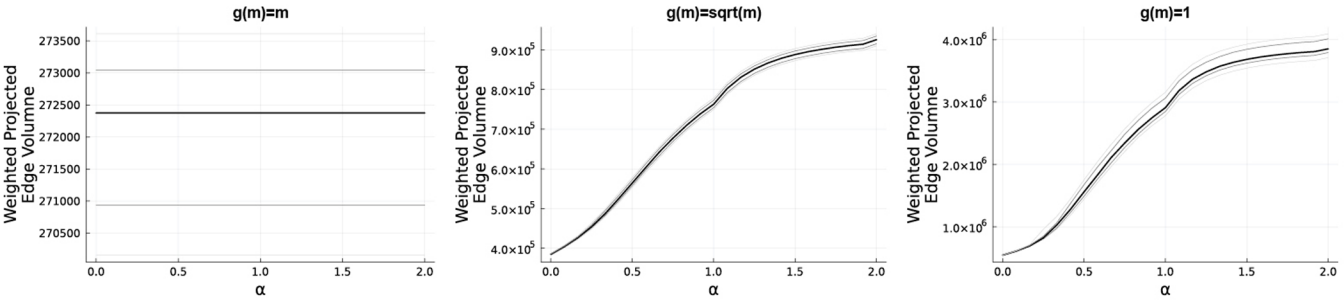

Fig 15. Projected edge volumes of weighted clique expansions from our model. A single hyperedge of size m is projected to a clique with pairwise edge weights $1/g(m)$. Gray and black lines indicate the 10th, 50th, and 90th percentiles. This quantity is constant for $g(m) = m$ while increasing for other choices of $g(m)$.

https://doi.org/10.1371/journal.pcsy.0000066.g015

unable to identify any large properties or differences by studying the epidemic propagation in detail, or looking at which hyperedges were responsible, at least beyond similarly small differences. For instance, when we studied the fraction of infection transmissions along edges or hyperedges, we could see that higher-order edges were more likely to transmit infections in the $\delta = 0.01$ and $\delta = 0.02$ scenario compared with the $\delta = 0.03$ scenario. But given that these edges account for a fairly small fraction of the overall transmissions in the highly ventilated case, we did not believe this finding to be mechanistically conclusive.

5.6. Changes in hyperedge transmissions by size below the pairwise threshold

Lastly, we examine transmissions by hyperedges of a given size in the steady state. What we call a transmission is when node i infects node j over a hyperedge. We focus on the case where there is no ventilation $g(m) = 1$. Recall that this is not a pairwise epidemic as a pair of nodes can interact in multiple hyperedges. We choose $\beta = 10^{-3}$, which is below the threshold in the pairwise case ($\alpha = 0$), and causes a large steady state epidemic in the higher-order case ($\alpha \to 2$). This will let us observe the impact of hyperedge transmission as total infections dramatically increases. The data are shown in Fig 16. For each epidemic trajectory, we compute the average transmissions by hyperedges of a given size over the last 1000 time steps (similar to Fig 10). We then max-normalize this information for each value of α. We find that initially, traditional pairwise edges are responsible for most transmissions but as the total number of infected nodes increases there is a transition to larger hyperedges transmitting infections. For $0.3 < \alpha < 0.6$, transmissions primarily occur due to large hyperedges or small hyperedges. However, as we move to larger values of $1 \leq \alpha \leq 2$, medium-sized hyperedges become responsible for most transmissions. Though as shown in the rightmost Fig 16, the effect at $\alpha = 2$ can partially be explained by projected edge volumes.

In this case, we again see the utility of our model. Note the two regimes: for small α, which cause a transition to an epidemic, we see that *large hyperedges* matter more than medium size hyperedges. Whereas in a strong, steady state epidemic at large α, then its the medium-size hyperedges that matter.

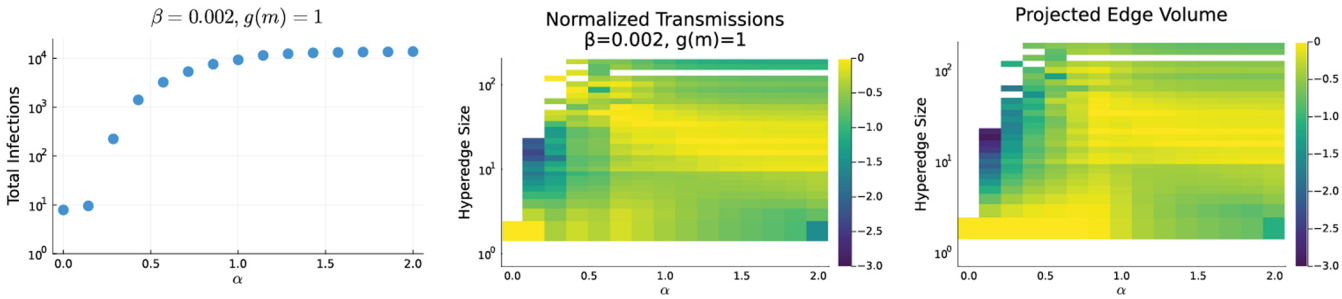

Fig 16. Without ventilation effects, higher order structure can dominate transmissions. (Left) As we increase α, the underlying higher-order structure increases causing a dramatic increase in average trailing infected nodes in the $g(m) = 1$ case to model no ventilation. (Middle) We collect statistics on which hyperedge size was responsible for transmissions of an infection. We display the probability that a transmission occurred with a hyperedge of a given size in the steady state (max-normalized for each α). At $\alpha = 0$, all infections are transmitted by pairwise edges because these are all that exist. For α a little bit larger (0.3 to 0.6), the larger hyperedges around 10^2 dominate transmission right around the threshold. Then for α larger than around 1.0, we see a transition to medium size hyperedges dominating. This persists until $\alpha = 2$. (Right) The effect at $\alpha = 2$ can partially be explained by projected edge volumes of those higher-order edges accounting for an out-sized portion of edges.

https://doi.org/10.1371/journal.pcsy.0000066.g016

6. Discussion

Pairwise interactions have enabled a large body of research with applications ranging from power grid robustness [56] to heterogeneity and disease spread [57–59] to accelerating materials and molecule discovery [60]. The non-trivial effects of higher-order structure coupled with competing generalizations for intuitive pairwise concepts have posed a significant challenge for researchers with a grounded understanding in pairwise data [10,61]. This manifests in a number of ways including inconclusive and contradictory findings depending on exactly how the higher-order problem is realized. We saw a key illustration of this in terms of the clustering coefficients (Sect 3), where we investigated and saw a number of distinct properties of various generalizations of clustering coefficients to hypergraphs [32,33,36,62]. As an example beyond what we have looked at, consider the impact of representation in the behavior of higher-order synchronization [13]. In this case, the choice of data representation (simplex versus hypergraph) can alter the behavior of synchronization.

The utility of our model is that we can isolate effects to the impact of the higher-order pieces from the rest of the graph model. We presented a few studies of this behavior but many more are possible. As it relates to epidemics and other dynamics, our model gives a parametric way to couple dynamics to higher-order structure through the parameter α. This leads to some interseting questions about whether even simplier epidemic models such as SIS would yield similar results where the impact of higher-order structure is intricately coupled with epidemic parameters. On the network analysis side, a quantitative understanding of how the interpolation parameter, α, maps to popular network statistics or fitting the model to data would be an interesting avenue of exploration. We are optimistic that our model will be broadly useful in the future to help understand the origin of ambiguities in the behavior of higher-order network models in the future.

There are many possible extensions of this model. For instance, we could consider extensions to *gravity-like* model wherein the link or move around based on gravity-like terms with their neighbors [63,64]. In terms of epidemics, pairwise epidemics behave differently in the presence of interventions [64], extending these types of interventions to hypergraphs would shed further insights into the role of higher-order spreading in more realistic scenarios. One avenue of future work we plan to investigate is *fitting* our model to a dataset. In this case, we anticipate that a combination of modern vertex embedding techniques [65–67] will yield interesting models that are broadly similar to input networks.

Supporting information

S1 Text. Text discussing the choice of clustering distance parameter and differences from a conference version of this work.
(PDF)

S1 Fig. Figure showing the total number of hyperedges using different radius functions for scaling the radius parameter in DBSCAN as we vary dimensions. The functions used for ε_α are: Eq (1) (leftmost column), linear scaling with $\varepsilon_\alpha = \alpha r_v/2$ (middle column), and scaling by α instead of $\alpha^{1/d}$ when $\alpha \leq 1$ (rightmost column). Gray to black to gray bands (only visible when zoomed in) in the middle and right columns indicate 10th, 25th, 50th, 75th, and 90th percentiles. This shows that our choice of Eq (1) gives the smoothest interpolation between pairwise effects and pure hyperedge effects as α varies from 0 to 2. Note that the leftmost column corresponds to the subplots from Fig 5.
(TIF)

S2 Text. Text discussion a higher order generalization of PageRank to a hypergraph. It closerly follows [17].
(PDF)

S3 Text. Text containing our psudeo-code for our implementation of a hypergraph SIRS model. Our code is also publicly available at https://github.com/oeldaghar/spatial-hypergraph-epidemics.
(PDF)

S4 Text. Supplemental text discussing the lack of bistability in our higher-order epidemic model.
(PDF)

S2 Fig. Figure showing that the average trailing number of infected nodes is a reliable quantity and we are not in a bistable regime for our epidemic simulations. In this case, epidemics are seeded using an initial number of infected nodes of 5%, 10%, 15%, 20%, 25%, 50%, 75%, and 95% of the graph. Other epidemic parameters are fixed with $\gamma = \delta = 0.05$. Each bin corresponds to 80 simulations (10 per each initial infected fraction). We compute the range of total infections for those simulations in the steady state. The maximum difference is less than 600 infections (which reflects a worse-case fluctuation of about 1% of nodes) right around the epidemic threshold.
(TIF)

Author contributions

Conceptualization: Omar Eldaghar, Yu Zhu, David F. Gleich.

Data curation: Omar Eldaghar, Yu Zhu, David F. Gleich.

Formal analysis: Omar Eldaghar, Yu Zhu, David F. Gleich.

Funding acquisition: David F. Gleich.

Investigation: Omar Eldaghar, Yu Zhu, David F. Gleich.

Methodology: Omar Eldaghar, Yu Zhu, David F. Gleich.

Project administration: David F. Gleich.

Resources: David F. Gleich.

Software: Omar Eldaghar, Yu Zhu, David F. Gleich.

Supervision: David F. Gleich.

Validation: Omar Eldaghar, Yu Zhu, David F. Gleich.

Visualization: Omar Eldaghar, Yu Zhu, David F. Gleich.

Writing – original draft: Omar Eldaghar, Yu Zhu, David F. Gleich.

Writing – review & editing: Omar Eldaghar, Yu Zhu, David F. Gleich.

References

1. Watts DJ, Strogatz SH. Collective dynamics of “small-world” networks. Nature. 1998;393(6684):440–2. https://doi.org/10.1038/30918 PMID: 9623998
2. Kleinberg J. The small-world phenomenon: an algorithmic perspective. In: Proceedings of the Thirty-Second Annual ACM Symposium on Theory of Computing, 2000. p. 163–70.

3. Lancichinetti A, Fortunato S, Radicchi F. Benchmark graphs for testing community detection algorithms. Phys Rev E Stat Nonlin Soft Matter Phys. 2008;78(4 Pt 2):046110. https://doi.org/10.1103/PhysRevE.78.046110 PMID: 18999496
4. Erdos P, Renyi A. On random graphs I. Publicationes Mathematicae Debrecen. 1959;6:290–7.
5. Erdös P, Rényi A. On the evolution of random graphs. The Structure and Dynamics of Networks. Princeton University Press; 2011. p. 38–82. https://doi.org/10.1515/9781400841356.38
6. Erdős P, Rényi A. On the strength of connectedness of a random graph. Acta Mathematica Academiae Scientiarum Hungaricae. 1964;12(1–2):261–7. https://doi.org/10.1007/bf02066689
7. Gilbert EN. Random graphs. Ann Math Statist. 1959;30(4):1141–4. https://doi.org/10.1214/aoms/1177706098
8. Kuramoto Y. Self-entrainment of a population of coupled non-linear oscillators. In: Kyoto University, Kyoto/Japan; 1975. p. 420–2.
9. Strogatz SH. From Kuramoto to Crawford: exploring the onset of synchronization in populations of coupled oscillators. Physica D: Nonlinear Phenomena. 2000;143(1–4):1–20. https://doi.org/10.1016/s0167-2789(00)00094-4
10. Benson AR, Gleich DF, Leskovec J. Higher-order organization of complex networks. Science. 2016;353(6295):163–6. https://doi.org/10.1126/science.aad9029 PMID: 27387949
11. Iacopini I, Petri G, Barrat A, Latora V. Simplicial models of social contagion. Nat Commun. 2019;10(1):2485. https://doi.org/10.1038/s41467-019-10431-6 PMID: 31171784
12. Boccaletti S, De Lellis P, del Genio CI, Alfaro-Bittner K, Criado R, Jalan S, et al. The structure and dynamics of networks with higher order interactions. Physics Reports. 2023;1018:1–64. https://doi.org/10.1016/j.physrep.2023.04.002
13. Zhang Y, Lucas M, Battiston F. Higher-order interactions shape collective dynamics differently in hypergraphs and simplicial complexes. Nat Commun. 2023;14(1):1605. https://doi.org/10.1038/s41467-023-37190-9 PMID: 36959174
14. Veldt N, Benson AR, Kleinberg J. Combinatorial characterizations and impossibilities for higher-order homophily. Sci Adv. 2023;9(1):eabq3200. https://doi.org/10.1126/sciadv.abq3200 PMID: 36608141
15. Battiston F, Cencetti G, Iacopini I, Latora V, Lucas M, Patania A. Networks beyond pairwise interactions: structure and dynamics. Physics Reports. 2020;874:1–92.
16. Eldaghar O, Zhu Y, Gleich DF. A spatial hypergraph model where epidemic spread demonstrates clear higher-order effects. In: Proceedings of Complex Networks and their Applications, 2024.
17. Liu M, Veldt N, Song H, Li P, Gleich DF. Strongly local hypergraph diffusions for clustering and semi-supervised learning. In: Proceedings of the Web Conference 2021. 2021. p. 2092–103.
18. Landry NW, Restrepo JG. The effect of heterogeneity on hypergraph contagion models. Chaos. 2020;30(10):103117. https://doi.org/10.1063/5.0020034 PMID: 33138447
19. Ester M, Kriegel HP, Sander J, Xu X. A density-based algorithm for discovering clusters in large spatial databases with noise. In: KDD; 1996. p. 226–31.
20. Diaz J, Mitsche D, Perez X. Dynamic random geometric graphs. arXiv preprint 2007. https://arxiv.org/abs/cs/0702074
21. Zhu Y, Gleich DF. A survey on random hypergraph models with connections to bipartite graphs and simplicial complexes. 2025.
22. Bonato A, Janssen J, Prałat P. Geometric protean graphs. Internet Mathematics. 2012;8(1–2):2–28. https://doi.org/10.1080/15427951.2012.625246
23. Bonato A, Gleich DF, Kim M, Mitsche D, Prałat P, Tian Y, et al. Dimensionality of social networks using motifs and eigenvalues. PLoS One. 2014;9(9):e106052. https://doi.org/10.1371/journal.pone.0106052 PMID: 25188391
24. Barthelemy M. Spatial networks: a complete introduction: from graph theory and statistical physics to real-world applications. Springer Nature; 2022.
25. Barthelemy M. Class of models for random hypergraphs. Phys Rev E. 2022;106(6–1):064310. https://doi.org/10.1103/PhysRevE.106.064310 PMID: 36671196
26. de Kergorlay H-L, Higham DJ. Connectivity of random geometric hypergraphs. Entropy (Basel). 2023;25(11):1555. https://doi.org/10.3390/e25111555 PMID: 37998246
27. Kahle M. Topology of random clique complexes. Discrete Mathematics. 2009;309(6):1658–71. https://doi.org/10.1016/j.disc.2008.02.037
28. Kahle M. Topology of random simplicial complexes: a survey. Contemporary Mathematics. American Mathematical Society; 2014. p. 201–21. https://doi.org/10.1090/conm/620/12367
29. Bobrowski O, Kahle M. Topology of random geometric complexes: a survey. J Appl and Comput Topology. 2018;1(3–4):331–64. https://doi.org/10.1007/s41468-017-0010-0

30. Lunagómez S, Mukherjee S, Wolpert RL, Airoldi EM. Geometric representations of random hypergraphs. Journal of the American Statistical Association. 2017;112(517):363–83. https://doi.org/10.1080/01621459.2016.1141686
31. Turnbull K, Lunagómez S, Nemeth C, Airoldi E. Latent space modeling of hypergraph data. Journal of the American Statistical Association. 2024;119(548):2634–46.
32. Miyashita R, Hironaka S, Shudo K. Clustering coefficient reflecting pairwise relationships within hyperedges. arXiv preprint 2024. https://arxiv.org/abs/2410.23799
33. Ha G-G, Neri I, Annibale A. Clustering coefficients for networks with higher order interactions. Chaos. 2024;34(4):043102. https://doi.org/10.1063/5.0188246 PMID: 38558051
34. Hagberg AA, Schult DA, Swart PJ. Exploring network structure, dynamics, and function using NetworkX. In: Proceedings of the Python in Science Conference. 2008. p. 11–5. https://doi.org/10.25080/tcwv9851
35. Fagiolo G. Clustering in complex directed networks. Phys Rev E Stat Nonlin Soft Matter Phys. 2007;76(2 Pt 2):026107. https://doi.org/10.1103/PhysRevE.76.026107 PMID: 17930104
36. Robins G, Alexander M. Small worlds among interlocking directors: network structure and distance in bipartite graphs. Computational & Mathematical Organization Theory. 2004;10:69–94.
37. Aksoy SG, Kolda TG, Pinar A. Measuring and modeling bipartite graphs with community structure. Journal of Complex Networks. 2017;5(4):581–603. https://doi.org/10.1093/comnet/cnx001
38. Li P, He N, Milenkovic O. Quadratic decomposable submodular function minimization: theory and practice. Journal of Machine Learning Research. 2020;21(106):1–49.
39. Takai Y, Miyauchi A, Ikeda M, Yoshida Y. Hypergraph clustering based on PageRank. In: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2020. p. 1970–8. https://doi.org/10.1145/3394486.3403248
40. Higham DJ, de Kergorlay H-L. Epidemics on hypergraphs: spectral thresholds for extinction. Proc Math Phys Eng Sci. 2021;477(2252):20210232. https://doi.org/10.1098/rspa.2021.0232 PMID: 35153574
41. Li Z, Deng Z, Han Z, Alfaro-Bittner K, Barzel B, Boccaletti S. Contagion in simplicial complexes. Chaos, Solitons & Fractals. 2021;152:111307.
42. Lu J, Gu J, Li K, Xu C, Su W, Lai Z, et al. COVID-19 outbreak associated with air conditioning in restaurant, Guangzhou, China, 2020. Emerg Infect Dis. 2020;26(7):1628–31. https://doi.org/10.3201/eid2607.200764 PMID: 32240078
43. Chang S, Pierson E, Koh PW, Gerardin J, Redbird B, Grusky D, et al. Mobility network models of COVID-19 explain inequities and inform reopening. Nature. 2021;589(7840):82–7. https://doi.org/10.1038/s41586-020-2923-3 PMID: 33171481
44. Dixit AK, Espinoza B, Qiu Z, Vullikanti A, Marathe MV. Airborne disease transmission during indoor gatherings over multiple time scales: modeling framework and policy implications. Proc Natl Acad Sci U S A. 2023;120(16):e2216948120. https://doi.org/10.1073/pnas.2216948120 PMID: 37036987
45. ASHRAE. Ventilation for Acceptable Indoor Air Quality. ANSI/ASHRAE standard 62.1–2019 ed. ASHRAE; 2019.
46. Bodó Á, Katona GY, Simon PL. SIS epidemic propagation on hypergraphs. Bull Math Biol. 2016;78(4):713–35. https://doi.org/10.1007/s11538-016-0158-0 PMID: 27033348
47. Suo Q, Guo J-L, Shen A-Z. Information spreading dynamics in hypernetworks. Physica A: Statistical Mechanics and its Applications. 2018;495:475–87. https://doi.org/10.1016/j.physa.2017.12.108
48. Bansal S, Grenfell BT, Meyers LA. When individual behaviour matters: homogeneous and network models in epidemiology. J R Soc Interface. 2007;4(16):879–91. https://doi.org/10.1098/rsif.2007.1100
49. Volz EM, Miller JC, Galvani A, Ancel Meyers L. Effects of heterogeneous and clustered contact patterns on infectious disease dynamics. PLoS Comput Biol. 2011;7(6):e1002042. https://doi.org/10.1371/journal.pcbi.1002042 PMID: 21673864
50. Großmann G, Backenköhler M, Wolf V. Heterogeneity matters: contact structure and individual variation shape epidemic dynamics. PLoS One. 2021;16(7):e0250050. https://doi.org/10.1371/journal.pone.0250050 PMID: 34283842
51. Ferraz de Arruda G, Tizzani M, Moreno Y. Phase transitions and stability of dynamical processes on hypergraphs. Commun Phys. 2021;4(1). https://doi.org/10.1038/s42005-021-00525-3
52. Ferraz de Arruda G, Aleta A, Moreno Y. Contagion dynamics on higher-order networks. Nat Rev Phys. 2024;6(8):468–82. https://doi.org/10.1038/s42254-024-00733-0
53. Chakrabarti D, Wang Y, Wang C, Leskovec J, Faloutsos C. Epidemic thresholds in real networks. ACM Trans Inf Syst Secur. 2008;10(4):1–26. https://doi.org/10.1145/1284680.1284681
54. Prakash BA, Chakrabarti D, Faloutsos M, Valler N, Faloutsos C. Threshold conditions for arbitrary cascade models on arbitrary networks. In: 2011 IEEE 11th ICDM. 2011. p. 537–46.

55. Castellano C, Pastor-Satorras R. Thresholds for epidemic spreading in networks. Phys Rev Lett. 2010;105(21):218701. https://doi.org/10.1103/PhysRevLett.105.218701 PMID: 21231361
56. Amani AM, Jalili M. Power grids as complex networks: resilience and reliability analysis. IEEE Access. 2021;9:119010–31.
57. Newman MEJ. Spread of epidemic disease on networks. Phys Rev E Stat Nonlin Soft Matter Phys. 2002;66(1 Pt 2):016128. https://doi.org/10.1103/PhysRevE.66.016128 PMID: 12241447
58. Pastor-Satorras R, Castellano C, Van Mieghem P, Vespignani A. Epidemic processes in complex networks. Reviews of Modern Physics. 2015;87(3):925–79.
59. Moreno Y, Pastor-Satorras R, Vespignani A. Epidemic outbreaks in complex heterogeneous networks. Eur Phys J B. 2002;26(4):521–9. https://doi.org/10.1140/epjb/e20020122
60. Yang N, Wu H, Zeng K, Li Y, Bao S, Yan J. Molecule generation for drug design: a graph learning perspective. Fundamental Research. 2024. https://doi.org/10.1016/j.fmre.2024.11.027
61. Benson AR, Gleich DF, Higham DJ. Higher-order network analysis takes off, fueled by classical ideas and new data. arXiv preprint 2021. https://arxiv.org/abs/2103.05031
62. Yin H, Benson AR, Leskovec J. Higher-order clustering in networks. Phys Rev E. 2018;97(5–1):052306. https://doi.org/10.1103/PhysRevE.97.052306 PMID: 29906904
63. Cabanas-Tirapu O, Danu´s L, Moro E, Sales-Pardo M, Guimer a R. Human mobility is well described by closed-form gravity-like models learned automatically from data. arXiv preprint 2023. https://arxiv.org/abs/2312.11281
64. Eldaghar O, Mahoney MW, Gleich DF. Multi-scale local network structure critically impacts epidemic spread and interventions. arXiv preprint 2023. https://arxiv.org/abs/2312.17351
65. Grover A, Leskovec J. node2vec: scalable feature learning for networks. KDD. 2016;2016:855–64. https://doi.org/10.1145/2939672.2939754 PMID: 27853626
66. Perozzi B, Al-Rfou R, Skiena S. DeepWalk. In: Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2014. p. 701–10. https://doi.org/10.1145/2623330.2623732
67. Tang J, Qu M, Wang M, Zhang M, Yan J, Mei Q. LINE. In: Proceedings of the 24th International Conference on World Wide Web, 2015. p. 1067–77. https://doi.org/10.1145/2736277.2741093