H^{2}ABM: Heterogeneous Agent-based Model on Hypergraphs to Capture Group Interactions

Vivek Anand*† Jiaming Cui*† Jack Heavey‡ Anil Vullikanti‡
B. Aditya Prakash†

Abstract

Heterogeneous agent-based models (HABMs) can simulate the dynamics of multiple types of entities and their interactions on contact networks. In recent years, they have gathered great interest and are widely applied in multiple fields, such as personalized recommendations, publication ranking, and epidemic modeling. Nevertheless, conventional HABMs on graphs can only capture pair-wise interactions between agents but fail to capture the more complex dynamics of group interactions (e.g., multiple people in the same location simultaneously), consequently leading to suboptimal performance. To address this, we propose using hypergraphs to capture such group interactions better and extend the current graph-based HABMs to hypergraphs. Specifically, we use MRSA (Methicillin-resistant *Staphylococcus aureus*, a kind of infectious disease acquired by patients during treatment at healthcare facilities) spread in the University of Virginia hospital as an example to showcase how we extend an existing graph-based HABM, Graph-HeterSIS, to a hypergraph-based HABM (H^{2}ABM), Hypergraph-HeterSIS. We show how the hypergraphs can capture the structural difference between contacts before and during the first wave of COVID-19 outbreak in Virginia better than graphs. Our experiments show that H^{2}ABM better captures the underlying group interactions and better fits and forecasts MRSA cases.

Keywords: Heterogeneous agent-based models, Hypergraphs, Modeling, Forecasting

1 Introduction

A **H**eterogeneous **A**gent-**b**ased **M**odel (HABM) is a computational model used to simulate the actions and interactions of multiple types of entities [9,10,13,16,20]. They are "heterogeneous" since they explicitly account for differences in characteristics, behaviors, or decision-making rules among different kinds of entities. For example, a HABM to model Healthcare-associated infections (HAI, infections acquired by patients during treatment at healthcare facilities such as hospitals and nursing homes [12,19,23]) may model the actions of patients, healthcare workers, locations as well as their interactions [17]. In recent years, HABMs have gathered great interest with applications in personalized recommendation [15], publication ranking [28], and drug design [31].

Current HABMs usually use graphs to capture the interactions between entities. However, graphs can only represent pair-wise interactions, while in the real world, many interactions are in groups. For example, as shown in Figure 1(a), only vertices v_1 and v_2 are at location L_1 from $t = 10$ to $t = 15$. However in Figure 1(b), the graph can only break the co-locations between v_1, v_2, and L_1 as 3 pair-wise interactions (edges between v_1 and v_2, v_1 and L_1, v_2 and L_1). Additionally, one cannot determine if v_2 interacted with both v_1 and v_3 together or with each of them separately from the graph. This makes the graphs less preferable in capturing complex dynamics of group interactions.

In this work, we propose to use hypergraphs to better capture such group interactions. As shown in Figure 1(c), the interactions between v_1, v_2, and L_1 from $t = 10$ to $t = 15$ can be easily represented by the yellow hyperedge. Similarly, the interactions between v_1, v_2, v_3, and L_1 from $t = 15$ to $t = 30$ can also be represented by the green hyperedge. From these two hyperedges, one can easily determine that v_2 interacted with v_1 and v_3 together once and interacted with only v_1 once. This shows that hypergraphs can capture the more complex group interactions more effectively. We also provide a more detailed explanation for this example in Section 4.

However, switching from graphs to hypergraphs also requires us to extend the HABMs to hypergraphs. This is non-trivial since the dynamics of group interactions cannot simply be considered as the linear combination of each pair-wise interaction. In this work, we use MRSA (Methicillin-resistant *Staphylococcus aureus*, a kind of HAI) spread at the University of Virginia

*Both authors contributed equally to this paper
†College of Computing, Georgia Institute of Technology, Email: {vivekanand, jiamingcui1997, badityap}@gatech.edu.
‡Department of Computer Science, University of Virginia, Email: {jch7jm, vsakumar}@virginia.edu

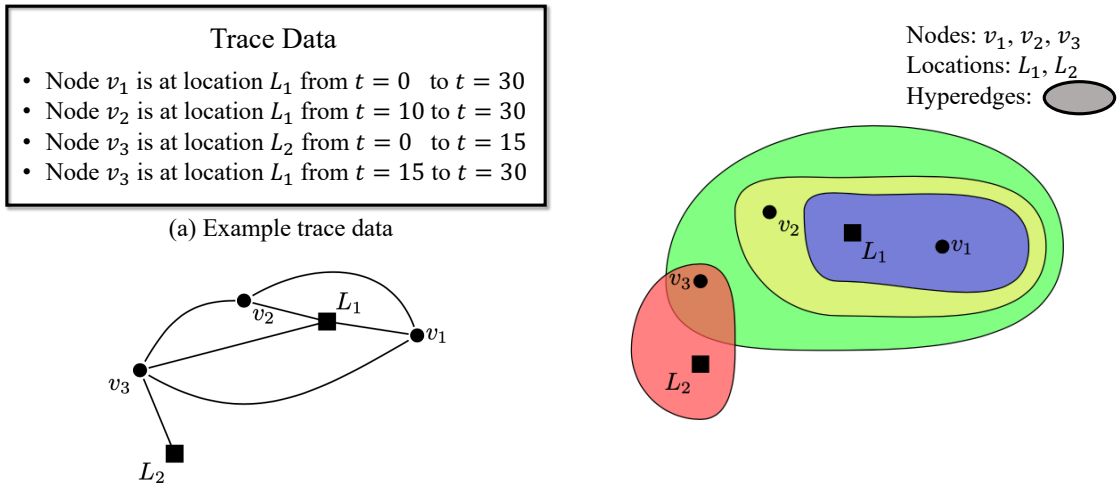

Figure 1: Example of graph and hypergraph generated from trace data to capture the interactions of vertices and locations. (a) An example trace data of vertices. It traces the location (L_1 and L_2) of each vertex (v_1 to v_3) for every timestep t. (b) The graph generated from example trace data in (a). Dots and squares represent vertices and locations, respectively. An edge between two entities represents the pair-wise relation between them. For example, only v_1 and v_2 are at the same location L_1 from $t = 10$ to $t = 15$. Therefore, there is an edge between v_1 and v_2. (c) The hypergraph generated from example trace data in (a). Each colored circle corresponds to a hyperedge, which represents a group interaction between the vertices and locations inside it. For example, only v_1 and v_2 are at the same location L_1 from $t = 10$ to $t = 15$, therefore there is a hyperedge (yellow one) among v_1, v_2, and L_1.

hospital and a widely-used graph-based HABM for HAI modeling, Graph-HeterSIS model [17], as an example to showcase how we can easily extend from a graph-based HABM to a **H**ypergraph-based **HABM** (H^2ABM). We refer to this H^2ABM as Hypergraph-HeterSIS model in later sections.

Our main contributions are summarized below.

1. We use the widely-used Graph-HeterSIS model as a simple example to show how we can extend from a graph-based HABM to a hypergraph-based HABM (H^2ABM) and propose the Hypergraph-HeterSIS model which incorporates group level nonlinear load discounting.
2. We show that hypergraphs can capture the structural difference better than graphs can. Specifically, we analyze the contact networks in the University of Virginia (UVA) hospital before and during the first wave of COVID-19 outbreak in Virginia and show that hypergraphs are more effective in capturing structural differences between the two contact networks than graphs.
3. We also show that the Hypergraph-HeterSIS leads to better fitting and forecasting of MRSA cases in the UVA hospital than Graph-HeterSIS. It also provides good interpretability of the MRSA outbreak in UVA hospital.

The rest of the paper is organized in the following way: Section 2 discusses the related works. In Section 3, we introduce the Graph-HeterSIS model. In Section 4, we motivate the need for hypergraphs for HABMs, introduce Hypergraph-HeterSIS model and demonstrate its additional expressivity to Graph-HeterSIS due to nonlinear load discounting. In section 5, we evaluate the performance of Hypergraph-HeterSIS and Graph-HeterSIS model on UVA hospital networks and data. We then discuss future work and conclude in Section 6.

2 Related work

Heterogeneous ABMs. HABMs have been widely used to study many application problems, such as personalized recommendation [15], publication ranking [28], drug design [31], and infectious disease modeling [17]. One of the most important applications among them is for HAI modeling [1,17,20]. Since both people (e.g., patients or healthcare workers (HCWs)) and contaminated surfaces can spread pathogens [3,27], their dynamics can be significantly different (e.g., patients or HCWs can move in hospital, while locations cannot move), making it less suitable for conventional homogeneous ABMs.

Hypergraphs. Hypergraphs have been used for many applications, such as rumor detection [18], business modeling [4], recommendation system design [29], and

Table 1: List of notations

Notation	Description
α	Pathogen shedding rate
β	Disease infectivity
δ	Recovery probability
τ_{ijt}	Pathogen transfer ratio from node j to node i at time t (Graph-HeterSIS model only)
Υ_{ijt}	Pathogen transfer ratio from node j to node i at time t (Hypergraph-HeterSIS model only)
$\boldsymbol{l}_t$	Pathogen load vector at time t sized $N \times 1$
$\boldsymbol{x}_t$	Infection state vector at time t sized $P \times 1$
$\boldsymbol{A}_t$	Adjacency matrix at time t sized $N \times N$
$\boldsymbol{R}_t$	Pathogen transfer matrix at time t sized $N \times N$
$\boldsymbol{H}_t$	Hypergraph matrix at time t sized $M_t \times N$ (Hypergraph-HeterSIS model only)
M_t	Number of hyperedges
$g(\cdot)$	Nonlinear discounting load accumulation function (Hypergraph-HeterSIS model only)
P	Total number of patients
H	Total number of healthcare workers (HCWs)
L	Total number of locations
N	Total number of entities ($P + H + L$)

financial transactions [11]. Recently, hypergraphs have been used for infectious disease modeling, most notably for COVID-19 [14]. However, in their work, they assume that every vertex is of the same type (i.e., the model is homogeneous). To the best of our knowledge, we are the first to use hypergraphs for heterogeneous ABMs.

3 Preliminaries

In this section, we will introduce a widely-used graph-based HABM, for HAIs, Graph-HeterSIS.

3.1 Graph-HeterSIS Graph-HeterSIS is a heterogeneous agent-based load-sharing SIS model over a graph [17]. It has been widely used to model HAIs such as MRSA in hospital settings. There are three different types of entities, each with different behaviors: Patients are mobile, can accumulate, transmit and get infected by pathogen; Healthcare Workers (HCWs) are mobile, can accumulate and transmit pathogen but cannot get infected; Locations cannot move but can accumulate and share pathogen.

Graph-HeterSIS is described in Algorithm 1. A vertex i can be either *susceptible* ($\mathbf{x}_t(i) = 0$) or *infected* ($\mathbf{x}_t(i) = 1$). In addition, the model keeps track of pathogen loads on all people and locations using a load vector $\boldsymbol{l}_t$ at each timestep t. Using the adjacency matrix $\mathbf{A}_t$ ($\mathbf{A}_t(i,j) = 1$ vertex if i and j interacted at time t and $\mathbf{A}_t(i,j) = 0$ otherwise) and the load transfer ratios τ_{ijt} (τ_{ijt} represents the ratio of loads transferred from vertex j to vertex i at time t), we can compute the pathogen transfer matrix $\boldsymbol{R}_t$ and write the pathogen load updates as a linear operation as seen in line 4 in Algorithm 1 (note that the column-sums of $\boldsymbol{R}_t$ are restricted to be less than or equal to 1, which implies that the total amount of pathogen cannot increase after transfer (i.e., $\|\boldsymbol{R}_t\boldsymbol{l}_t\|_1 \leq \|\boldsymbol{l}_t\|_1$)). The probability that a patient i gets infected from susceptible is a dose-response function proportional to $\boldsymbol{l}_t(i)$, as shown in line 7 in Algorithm 1 with β as the disease infectivity. Once infected, the patient i sheds α additional pathogen per timestep to his own load, which can later be transferred to neighbors (both people and locations) via edges; this shedding continues until the patient recovers with recovery probability δ. We list the notations in Table 1.

3.2 Hypergraphs Hypergraphs are generalizations of graphs, where each hyperedge can include any number of vertices [30]. Compared with the edges of a graph that can only connect exactly two vertices, hyperedges are more flexible in capturing complex, multiple-entity relations in addition to the pairwise relations.

4 Hypergraph-HeterSIS model

In this section, we motivate the need for hypergraphs in modelling HAIs and subsequently introduce Hypergraph-HeterSIS. We then showcase the additional expressivity of Hypergraph-HeterSIS compared to Graph-HeterSIS due to the presence of nonlinearity.

4.1 Hypergraph formulation Given a set of N vertices $V = \{v_1, v_2, \ldots v_N\}$ that interact with each other ($\|V\| = N$) at each timestep t, we use a incidence matrix $\boldsymbol{H}_t$ to represent the interactions between vertices at time t (i.e., for T timesteps, we have $\boldsymbol{H}_1$, $\boldsymbol{H}_2$, $\cdots$, $\boldsymbol{H}_T$, and we use the same vertex set V for all $\boldsymbol{H}_t$). The dimension of hypergraph matrix $\boldsymbol{H}_t$ is $M_t \times N$, where M_t is the number of hyperedges at time t. Each row in $\boldsymbol{H}_t$ is a separate hyperedge h, and nonzero entries indicate membership of the respective vertices. For example, for the hypergraph shown in Figure 1(c), the yellow hyperedge between v_1, v_2 and L_1 can be written as $[1, 1, 0, 1, 0]$, where the five elements correspond to v_1, v_2, v_3, L_1, and L_2 respectively.

4.2 Limitations of graph-based HABMs Although graph-based HABMs can capture the pair-wise interactions between entities well, they are less suitable to model complex group interactions. This is because they can only break group interactions into multiple pair-wise interactions (e.g., breaking an interaction among three people to 3 pair-wise edges between each of two), which is computationally expensive (since it needs n^2 pair-wise edges to capture a group interaction between n people). Meanwhile, graph-based HABMs are also less suitable for many real-world applications.

For example, for MRSA spread and GRAPH-HETERSIS model, indirect spread pathways (e.g., patient-HCW-patient) also play a significant role in the MRSA spread, as suggested by clinical studies [22]. However, although GRAPH-HETERSIS model can capture the pair-wise interactions well (e.g., from HCW to patients), they still struggle to model such indirect spread pathways.

Algorithm 1: GRAPH-HETERSIS model procedure

1 **Inputs:** $\alpha, \beta, \delta, \{\boldsymbol{A}_t\}_{t=1}^{T}, \{\tau_{ijt}\}_{i,j,t}$
/* Compute pathogen transfer matrix */
2 Compute $\boldsymbol{R}_t(i,j) = \begin{cases} \tau_{ijt}\boldsymbol{A}_t(i,j) & \text{if } i \neq j \\ \tau_{ijt} & \text{if } i = j \end{cases}$
3 **for** $t = 1, ..., T$ **do**
/* Add loads for each node */
4 Update loads $\boldsymbol{l}_{t+1} = \boldsymbol{R}_t\boldsymbol{l}_t + \alpha\boldsymbol{x}_t$
/* Calculate next state for patients */
5 **for** *each patient* i **do**
6 **if** i *is susceptible at time* t *(i.e.,* $\boldsymbol{x}_t(i) = 0$*)* **then**
7 i gets infected (i.e. $\boldsymbol{x}_{t+1}(i) = 1$ with probability $\min\{1, \beta\boldsymbol{l}_t(i)\}$
8 **else**
9 i gets susceptible (i.e. $\boldsymbol{x}_{t+1}(i) = 0$ with prob. δ

4.3 Why Hypergraph-HeterSIS model To overcome these limitations, we propose to use a hypergraph-based HABM. Here is an example to illustrate why hypergraphs can better capture complex group interactions compared to graphs. Consider an example trace data with 3 people (v_1, v_2, v_3) and 2 locations (L_1, L_2) as shown in Figure 1(a). In this example, the vertices are $V = \{v_1, v_2, v_3, L_1, L_2\}$. For the hypergraph representation, vertices v_1 and v_2 are at location L_1 between $t = 10$ to $t = 15$, which can be represented by the hyperedge in yellow. Instead, for the graph representation, we can only connect vertices v_1, v_2, and L_1 as a 3-clique. Here, if we focus on vertex v_2, we can clearly see from the hypergraph that v_2 has interacted with both v_1 and (v_1, v_3) at location L_1 across different times. However, from the graph representation, we cannot determine if v_2 interacted with both v_1 and v_3 together or with each of them separately. This example clearly indicates that hypergraphs are more suitable in capturing the group interactions than graphs.

Although people can also use more fine-grained graphs with much smaller intervals between each timestep, the computational cost also increases significantly, which makes this approach less applicable. Consequently, this motivates us to extend the graph-based HABMs to hypergraph-based HABMs. Next, we will extend GRAPH-HETERSIS introduced in Section 3 to utilize hypergraphs in HYPERGRAPH-HETERSIS.

4.4 Overview of Hypergraph-HeterSIS model Intuitively, the GRAPH-HETERSIS shown in Algorithm 1 contains two steps: (i) accumulate pathogen load from both contacts and/or self-shedding and (ii) stochastically transitioning patients between the susceptible and infected states. Specifically, we do not seek to drastically change GRAPH-HETERSIS but only modify it to include group interactions. Therefore, in HYPERGRAPH-HETERSIS, we seek to maintain these two steps, but only change how we accumulate load, keeping the state transition procedure exactly the same as before.

Instead of modeling the pathogen load spread using pairwise interactions, we aggregate load across all the group interactions for each vertex using the hypergraph matrix $\boldsymbol{H}_t$ at time t. After we aggregate loads within each hyperedge, we discount the accumulated load by a nonlinear yet invertible function $g(\cdot)$ to simulate the lost load between various interactions, which has the same size as $\boldsymbol{l}_t$. Intuitively, this makes sense since if a person moves from one location to another, some load is inevitably shed. Once we have discounted the load obtained by each vertex from each group interaction, we sum up the total load obtained across every group interaction to obtain the updated pathogen load vector. Then, we stochastically state transit patient states like GRAPH-HETERSIS model. We provide more details for HYPERGRAPH-HETERSIS in Supplementary Materials.

4.5 Design A concise summary of the notation can also be seen in Table 1 and the detailed procedure can be seen in Algorithm 2. Specifically, we first calculate the pathogen transfer matrices at each time step based on the various transfer ratios $\boldsymbol{\Upsilon}_{ijt}$ and their corresponding adjacency matrices. Then at each timestep, we sum up the loads transfered via each hyperedge together using the matrix operations, and update the pathogen load in line 4 of Algorithm 2. Specifically, Next, patients stochastically undergo state transitions (*susceptible* to *infected* or *infected* to *susceptible*) respectively.

4.6 Expressivity of Hypergraph-HeterSIS model In this section, we will elaborate on how HYPERGRAPH-HETERSIS model is more expressive than GRAPH-HETERSIS model and why $g(\cdot)$ is necessary. Recall that τ_{ijt} and $\boldsymbol{\Upsilon}_{ijt}$ represent the load transfer ratio from vertex j to i at time t for GRAPH-HETERSIS and HYPERGRAPH-HETERSIS, respectively.

Further assume m_{ijt} to be the number of hyperedges that vertex i and j colocate in at time t.

Algorithm 2: HYPERGRAPH-HETERSIS model procedure

Inputs: $\alpha, \beta, \delta, \{\boldsymbol{H}_t\}_{t=1}^{T}, \{\Upsilon_{ijt}\}_{i,j,t}$

/* Compute pathogen transfer matrix */

Compute $\boldsymbol{\Upsilon}_t(i,j) = \begin{cases} \Upsilon_{ijt}\boldsymbol{A}_t(i,j) & \text{if } i \neq j \\ \Upsilon_{ijt} & \text{if } i = j \end{cases}$

for $t = 1, ..., T$ **do**

  /* Add loads from each hyperedge */

  $\boldsymbol{l}_{t+1} = \alpha \boldsymbol{x}_t + \Sigma_{M_t} \boldsymbol{H}_t \cdot g(\Sigma_N(\boldsymbol{H}_t \circ \boldsymbol{l}_t \otimes \boldsymbol{\Upsilon}_t))$

  /* Calculate next state for patients */

  for *each patient* i **do**

    if i *is susceptible at time* t *(i.e.,* $\boldsymbol{x}_t(i) = 0$*)* **then**

      i gets infected (i.e. $\boldsymbol{x}_{t+1}(i) = 1$ with probability $\min\{1, \beta \boldsymbol{l}_t(i)\}$

    else

      i gets susceptible (i.e. $\boldsymbol{x}_{t+1}(i) = 0$ with prob. δ

THEOREM 4.1. (LINEAR EQUIVALENCE) *If* $g(\cdot)$ *is linear,* GRAPH-HETERSIS *and* HYPERGRAPH-HETERSIS *are equivalent if and only if* $\forall\, i, j \in V$, $\boldsymbol{\Upsilon}_{ijt} \propto \frac{\tau_{ijt}}{m_{ijt}}$.

Proof. Assume that pathogen loads are equal for both GRAPH-HETERSIS and HYPERGRAPH-HETERSIS at time t. Then, inductively solve for parameters (τ and $g(\cdot)$) needed to make pathogen loads at time $t+1$ equal. Note that all the information necessary to make the two models equivalent is known *statically* i.e., without knowing the state of the system and only relying on external information. This is because, even though $\boldsymbol{x}_t$ and $\boldsymbol{l}_t$ are unknown, if we know τ_{ijt} and m_{ijt}, we can reweight the HYPERGRAPH-HETERSIS (and vice versa) to make $\boldsymbol{\Upsilon}_{ijt}$ identical to the GRAPH-HETERSIS. ∎

THEOREM 4.2. (NONLINEAR EQUIVALENCE) *If* $g(\cdot)$ *is nonlinear, then* GRAPH-HETERSIS *and* HYPERGRAPH-HETERSIS *are equivalent if and only if* $g(\cdot)$ *is invertible and* $\forall i, j \in V$ $\boldsymbol{\Upsilon}_{ijt} = g^{-1}(\frac{\tau_{ijt}\boldsymbol{l}_t(j)}{m_{ijt}})$ / $\boldsymbol{l}_t(j)$.

Proof. Assume that pathogen loads are equal for both GRAPH-HETERSIS and HYPERGRAPH-HETERSIS at time t. Although we know τ_{ijt}, m_{ijt} and $g^{-1}(\cdot)$ statically, we do not know $\boldsymbol{l}_t(j)$ until we run our simulation. As this information is hidden, we cannot reweight HYPERGRAPH-HETERSIS to be equivalent to GRAPH-HETERSIS and vice versa. It is easy to see that this is because of $g^{-1}(\cdot)$, which prevents cancellation of $\boldsymbol{l}_t$. ∎

To summarize, if g is linear, we can map HYPERGRAPH-HETERSIS model to GRAPH-HETERSIS model directly with a transformation of the pathogen transfer ratios. However, if $g(\cdot)$ is nonlinear, we cannot. This shows that the nonlinearity of $g(\cdot)$ is what allows HYPERGRAPH-HETERSIS model to be more expressive than GRAPH-HETERSIS model. In this work, we use the $tanh(\cdot)$ function as the nonlinear function $g(\cdot)$.

5 Empirical studies

In this section, we answer the following research questions:

- **Question 1:** Can hypergraphs better capture the differences of structural patterns of group interactions better than graphs in a real-world, clinical hospital setting?
- **Question 2:** Can HYPERGRAPH-HETERSIS better fit the number of MRSA cases (i.e., better calibration) in UVA hospital than GRAPH-HETERSIS?
- **Question 3:** After calibration, can HYPERGRAPH-HETERSIS better forecast the number of MRSA cases in UVA hospital than GRAPH-HETERSIS?
- **Question 4:** Can HYPERGRAPH-HETERSIS provide a good interpretation to help explain MRSA cases in UVA hospital?

5.1 Setup

5.1.1 Dataset We evaluate the performance of graphs and hypergraphs on the UVA dataset. It contains a large suite of clinical metadata from the Epic-based SQL database at the University of Virginia (UVA) hospital and the weekly number of MRSA cases in the hospital. From the dataset, we construct the graphs and hypergraphs for two time periods, one representing 28 weeks before COVID-19 (Jun 18, 2019, to Dec 31, 2019), referred as the UVA-PRECOVID time period, and another representing 28 weeks since the pandemic (Jun 18, 2020, to Dec 31, 2020), referred as the UVA-COVID time period. The interactions are derived from Electronic Healthcare Records (EHRs), which record the time and location of interactions between patients and HCWs. We run experiments on UVA's Rivanna server, which has 595 vertices with over 22598 cores and 8 PB of RAM. On each dataset, we run 100 simulation runs and report the average to evaluate the performance of graphs and hypergraphs. The basic statistics of two time periods are listed in Table 2. We release our code at https://github.com/AdityaLab/H2ABM. We also release an anonymized version of the hypergraphs at https://github.com/AdityaLab/UVA-Hypergraph

Table 2: Summary statistics of UVA-PRECOVID and UVA-COVID

Time periods	UVA-PRECOVID	UVA-COVID
Number of patients	6520	6102
Number of HCWs	6614	6393
Number of Locations	586	579
Number of edges	786767	800278
Number of hyperedges	762135	790060

5.1.2 Calibration Specifically, we use the Ensemble Adjustment Kalman Filter (EAKF) [2] to estimate the parameters in GRAPH-HETERSIS and HYPERGRAPH-HETERSIS model (i.e., α, β, δ, τ_{ijt}, Υ_{ijt} in Algorithm 1 and 2). EAKF has been widely used for heterogeneous agent-based model calibration on various problems, including healthcare-associated infections [16,17] (e.g., MRSA studied in this work), demonstrating high fidelity in recovering model parameters. We calibrate the number of patients in infected state to the weekly number of calibration MRSA cases collected from EHR data in the UVA hospital.

5.1.3 Metrics To better showcase which model leads to a more accurate fit that is well-correlated to real-world MRSA outbreak trends, we use two types of measures to evaluate the performance of GRAPH-HETERSIS and HYPERGRAPH-HETERSIS: Error and trend correlation. For error metrics, following the previous work in HAI modeling [7,8,24,25], we use both Normalized Root Mean Squared Error (NRMSE) and Normal Deviation (ND) as metrics for evaluation. For correlation metrics, we use Pearson correlation following the previous works [24]. Note that for both categories of measures, we use them to evaluate the performance on both calibration and forecasting period. We provide more details for these metrics in the Supplementary Materials.

5.2 Q1: Capturing structural differences In this section, we show that hypergraphs better capture the structural difference than graphs can. Specifically, we compare the contact networks in the University of Virginia (UVA) hospital before and during the first wave of COVID-19 outbreak in Virginia.

We focus on the difference between the contact networks for April 2019 and April 2020. For graphs, we compare basic statistics like the number of entities (patients, HCWs, and locations) and edges. For hypergraphs, we quantify hypergraphs using their fine-grained structural signatures following recent research by [21]. In their work, inspired by *graph motifs*, a combination of repeatable edge connection patterns, they propose a corresponding set of *hypergraph motifs*. Each motif represents all the connectivity patterns possible

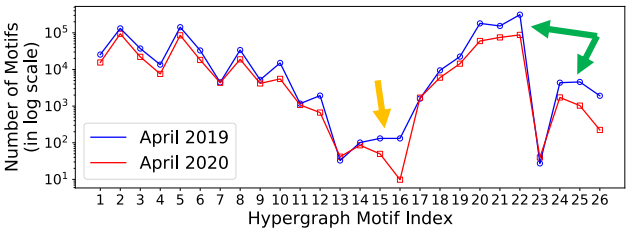

Figure 2: Number of motifs for April 2019 (blue line) and 2020 (red line). The x-axis corresponds to the index of hypergraph motifs, and the y-axis is the daily averaged number of motifs in the log scale. Note that April 2020 corresponds to the first wave of COVID-19 outbreak in Virginia. We see that motifs 15 and 16 (yellow arrow) are significantly lower in 2020 which is explained by the fact that people reduce their interactions with each others. Also note that Motifs 20-22 and 24-26 (green arrows) are also significantly lower due to fewer indirect contacts.

for 3 hyperedges h_1, h_2, h_3 and the emptiness or not of the seven subsets of their overlap, $h_1 \setminus h_2 \setminus h_3$, $h_2 \setminus h_1 \setminus h_3$, $h_3 \setminus h_2 \setminus h_1$, $h_1 \cap h_2 \setminus h_3$, $h_2 \cap h_3 \setminus h_1$, $h_3 \cap h_1 \setminus h_2$ and $h_1 \cap h_2 \cap h_3$. In total, there are 26 possible motifs from all possible combinations of these 7 sets (See Supplementary Materials for a visualization of these). The relative counts for all 26 motifs serve as a signature for a particular type of hypergraph and has been effective in distinguishing different kinds of contact networks, such as co-authorship hypergraphs and email hypergraphs solely based on relative frequencies of the motifs. We use these motifs to analyze the UVA hypergraphs.

In Figure 2 we see that the relative counts of the 26 motifs are vastly different for April 2019 in blue and April 2020 in red. Here, the x-axis corresponds to 26 different kinds of hypergraph motifs, and the y-axis is the daily averaged number of motifs (in log scale) which implies that even small deviation in the plot is highly significant. The reason for choosing these months in particular, is because April 2020 corresponds to the first wave of COVID-19 outbreak in Virginia and the previous April serves as a useful baseline to compare against.

For example, the counts of motifs 15 and 16 (yellow arrow) are significantly lower in 2020 (62.2% and 92.5%) compared to 2019. These motif IDs are those with only one or no empty subsets which implies high interconnectivity between the vertices in h_1, h_2 and h_3. This means that the number of individuals who met in groups was much smaller during COVID than before. This is easily explained by the fear of the pandemic and social distancing guidelines. Additionally, the counts of Motif 20-22 and 24-26 (green arrows) also decrease significantly (66.9%, 51.4%, 71.9% for motif 20-22, and 60.4%, 77.4%, 88.3% for motif 24-26). These motifs cor-

Table 3: Performance of Graph-HeterSIS and Hypergraph-HeterSIS calibration on UVA MRSA cases

Calibration Performance						
	UVA-Precovid			UVA-Covid		
Model	**NRMSE**	**ND**	**Pearson correlation**	**NRMSE**	**ND**	**Pearson correlation**
Hypergraph-HeterSIS	**0.1319**	**0.1159**	**0.4294**	**0.1644**	**0.1494**	**0.3304**
Graph-HeterSIS	0.2512	0.2427	-0.1093	0.3863	0.3380	-0.4381
Forecasting Performance						
	UVA-Precovid			UVA-Covid		
Model	**NRMSE**	**ND**	**Pearson correlation**	**NRMSE**	**ND**	**Pearson correlation**
Hypergraph-HeterSIS	**0.1262**	**0.1091**	**0.2660**	**0.1213**	**0.1061**	**-0.1793**
Graph-HeterSIS	0.4215	0.3983	0.1228	0.4733	0.3910	-0.5355

respond to those with 1-2 empty sets in addition to an empty $h_1 \cap h_2 \cap h_3$. These indicate fewer indirect (e.g., patient-patient-patient) contacts. As shown in later experiments, Hypergraph-HeterSIS captures these structural differences in the contact patterns between UVA-Precovid to UVA-Covid leading to better fits and forecasts. Both models are equivalent if $g(.)$ is linear, warranting the use of nonlinear group load aggregation in Hypergraph-HeterSIS.

Table 4: Basic statistics of graphs for UVA hospital in April 2019 and April 2020

Dataset	April 2019	April 2020
Number of patients	1306	795
Number of HCWs	3998	3685
Number of Locations	467	421
Number of edges	109418	77353

The fact that we are able to identify such complex dynamics using a simple characterization of the connection patterns showcases the effectiveness of hypergraphs in teasing out detailed structural patterns. However, it is hard to observe such dynamics intuitively from the graphs for April 2019 and 2020. We list summary statistics in Table 4. As shown in the Table, we see that the number of patients decreases by around 40% (from 1306 to 795), while the number of HCWs and locations decreases by around 8% and 10% (from 3998 to 3685, and from 421 to 467, respectively). The edges (pair-wise interactions) decrease by around 30% (from 77353 to 109418). However, whether these pair-wise interactions are from group or one-to-one interactions are hard to identify only from the graphs.

5.3 Q2: MRSA cases calibration We have already shown that hypergraphs can capture the structural difference better than graphs can. Next, we show that hypergraph-based HABMs can better fit and forecast MRSA cases in the UVA hospital.

To compare the performance of Graph-HeterSIS and Hypergraph-HeterSIS in calibration, we use the MRSA spread in the UVA hospital. As mentioned in Section 5.1.1, the UVA dataset also includes the weekly number of MRSA cases in the hospital. Here, we follow a real-world prediction scenario that make 4-week ahead predictions every 4 weeks. This means we first calibrate on week 1-4 to predict the MRSA cases in week 5-8, then calibrate on week 1-8 and predict the MRSA cases in week 9-12 and repeat this procedure throughout the whole period.

As shown in Table 3, for both the UVA-Precovid and UVA-Covid dataset, Hypergraph-HeterSIS achieves lower NRMSE and ND in calibration, indicating that Hypergraph-HeterSIS better fits the real-world MRSA cases in UVA hospital. Moreover, we notice that the Pearson correlation for Hypergraph-HeterSIS is significantly higher than Graph-HeterSIS which shows that the Hypergraph-HeterSIS estimated MRSA curves are closely correlated to real-world MRSA outbreak trends.

5.4 Q3: Forecasting MRSA cases We show the performance of 4-weeks ahead predictions in Table 3. As shown in the table, Hypergraph-HeterSIS achieves lower NRMSE, ND and higher forecast Pearson correlation. Note that both NRMSE and ND are normalized metrics, indicating that Hypergraph-HeterSIS better forecasts future MRSA cases in UVA hospital than Graph-HeterSIS can.

Additionally, a recent study [6] reveals that only 11.1% of MRSA colonization (i.e., carry MRSA pathogens) of patients will later develop to infections and be tested. Therefore, if Hypergraph-HeterSIS

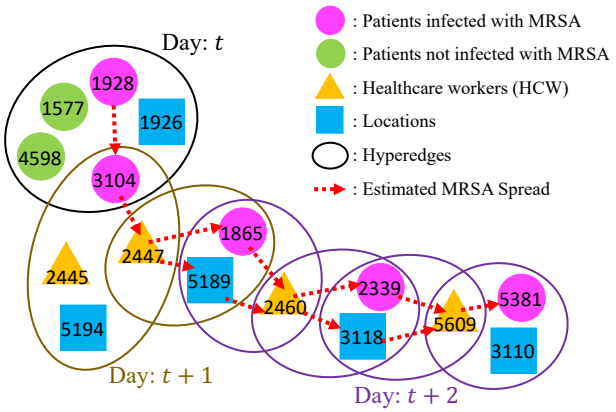

Figure 3: HYPERGRAPH-HETERSIS provides a reasonable interpretation of real-world MRSA spread in the UVA hospital. Pink and green rings, represent infected and non infected patients, yellow triangles represent healthcare workers while blue squares represent locations. Black, brown, and purple lines represent hyperedges on day t, $t+1$, and $t+2$, respectively. From the hypergraph, we can estimate how MRSA spread from Patient 1928 to 3104, 1865, 2339, and 5381 based on the time and connectivity of the hyperedges through the red dash arrows shown in the figure.

can forecast MRSA cases better by even 1 case, we have forecast colonized cases by almost 10 cases better. Therefore, every case counts even if the perceived improvement is marginal.

5.5 Q4: Interpretability We find that HYPERGRAPH-HETERSIS is easily interpretable for MRSA spread. In Figure 3 we showcase a real example, with anonymized IDs, to explain the positive MRSA cases in the UVA hospital over 3 continuous days t to $t+2$ (due to privacy concerns, we cannot release the exact dates of the example, but use t, $t+1$, and $t+2$ instead). Pink and green rings, represent infected and non infected patients, yellow triangles represent healthcare workers while blue squares represent locations. The circles represent the various hyperedges over different days. From Figure 3, we could explain the MRSA infections of 3104, 1865, 2339, and 5381 using the fact that patient 1928 tested positive for MRSA on day t and was indirectly connected to the others via the hyperedges with the estimated spread depicted by the dashed red arrows. This shows that HYPERGRAPH-HETERSIS is effective in capturing group interactions and indirect spread pathways (e.g., HCWs 2447, 2470, and 5609).

6 Conclusion

This work proposes a new class of hypergraph-based HABMs (H^2ABM) by extending the current graph-based heterogeneous agent-based models (HABMs) to better capture the more complex dynamics of group interactions. We show that hypergraphs capture the group interactions better than graphs. We also use a widely used graph-based HABM, GRAPH-HETERSIS as an example to showcase how SIS graph-based HABM can be extended to a H^2ABM (HYPERGRAPH-HETERSIS). Our experiments show that HYPERGRAPH-HETERSIS leads to better interpretability, fitting and forecasting of MRSA cases in the UVA hospital than GRAPH-HETERSIS. Note that we choose a SIS model since it is the most suitable to describe MRSA spread. However, one could extend to other more complex compartment models (e.g., SIR, SIRD, SIRV) to hypergraph-based HABMs for other diseases. Based on our results and other literature (not just epidemics but other contagion phenomena, e.g., smoking, where smokers are known to start and stop in groups [5]), we suggest that any phenomenon that occurs in groups, be modelled as such via hypergraphs. Another interesting followup study would be to use this model to evaluate various interventions to minimize spread such as increased hand washing etc. Additionally, extending the model to account for the several asymptomatic cases, a common occurrence in practice [22], would be a useful direction.

Acknowledgements: This paper was supported in part by the NSF (Expeditions CCF-1918770 and CCF-1918656, CAREER IIS-2028586, RAPID IIS-2027862, Medium IIS-1955883, Medium IIS-2106961, IIS-1931628, IIS-1955797, IIS-2027848, PIPP CCF-2200269), NIH 2R01GM109718, CDC MInD program U01CK000589, ORNL and funds/computing resources from Georgia Tech and GTRI.

References

[1] B. Adhikari, B. Lewis, A. Vullikanti, J. M. Jiménez, and B. A. Prakash. Fast and near-optimal monitoring for healthcare acquired infection outbreaks. *PLoS computational biology*, 15(9):e1007284, 2019.

[2] J. L. Anderson. An ensemble adjustment kalman filter for data assimilation. *Monthly weather review*, 129(12):2884–2903, 2001.

[3] A. Bhalla, N. J. Pultz, D. M. Gries, A. J. Ray, E. C. Eckstein, D. C. Aron, and C. J. Donskey. Acquisition of nosocomial pathogens on hands after contact with environmental surfaces near hospitalized patients. *ICHE*, 25(2):164–167, 2004.

[4] K. Bouafia and B. Molnár. Hypergraph application on business process performance. *Information*, 12(9):370, 2021.

[5] N. A. Christakis and J. H. Fowler. The collective dynamics of smoking in a large social network. *New England journal of medicine*, 358(21):2249–2258, 2008.

[6] R. Coello, J. Glynn, C. Gaspar, J. Picazo, and J. Fereres. Risk factors for developing clinical in-

fection with methicillin-resistant staphylococcus aureus (mrsa) amongst hospital patients initially only colonized with mrsa. *Journal of Hospital Infection*, 37(1):39–46, 1997.

[7] E. Y. Cramer, Y. Huang, Y. Wang, E. L. Ray, M. Cornell, J. Bracher, A. Brennen, A. J. C. Rivadeneira, A. Gerding, K. House, et al. The united states covid-19 forecast hub dataset. *Scientific data*, 9(1):462, 2022.

[8] E. Y. Cramer, E. L. Ray, V. K. Lopez, J. Bracher, A. Brennen, A. J. Castro Rivadeneira, A. Gerding, T. Gneiting, K. H. House, Y. Huang, et al. Evaluation of individual and ensemble probabilistic forecasts of covid-19 mortality in the united states. *PNAS*, 119(15):e2113561119, 2022.

[9] J. Cui, S. Cho, M. Kamruzzaman, M. Bielskas, A. Vullikanti, and B. A. Prakash. Using spectral characterization to identify healthcare-associated infection (hai) patients for clinical contact precaution. *Scientific Reports*, 13(1):16197, 2023.

[10] J. Cui, J. Heavey, L. Lin, , E. Klein, G. Madden, C. Sifri, L. Bryan, A. Vullikanti, and B. A. Prakash. Modeling relaxed policies for discontinuation of methicillin resistant staphylococcus aureus contact precautions. *ICHE*, 2024.

[11] J. Cui, H. Wu, L. Fu, and X. Gan. De-anonymizing bitcoin networks: An ip matching method via heuristic approach: Poster. In *Proceedings of the ACM Turing Celebration Conference-China*, pages 1–2, 2019.

[12] J. Fernández-Gracia, J.-P. Onnela, M. L. Barnett, V. M. Eguíluz, and N. A. Christakis. Influence of a patient transfer network of us inpatient facilities on the incidence of nosocomial infections. *Scientific reports*, 7(1):2930, 2017.

[13] J. Heavey, J. Cui, C. Chen, B. A. Prakash, and A. Vullikanti. Provable sensor sets for epidemic detection over networks with minimum delay. In *AAAI*, volume 36, pages 10202–10209, 2022.

[14] D. Hong, R. Dey, X. Lin, B. Cleary, and E. Dobriban. Group testing via hypergraph factorization applied to covid-19. *Nature communications*, 13(1):1837, 2022.

[15] B. Hu, C. Shi, W. X. Zhao, and P. S. Yu. Leveraging meta-path based context for top-n recommendation with a neural co-attention model. In *KDD*, pages 1531–1540, 2018.

[16] H. Jang, A. Fu, J. Cui, M. Kamruzzaman, B. A. Prakash, A. Vullikanti, B. Adhikari, and S. V. Pemmaraju. Detecting sources of healthcare associated infections. In *AAAI*, 2023.

[17] H. Jang, S. Justice, P. M. Polgreen, A. M. Segre, D. K. Sewell, and S. V. Pemmaraju. Evaluating architectural changes to alter pathogen dynamics in a dialysis unit. In *ASONAM*, pages 961–968. IEEE, 2019.

[18] U. Jeong, K. Ding, L. Cheng, R. Guo, K. Shu, and H. Liu. Nothing stands alone: Relational fake news detection with hypergraph neural networks. In *2022 IEEE International Conference on Big Data (Big Data)*, pages 596–605. IEEE, 2022.

[19] A. J. Kallen, Y. Mu, S. Bulens, A. Reingold, S. Petit, K. Gershman, S. M. Ray, L. H. Harrison, R. Lynfield, G. Dumyati, et al. Health care–associated invasive mrsa infections, 2005-2008. *Jama*, 304(6):641–647, 2010.

[20] L. Kong, J. Cui, H. Sun, Y. Zhuang, B. A. Prakash, and C. Zhang. Autoregressive diffusion model for graph generation. In *ICML*, 2023.

[21] G. Lee, J. Yoo, and K. Shin. Mining of real-world hypergraphs: Patterns, tools, and generators. In *CIKM*, pages 5144–5147, 2022.

[22] S. Pei, F. Liljeros, and J. Shaman. Identifying asymptomatic spreaders of antimicrobial-resistant pathogens in hospital settings. *PNAS*, 118(37):e2111190118, 2021.

[23] L. E. Rocha, V. Singh, M. Esch, T. Lenaerts, F. Liljeros, and A. Thorson. Dynamic contact networks of patients and mrsa spread in hospitals. *Scientific reports*, 10(1):1–10, 2020.

[24] A. Rodríguez, J. Cui, N. Ramakrishnan, B. Adhikari, and B. A. Prakash. Einns: epidemiologically-informed neural networks. In *AAAI*, volume 37, pages 14453–14460, 2023.

[25] A. Rodriguez, A. Tabassum, J. Cui, J. Xie, J. Ho, P. Agarwal, B. Adhikari, and B. A. Prakash. Deepcovid: An operational deep learning-driven framework for explainable real-time covid-19 forecasting. In *AAAI*, volume 35, pages 15393–15400, 2021.

[26] P. Roy, S. Sarkar, S. Biswas, F. Chen, Z. Chen, N. Ramakrishnan, and C.-T. Lu. Deep diffusion-based forecasting of covid-19 by incorporating network-level mobility information. In *ASONAM*, pages 168–175, 2021.

[27] G. M. Snyder, K. A. Thorn, J. P. Furuno, E. N. Perencevich, M.-C. Roghmann, S. M. Strauss, G. Netzer, and A. D. Harris. Detection of methicillin-resistant staphylococcus aureus and vancomycin-resistant enterococci on the gowns and gloves of healthcare workers. *ICHE*, 29(7):583–589, 2008.

[28] Y. Sun, J. Han, X. Yan, P. S. Yu, and T. Wu. Pathsim: Meta path-based top-k similarity search in heterogeneous information networks. *Proceedings of the VLDB Endowment*, 4(11):992–1003, 2011.

[29] J. Wang, K. Ding, Z. Zhu, and J. Caverlee. Session-based recommendation with hypergraph attention networks. In *SDM*, pages 82–90. SIAM, 2021.

[30] M. M. Wolf, A. M. Klinvex, and D. M. Dunlavy. Advantages to modeling relational data using hypergraphs versus graphs. In *HPEC*, pages 1–7. IEEE, 2016.

[31] F. Zhang, M. Wang, J. Xi, J. Yang, and A. Li. A novel heterogeneous network-based method for drug response prediction in cancer cell lines. *Scientific reports*, 8(1):1–9, 2018.