# Spreader events and the limitations of projected networks for capturing dynamics on multipartite networks

Hyojun A. Lee [1,*] Luiz G. A. Alves [1,*] and Luís A. Nunes Amaral [1,2,3,†]

[1]*Department of Chemical and Biological Engineering, Northwestern University, Evanston, Illinois 60208, USA*
[2]*Department of Physics and Astronomy, Northwestern University, Evanston, Illinois 60208-3112, USA*
[3]*Northwestern Institute on Complex Systems, Northwestern University, Evanston, Illinois 60208-4057, USA*

Many systems of scientific interest can be conceptualized as multipartite networks. Examples include the spread of sexually transmitted infections, scientific collaborations, human friendships, product recommendation systems, and metabolic networks. In practice, these systems are often studied after projection onto a single class of nodes, losing crucial information. Here, we address a significant knowledge gap by comparing transmission dynamics on temporal multipartite networks and on their time-aggregated unipartite projections to determine the impact of the lost information on our ability to predict the systems' dynamics. We show that the dynamics of transmission models can be dramatically dissimilar on multipartite networks and on their projections at three levels: final outcome, the magnitude of the variability from realization to realization, and overall shape of the temporal trajectory. We find that the ratio of the number of nodes to the number of active edges over the time-aggregation scale determines the ability of projected networks to capture the dynamics on the multipartite network. Finally, we explore which properties of a multipartite network are crucial in generating synthetic networks that better reproduce the dynamical behavior observed in real multipartite networks.

## I. INTRODUCTION

The theoretical study of transmission processes, whether infectious diseases, innovations, or mores, has typically followed one of two approaches: compartment models [1] and network models [2–4]. In compartment models, individuals belong to a small number of compartments, and they transit in and out of compartments according to specific rates. Compartment models are mean-field models in which all members of a compartment are equally likely to interact with members of another compartment. In network models, individuals interact pairwise according to the connections defined in the underlying social network. Both approaches have produced significant advances both separately and when integrated as can be seen from the studies and forecasting of the current SARS-CoV-2 pandemic and of other recent pandemics [5,6].

Research published in the past two decades has demonstrated that network representations provide a fruitful abstraction of many complex systems [7,8]. Despite the remarkable advances, the task of determining the simplest network representation able to fully capture the characteristics of the system remains more of an art than a science. Indeed,

although limitations in data collection procedures and less developed theoretical frameworks lead many scholars to settle for static, unipartite, unweighted, and undirected network representations, recent research has demonstrated that such representations may be inadequate [9–12].

For example, in the context of human mobility or international trade, one cannot disregard the weights (i.e., number of passengers in transportation networks, or exchange volume in international trade flows) of the connections between system components [13,14]. In cases such as these, the system is more accurately conceptualized through *weighted* networks [9]. Similarly, in the context, for example, of the spread of information within social systems, one cannot ignore the multiple media—email, Twitter, WhatsApp, and so on—that are used by individuals to communicate. In cases such as these, the system is more accurately conceptualized through *multilayer* networks [11]. As a result, in recent years there has been an important and thriving research stream focused on multiplexed systems [12,15].

Another significant characteristic of many nonequilibrium complex systems is that interactions among components are not immutable [16–19]. For example, direct flight connections to a city can be added or dropped over time; friendships flourish and wane; trophic relations in an ecosystem change with the seasons. In cases such as these, the system is more accurately conceptualized through *temporal* networks [11,20]. The temporal aggregation of evolving networks can significantly affect processes, such as transmission phenomena [20,21]. As a result, a significant research thrust has focused on the study of temporal networks and its implications to dynamical processes [22–33].
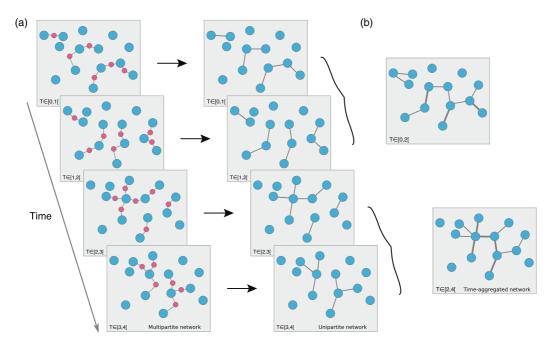
---

FIG. 1. Temporal multipartite network, unipartite projections, and time-aggregated networks. (a) Temporal multipartite network (left) composed of a class of agents nodes (larger blue circles) and a class of event nodes (pink circles) and its temporal agent projection (right). Event nodes have specific creation times and finite temporal duration (such as a day or a week). (b) Time-aggregated unipartite projection over the agent nodes shows simpler structures and dynamics. When the network is projected onto the agent network, nodes with temporal properties can be aggregated over different periods (a week, a month, and so on), losing crucial information on the ordering of events.

In contrast to the deserved attention provided by these important and flourishing research streams to weighted, multilayer, and temporal networks, multipartite networks remain overlooked (see Ref. [34] for exceptions and Refs. [35–39] for epidemics on bipartite networks). However, many important biological and social systems are most naturally conceptualized as multipartite networks. Examples include cellular metabolism [40], ecosystems [41], collaborations [42], heterosexual romantic relationships [3,43,44], research collaborations [45], or product recommendation systems [46–48]. These systems comprise multiple classes of components—for example, movies, actors, directors, producers, agents, studios, and so on in movie-production networks—and edges can only connect nodes of different types—producers connect through movies or through studios.

Despite the diversity of ways complex systems can be represented and projected onto complex networks, our understanding of how modeling choices affect the dynamical process remains limited. Presumably because of the difficulty in obtaining detailed data, or for simplicity of analysis [49]. In particular, multipartite networks are typically studied after projection onto a single type of node considered to be the most important, such as actors in movie-production networks. However, by construction, the unipartite projection of a multipartite network is dramatically less information rich than the real network. For this reason, there is growing interest and effort in the community to collect open-access, large-scale, and information-rich datasets on different network projections with time dependence to better understand real networks (see, for instance, Ref. [50]).

A particularly important aspect of some multipartite networks is the presence of strong temporal discontinuities for, at least, one type of node. These nodes present discrete timestamps and confined temporal existence: Movies have production schedules, and romantic relationships span finite periods [Fig. 1(a)]. For this reason, these temporal multipartite networks have a stronger and more complex temporal structure than the typical temporal network, and detailed temporal information is frequently unavailable [Fig. 1(b)]. Thus, it remains unclear whether the approach of temporal fine-graining used in the context of temporal networks to obtain static snapshots [51] would be sufficient to generate an accurate conceptualization of multipartite networks that include a class of nodes with heterogeneous and discontinuous event dynamics.

To address this question, we investigate the extent to which unipartite projections of a multipartite network can be used to accurately describe and predict transmission dynamics [12,52–55] taking place within multipartite networks in which one class of nodes has discontinuous dynamics. As we mentioned earlier, we focus on transmission dynamics because they have been used to model a broad class of important phenomena, such as epidemic outbreaks, opinion formation, and the spread of innovations [33,56,57]. Because of the broad range of applications of interest, we will consider model predictions for the full range of infectability values from the low infection rate of viruses, such as HIV [44] to the high infection rate of "unavoidable" memes [58].

We organize our analysis around two systems with very different sizes and characteristics—but for which we have high-quality data—and that have been studied in the context of the adoption of innovations among small teams of highly trained professions [59] and in the context of the assembly of creative teams [60]. We simulate transmission dynamics on

these networks and on synthetic networks derived from them using two well-known transmission models—susceptible infectious (SI) and general contagion (GC) [61–63].

Significantly, we find under real-world conditions that the time-aggregated unipartite projections are broadly unable to predict the dynamics observed on the full temporal multipartite networks. Using model networks, we show that when the ratio of number of agents to the number of events/teams is less than one, unipartite projections will fail to reproduce the final number of infected agents, the magnitude of the variability from different realizations, and the overall shape of the temporal trajectories observed on multipartite networks. Furthermore, to address the cases in which unipartite projections are inadequate, we investigate the minimal amount of information that must be retained from a multipartite network in order to generate synthetic multipartite networks that properly reproduce the dynamics observed in the real networks.

## II. EMPIRICAL MULTIPARTITE NETWORKS AND THEIR PROJECTIONS

A multipartite network is a network whose nodes are partitioned into $K$ distinct nonoverlapping sets. The nodes in each set have edges can connect only to nodes in the other $K - 1$ sets [Fig. 1(a)]. We consider two contexts where connectivity is well described by multipartite networks, physician coverage teams in a hospital, and movie-producing teams.

For the former, we consider the multipartite networks formed by the physicians providing high-intensity critical care coverage in the medical intensive care unit (MICU) of a Chicago hospital that we studied previously [59]. Specifically, we obtained the MICU coverage schedules of 76 physicians from July 2012 to August 2016. Each team is composed of one attending and one fellow and has confined temporal existence (see the Supplemental Material [64] and Ref. [59] for details). We construct a temporal multipartite network for each academic year composed of three sets of nodes: coverage teams, attendings, and fellows. Attendings and fellows are only connected to team nodes. Edges between nodes are created when a fellow and an attending are together in the same team. Since the team has finite temporal existence, the edges exist only while the teams are active, typically a week.

For the latter, we construct a network of producer collaborations on U.S. movies, that we obtained by crawling the Internet Movie Database [65]. We focus here on production teams for the 2009 U.S.-produced movies released in the period 1990–2000. We identify 5758 active producers during the considered period. Unlike the physician teams, we lack information on the exact production period for each movie.

Because the temporal aggregation timescale is known to affect static network structure [20,21,30], we compare the temporal multipartite physician network with different unipartite projected networks aggregated over different periods of time, spanning from 7 days to 12 months. In the unipartite projected networks, the edges are static during the respective time period and, therefore, lose information about the temporal ordering of interactions [47]. In addition to the unweighted unipartite network projection, we consider two weighting approaches for defining edge weight: the number of shared teams and the number of days in shared teams. For the movie producers, we aggregate time over four timescales 1, 2, 5, and 10 yrs. We model the edges as unweighted for the producer projected network because the vast majority of producers participate in only one movie per year.

## III. TRANSMISSION DYNAMICS

Here, we focus on transmission dynamics and use the language of infectious diseases to investigate the limitations of time-aggregated unipartite projected networks for capturing the dynamics of temporal multipartite networks. The physicians' network is a tripartite network composed of two classes of agent nodes (attendings and fellows) and one class of event nodes (teams). At time zero, two agent nodes are randomly chosen to become infectious (for the multipartite network, we chose the two nodes that are connected to the first event node), whereas all other nodes are deemed susceptible and to have no prior exposure (see Methods for details). We then track the number of infectious individuals over time.

Figure 2(a) shows the mean number of infectious agents over time for two different parameter values for the SI model on multipartite and projected networks. It is visually apparent that the dynamics on the projected networks yield different dynamics from those observed on the multipartite network [Fig. 2(a)].

We further investigate the dynamic differences using the GC model to compare the temporal trajectories of the four multipartite networks. Our benchmark multipartite networks (i.e., the networks for the different academic years) show similar temporal trajectories for a wide range of parameters, $0.1 < d < 1$ and $0.1 < p < 1$; Fig. S1 of the Supplemental Material [64]. In contrast, we can observe that the dynamics of time-aggregated unipartite projected networks compared with our benchmark multipartite network can overestimate the number of infectious agents over the course of time for a wide range of parameters for asynchronous dynamics simulations (see Figs. S2–S4 in the Supplemental Material [64]), and it can overestimate or underestimate the number of infected agents depending on the time-aggregation interval and parameter range for synchronous dynamics simulations (see Figs. S5–S7 in the Supplemental Material [64] for synchronous dynamics). We will provide a quantification of these differences in the following sections.

Transmission dynamics on projected networks also lack the trajectory "diversity" displayed by the dynamics on real networks [compare 95% confidence intervals in Fig. 2(a)], pointing at deep differences of the dynamics.

The producers' network is a bipartite network where there are one class of agent nodes (producers) and one class of event nodes (movies). Using the SI model to simulate the transmission process on temporal multipartite networks and compare with their projections, we find that the movie producers' network exhibit less variable temporal trajectories than the physicians' networks [Fig. 2(b)]. Nonetheless, it is visually apparent that the dynamics on the projected movie-producers' networks yield different dynamics from those observed on the multipartite network. Transmission dynamics taking place on projected networks consistently overestimate the number of
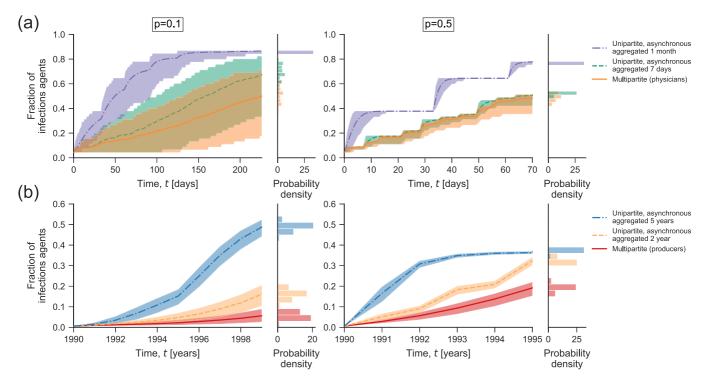
FIG. 2. Spreading dynamics on temporal multipartite networks and on their time-aggregated unipartite projections. (a) Mean fractions of infectious agents (full lines) and 95% confidence interval for the dynamics (shaded regions) on the physicians' multipartite networks and unweighted unipartite projected networks using 7 days and 1 month temporal aggregation (initial infectious agents: 2). The results for the projected network are obtained for the asynchronous status update. (b) Mean fractions of infectious agents (full lines) and 95% confidence interval for the dynamics (shaded regions) on the producers' multipartite networks and unweighted unipartite projected networks using 2 and 5 year temporal aggregation (initial infectious agents: 5% of producers appearing in the first year). The results for projected networks are obtained for the asynchronous status update. The right side of each panel shows the distributions of the final fraction of infectious agents for each set of parameters and conditions.

infected agents over the course of time. These results are less striking than those obtained for the physicians' network but still consistent.

### A. Factors controlling stability of unipartite projections

The movie producers' network and physicians' network differ significantly in terms of the number of agents, the team sizes, the aggregation timescale, and the number of teams per aggregation cycle. Thus, we investigate which of these factors can explain the ability of projected networks to make better predictions. Additionally, we study a simple bipartite network model with community structure [66,67] in order to account for the possibility of modular structure in real-world multipartite networks (see Methods).

We find that, as the degree of temporal fine graining increases, the transmission dynamics of unipartite projected networks become more similar to those of temporal bipartite networks (Fig. 3). This is to be expected because the smaller the temporal graining is, the more the projected networks become an accurate snapshot of the bipartite networks.

We also find that the number of agent $N$ increases whereas the number of teams per aggregation cycle $T_c$ remain fixed and the difference between the transmission dynamics of the bipartite networks and their projected networks also increases (Fig. 3).

Our analysis, thus, reveals that the determining factor is the ratio of the number of agents to the number of teams per aggregation cycle. It is important to note, however, that the trajectory of the transmission dynamics for the projected and the bipartite networks have distinct functional forms. Thus, even the temporal fine-grained networks used in the context of temporal networks to obtain static snapshots [51] would not be sufficient to generate an accurate conceptualization of temporal multipartite networks that include a class of nodes with discontinuous dynamics.

### IV. BUILDING REALISTIC MULTIPARTITE NETWORKS

Since simulation on unipartite projections cannot, for many conditions, replicate the transmission dynamics observed on multipartite networks, we ask whether there are suitable null models such that the dynamics on the generated networks better replicate the dynamics on real temporal multipartite networks. Specifically, we aim to identify the crucial information needed for creating adequate synthetic multipartite networks, and we focus here on the physicians' networks because they show the largest dynamical differences between projected and multipartite networks.

In building our multipartite network models, we focus on three types of aggregate information: the number of agents in each agent class of nodes; distribution of the number of teams
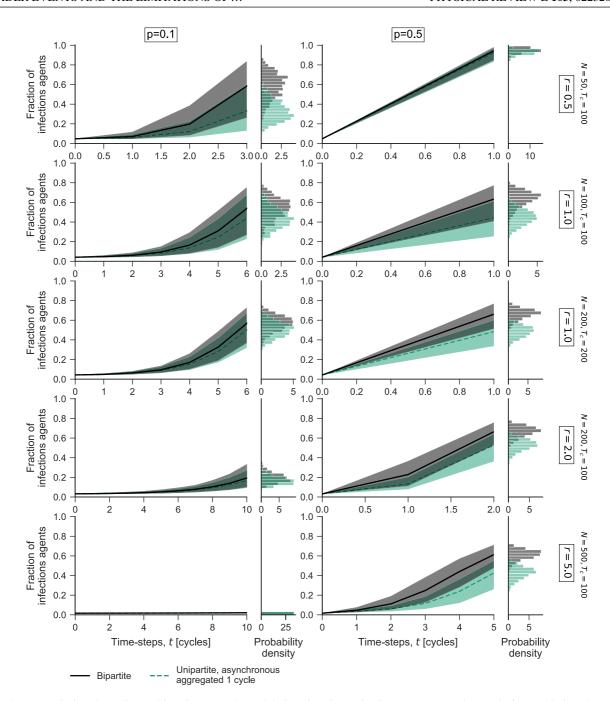
FIG. 3. Transmission dynamics on bipartite networks and their unipartite projections. Average and spread of transmission dynamics for the GC model using infection probabilities $p = 0.1$ (left column) and $p = 0.5$ (right column) and dose $d = 0.5$. Note that as the ratio agents and teams (i.e., $r \equiv \frac{N}{T_c}$) increases, the difference in the dynamics becomes less pronounced.

that an agent participates in; and distribution of gap durations between consecutive participations in a team by an agent. We construct three null models by using increasing amounts of information about the real network [Fig. 4(a)]. The first generative model, denoted **N**, makes use only of the number of agents $N$ in the system. The second model, denoted **NT**, adds also information on the distribution of the number of teams $T$ in which agents participate [Fig. 4(b) top]. The third model, denoted **NTG**, adds information on the distribution of gap $G$ durations between participation in consecutive teams as well [Fig. 4(b) bottom].

Figure 4(c) shows the level of information that is used in each model and the steps for generating a synthetic multipartite network. We show the distributions of the number of teams and gap durations for the three types of synthetic networks and the real networks in Fig. S8 of the Supplemental Material [64].

### A. Quantification of the similarity of two ensembles of dynamical trajectories

Next, we address how to quantify the differences in the outcomes between the transmission dynamics on the synthetic
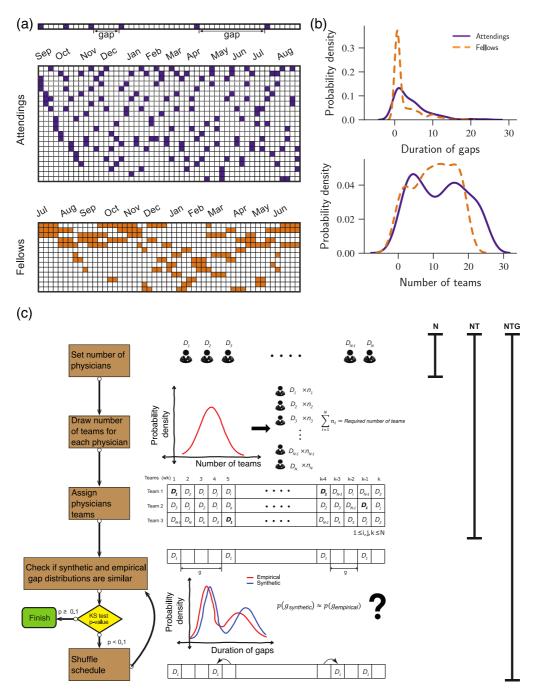
FIG. 4. Generation of synthetic multipartite networks. (a) One year schedule for attendings and fellows in the MICU. Each row displays the entire year schedule for a given physician. Each square represents 1 week, and filled squares indicate participation in a coverage team by a given physician. Consecutive white squares show gap durations between participation in coverage teams for each physician. (b) Distribution of the number of coverage teams a physician is part of during a year and distribution of the duration of gaps between participation in consecutive coverage teams. We first note that whereas coverage teams are reportedly assigned randomly by the hospital based on the number of teams that each physician has to be assigned to, we find that the distribution of the number of teams in which physicians participate does not obey a Poisson distribution [68]. Therefore, physicians' team assignments cannot be purely random. The distributions of gap durations are also not exponential, which is the expected distribution for interevent times that occur at a constant rate [69]. These two findings suggest that there are other constraints that govern how physicians are assigned to teams. (c) Algorithm for generating synthetic multipartite networks according to levels of information incorporated. We create $N$ physician nodes as in the multipartite network we are trying to replicate (the **N** model). We can then either draw the number of coverage teams assigned to a physician from a Poisson distribution or from the distribution of the number of teams in the multipartite network without replacement (the **NT** model). Note that the total number of a team must be equal to the number in the multipartite network. We assign physicians' team participation at random. If we are including information about the distribution of gap durations, then we calculate the gap duration distribution for the synthetic network and compare it to that of the multipartite network. The teams are adjusted to match the gap duration distribution until the difference between the synthetic network and the multipartite network is not statistically significant (the **NTG** model).

networks and that on the real networks. To this end, we define a new metric, the dynamic overlap of two ensembles of dynamical trajectories, which we denote as $\Omega$. We focus on the GC model since the SI model is recovered for $D = d$. We compare both the final number of infectious nodes as well as the similarity of their trajectories when using the same parameters on either synthetic or empirical networks. Quantifying such a comparison is not trivial, and different methods have been introduced to compare the outcomes of transmission dynamics in networks [70], however, there has been less focus on comparing the transmission trajectories over the entire period. Inspired by the weak Fréchet distance [71], we first define the distance between two dynamic trajectories as the area between the curves for an observation period $T$,

$$A_j^i(\mathcal{N}_1, \mathcal{N}_2) \equiv \int_0^T |f_i(t|\mathcal{N}_1) - f_j(t|\mathcal{N}_2)|dt, \qquad (1)$$

where $f_i(t|\mathcal{N}_1)$ is the $i$th realization of the dynamics on network $\mathcal{N}_1$. We control for the potentially different number of agents in each network by scaling the trajectories by the number of agents in the network. We demonstrate how individual realizations of the dynamics on the same network can yield quite distinct trajectories in the top panel of Fig. 5(a). Thus, our aim must be to quantify whether a particular trajectory on a network falls within the expected range for trajectories obtained for simulations of the dynamics on a set of reference networks, $\mathcal{R} = \{\mathcal{R}_1, \mathcal{R}_2, \ldots\}$. To this end, we define the distance of trajectory $i$ to the mean trajectory $R$ on a set of reference networks as

$$A_R^i(\mathcal{N}_k, \mathcal{R}) \equiv \int_0^T |f_i(t|\mathcal{N}_k) - \overline{f_R(t|\mathcal{R})}|dt , \qquad (2)$$

where $\overline{f_R(t|\mathcal{R})}$ represents an average over realizations of the simulations and over the set of networks in $\mathcal{R}$ [middle panel, Fig. 5(a)]. By calculating $A_R^i(\mathcal{N}_k, \mathcal{R})$ for the $i$th trajectory, we can estimate the percentiles for the distances to the mean trajectory, $P_{2.5}\{A_R^i(\mathcal{N}_k, \mathcal{R})\}$ and $P_{97.5}\{A_R^i(\mathcal{N}_k, \mathcal{R})\}$, that encompass 95% of the dynamical trajectories generated on the set of reference networks [bottom panel, Fig. 5(a)]. We define the dynamic overlap $\omega$ for a given network $\mathcal{N}_k$ as

$$\omega(\mathcal{N}_k) \equiv \int_{P_{2.5}}^{P_{97.5}} A_R \, p(A_R|\mathcal{N}_k) dA_R, \qquad (3)$$

where $p(A_R|\mathcal{N}_k)$ is the probability of observing a specific value of $A_R^i(\mathcal{N}_k, \mathcal{R})$. The dynamic overlap for a set of multipartite networks is $\Omega \equiv P_{50}\{\omega(\mathcal{N}_k)\}$, which takes values in [0,1]. In practice, we define $\mathcal{R}$ for multipartite networks as the set of networks for 2012, 2013, and 2015 and use the real network for 2014 as a control for estimating the similarity of the trajectories for different parameter values. We chose $\mathcal{R}_{2014}$ as the control because its number of physicians takes a value close to the mean for the four networks. We then compare the trajectories of the GC model for the control network with results obtained for synthetic multipartite network generated using information from $\mathcal{R}_{2014}$ [Fig. 5(b)].

We show $\Omega$ for the control multipartite network $\mathcal{R}_{2014}$—multipartite network from the hospital for the year 2014 in Fig. 5(c). Similar results are obtained for $\Omega$ when using $\mathcal{R}_{2012}$, $\mathcal{R}_{2013}$, and $\mathcal{R}_{2015}$ as control networks (see

Figs. S9–S11 in the Supplemental Material [64]). Even though $\mathcal{R}_{2014}$ is not structurally different from the networks in $\mathcal{R}$, we still found that $\Omega$ is close to zero when $p$ is large. This is due to the fact that the stochasticity of the transmission dynamics becomes negligible when almost all interactions lead to successful transmission. Under these conditions, even small differences in the network lead to statistically significant differences in the dynamics (Fig. S1 in the Supplemental Material [64]).

In contrast, for the projected networks, $\Omega$ is generally close to zero for nearly all regions of parameter space [Fig. 5(c), and Figs. S12–S15 in the Supplemental Material [64]]. The exceptions are those parameter values for which the variance in the trajectories for the multipartite networks is very large. Because the trajectory diversity for the dynamics on the projected networks are usually small (Fig. 2), it enables all the trajectories on the projected networks to fall within the 95% confidence interval of the dynamics on the multipartite networks (see the case of $p = 0.1$ in Fig. 2 and Figs. S1–S6 in the Supplemental Material [64]). This interpretation is supported by the fact that when we take the time-aggregated unipartite networks as the reference set and the multipartite networks as the ones being compared, then $\Omega$ drops to zero (Figs. S16 and S17 in the Supplemental Material [64]). If the transmission dynamics were truly similar, $\Omega$ would take similar values even when $\mathcal{R}$ and $\{\mathcal{N}_k\}$ are switched so that $\{\mathcal{N}_k\}$ are used as reference networks as indeed is the case for synthetic multipartite networks (Fig. S18 in the Supplemental Material [64]).

Thus, we find that as we increase the real-world information built into the null models, the more closely the infection trajectories tend to mirror those observed for the control network [Fig. 6(a)]. In fact, calculating the overlap $\Omega$ of the transmission dynamics in synthetic multipartite networks with the reference set of network $\mathcal{R}$ for different values of the parameter of the GC model, there is a strong agreement between the transmission trajectories and the control network for a wide range of parameters (Fig. 6(b), and Figs. S9–S11 in the Supplemental Material [64] for comparisons using other years than 2014 as control network). Thus, even simplistic synthetic networks can produce $\Omega$ values that are quite similar to the values obtained for the control network and can capture real dynamic much better than the projected networks (Fig. S19 in the Supplemental Material [64]).

Furthermore, to verify that this model is generalizable to other cases, we generate synthetic multipartite networks using the characteristics of the producers' network. We randomly selected 5% of the producers active in the first year as initial infectious agents. We found that the transmission dynamics on even the simplest synthetic network are much more similar to the transmission dynamics on the original multipartite network than on projected networks which is consistent with the results from the physicians' network. However, we observed that the infection rates are not as similar to the real multipartite network compared to the physicians' case, although the overall shape of the infection curves is quite similar (Fig. S20 of the Supplemental Material [64]). This difference may be due to the fact that unlike for the physicians' networks, producer networks are more heterogeneous. An alternative or even complementary explanation is that physician and producer teams have different formation mechanisms. In the
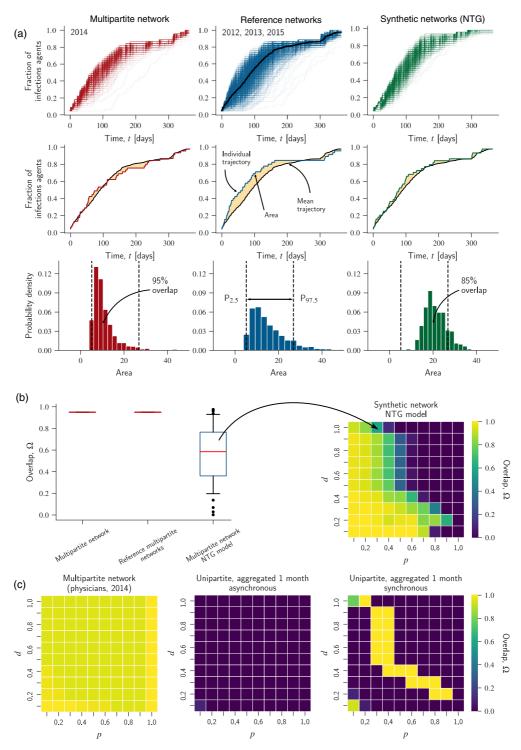
FIG. 5. Systematic quantification of the similarity of transmission dynamics on different networks. (a) The top panel shows the simulation of 1000 independent stochastic realizations of the GC model dynamics for each network: the 2014 control temporal multipartite network (red), the reference set of multipartite networks for 2012, 2013, and 2015 (blue), and 25 synthetic networks (green). The middle panel shows the areas between the mean of realizations from the reference networks (reference curve, i.e., thick black line in the top panel) and each realization from the reference networks. The bottom panel shows the area distribution $p(A_R|\mathcal{N}_k)$ where we identify the 95% confidence boundaries for the observed overlapping. The percentages indicate the overlapping between the areas from the simulations compared with those from the reference set. (b) The panel shows the box plot of the area overlapping for the control multipartite network, the reference set, and the synthetic network for a specific set of parameters as indicated in the figure. We calculate $\omega(\mathcal{N}_k)$ for each set of networks in multipartite networks and synthetic networks. Because there are 25 synthetic networks, we define $\Omega$ as the median of $\omega(\mathcal{N}_k)$ for the 25 networks. (c) The heat maps show the dynamic range overlap $\Omega$ for the entire parameter space for different networks. Each cell shows the $\Omega$ value of the GC model dynamics compared with those obtained on the reference set $\mathcal{R}$ for the multipartite network (left), for projected networks using synchronous (middle), and asynchronous (right) update.
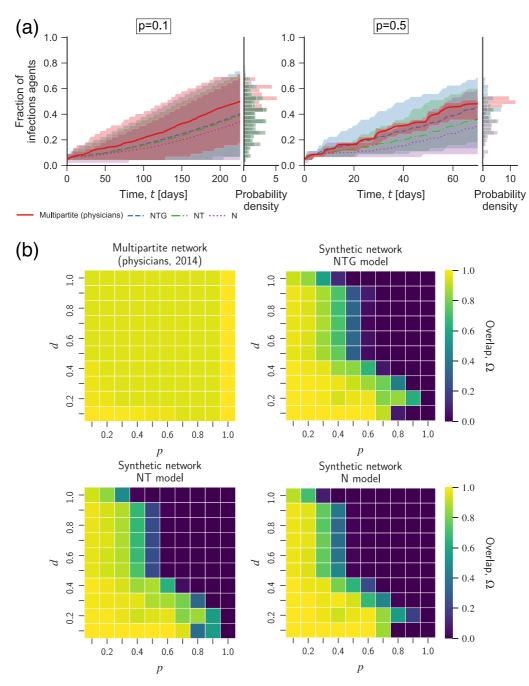
FIG. 6. Model dynamics on multipartite networks and on their synthetic networks. (a) Mean fraction of infectious agents over time (lines) and 95% confidence interval for the dynamics (shaded regions) on the multipartite and synthetic multipartite networks using SI model with infection probabilities $p = 0.1$ and $p = 0.5$. The histograms show the final fraction of infectious agents for each network. (b) The heat maps show the dynamic range overlap $\Omega$ for the entire parameter space for different networks. Each cell shows the $\Omega$ value of the GC model dynamics compared with those obtained on the reference set $\mathcal{R}$ for the multipartite network (top left) and the three synthetic multipartite model networks.

physicians' networks, one attending and one fellow form a team, whereas teams vary in size and composition for the producers' network (Figs. S21 and S22 in the Supplemental Material [64]).

## V. DISCUSSIONS

Transmission dynamics on systems best described by temporal multipartite networks show much greater variability across realization than one is led to believe from simulating dynamics on time-aggregated unipartite projection networks. This difference means that studies conducted on projected networks are inadequate both for obtaining estimates of expected outcomes and deciding whether an observed outcome is consistent with a given model. We show that the ratio of the number of agents to the number of teams per aggregation cycle has a significant impact on the resulting differences of simulating the transmission dynamics on multipartite

networks versus their unipartite projections. This means that there is the need to accomplish a trade-off between decreasing the number of teams per aggregation cycle whereas maintaining a small ratio of the number of agents to the number of teams per aggregation cycle.

Additionally, we show that simulations on very simplistic multipartite synthetic networks can capture real dynamics much better than simulations on projected networks. Our results inform about the types of information to focus on when there are limited resources for collecting data on complex multipartite networks.

Finally, because of the growing body of information-rich networks datasets [50], our results open new venues to explore a variety of systems that can be represented as temporal multipartite networks, including very large networks with different topological heterogeneity that could result in new behaviors that cannot be unveiled by simple time-aggregated unipartite projected networks.

## VI. METHODS

### A. Transmission models

We study the transmission dynamics as a contagion process diffusing either on the actual temporal multipartite network, or on different projections obtained for different levels of temporal aggregation, or on synthetic multipartite networks. We consider two contagion models: the SI model and the GC model.

*SI model.* This model agents can be in one of two states, susceptible ($S$) and infectious ($I$), which we represent as 0 and 1, respectively. At each event or time step $t \geqslant 1$, a susceptible node in contact with an infectious node becomes infected with probability $p$.

*GC model.* This model interpolates between two classes of transmission mechanisms: threshold [72] and infection [61]. When a susceptible agent is exposed to infectious agents, they receive an infection dose $d$ with probability $p$. Susceptible agents keep in memory the infection doses received over the previous $\tau$ time steps. Susceptible agents become infected when their cumulative remembered infection doses exceed a threshold, $D$ (see algorithm 1 in the Supplemental Material [64]). For simplicity, we restrict our attention to the case where $\tau = \infty$. For both transmission models, each simulation is stochastic and starts with two infectious nodes chosen randomly from a fully susceptible population. Note that the SI model is captured by the case where $d = D$ in the GC model.

### B. Simulation of dynamics

We systematically study the transmission dynamics for $p \in \{0.1, \ldots, 1.0\}$ (SI) and for 100 pairs of values $p \in \{0.1, \ldots, 1.0\} d \in \{0.1, \ldots, 1.0\}$ (GC) and for four different timescales of aggregation in the unipartite projection. For each set of parameter values and each network, we simulate 1000 independent realizations.

*Unipartite network dynamics.* In order to assure the robustness of our results for unipartite projected networks, we consider two methods to update the state of the agents—asynchronous update using a random walk procedure [73,74] and synchronous update [52,72,75,76].

In the asynchronous update, at the start of transmission, $t = 0$, we randomly select two nodes as initially infected nodes $n_1, n_2$. At each time step, a node $n_1$ is selected and one of its neighbors is randomly choose to calculate the transmission dynamic. Then we repeat the previous step $t$ times according to the time interval the network was projected and the number edges in the network. The probability of choosing the neighbor and the size of the influence two active nodes have on each other is proportional to the edge weights. At each step only the chosen nodes interact (see algorithm 2 in the Supplemental Material [64]). When one node is in the susceptible state and the other is in the infected state, the susceptible node receives a dose $d$ with probability $p$ for the GC model or is infected with probability $p$ for the SI model (see algorithm 1 in the Supplemental Material [64]). If both nodes are susceptible or infected (when nodes are in the same state), no exchange occurs. Asynchronous update behaves as if it has temporal properties because the edges are selectively active for a certain duration.

In the synchronous update, at time step $t$, the transmission of infection can happen across all connected nodes, and all nodes update their state simultaneously (see algorithm 3 in the Supplemental Material [64]). This updating strategy is inspired by a modified linear threshold model where a susceptible node is influenced by all of its infected neighbors at each time step [52,72,75,76]. The influence a node exerts on another depends on the edge weights. In the synchronous update, it is assumed that edges always existed during the interval of the simulation, therefore, in a network of finite size and nonzero transmission rate, all nodes will eventually become infected [77].

*Multipartite network dynamics.* For the multipartite networks, the transmission can progress through the active members of the coverage team; physicians who are part of a team at each time point interact and influence each other. When one node is in the susceptible state and the other is in the infected state in an active team, the susceptible node receives a dose $d$ with probability $p$ for the GC model or is infected with probability $p$ for the SI model. However, because a certain team can be active during different periods, for each day, the infection process is repeated until the team is no longer active (see algorithm 4 in the Supplemental Material [64]).

For the producers' multipartite network because we do not have data on exact production periods, we randomized the order of the movie production during each year to simulate the transmission dynamics. The number of movies produced per year stays approximately constant (Fig. S22 in the Supplemental Material [64]), therefore, we can assume the gap between produced movies remains approximately constant.

### C. Generating the synthetic bipartite network with a modular structure

We use the generative model for bipartite networks introduced by Guimerá *et al.* [66], because it yields an ensemble of random bipartite networks with a prescribed modular structure. To create a network instance, we first dived $N$ agents equally into five groups. We then sequentially create $N_T$ teams. Each time, we first select at random a primary group. Then, with probability $p_p$ (team homogeneity) we select an

agent at random from the primary group. With probability $1 - p_p$, we select an agent at random from one of the four non-primary groups. Thus, we use the generated bipartite networks to systematically investigate the properties of the bipartite networks that affect the dynamics of the projected networks.

[1] D. J. Daley and J. Gani, *Epidemic Modelling: An Introduction* (Cambridge University Press, Cambridge, UK, 2001), pp. 1–213.

[2] R. Pastor-Satorras and A. Vespignani, Epidemic Spreading in Scale-Free Networks, Phys. Rev. Lett. **86**, 3200 (2001).

[3] F. Liljeros, C. R. Edling, and L. A. N. Amaral, Sexual networks: Implication for the transmission of sexually transmitted infection, Microbes Infect. **5**, 189 (2003).

[4] A. Barrat, M. Barthelemy, and A. Vespignani, *Dynamical Processes on Complex Networks* (Cambridge University Press, Cambridge, UK, 2008).

[5] M. Chinazzi, J. T. Davis, M. Ajelli, C. Gioannini, M. Litvinova, S. Merler, A. P. y Piontti, K. Mu, L. Rossi, K. Sun *et al.*, The effect of travel restrictions on the spread of the 2019 novel coronavirus (covid-19) outbreak, Science **368**, 395 (2020).

[6] B. F. Maier and D. Brockmann, Effective containment explains subexponential growth in recent confirmed covid-19 cases in china, Science **368**, 742 (2020).

[7] L. A. N. Amaral and J. M. Ottino, Complex networks, Eur. Phys. J. B **38**, 147 (2004).

[8] M. Newman, *Networks: An Introduction* (Oxford University Press, Oxford, 2010), pp. 1–784.

[9] A. Barrat, M. Barthélemy, R. Pastor-Satorras, and A. Vespignani, The architecture of complex weighted networks. Proc. Natl. Acad. Sci. USA **101**, 3747 (2004).

[10] D. B. Stouffer, J. Camacho, R. Guimerà, C. A. Ng, and L. A. N. Amaral, Quantitative patterns in the structure of model and empirical food webs, Ecology **86**, 1301 (2005).

[11] S. V. Buldyrev, R. Parshani, G. Paul, H. E. Stanley, and S. Havlin, Catastrophic cascade of failures in interdependent networks, Nature (London) **464**, 1025 (2010).

[12] M. Kivelä, A. Arenas, M. Barthelemy, J. P. Gleeson, Y. Moreno, and M. A. Porter, Multilayer networks, J. Complex Networks **2**, 203 (2014).

[13] G. Fagiolo, J. Reyes, and S. Schiavo, On the topological properties of the world trade web: A weighted network analysis, Physica A **387**, 3868 (2008).

[14] L. G. A. Alves, A. Aleta, F. A. Rodrigues, Y. Moreno, and L. A. N. Amaral, Centrality anomalies in complex networks as a result of model over-simplification, New J. Phys. **22**, 013043 (2020).

[15] S. Boccaletti, G. Bianconi, R. Criado, C. I. Del Genio, J. Gómez-Gardenes, M. Romance, I. Sendina-Nadal, Z. Wang, and M. Zanin, The structure and dynamics of multilayer networks, Phys. Rep. **544**, 1 (2014).

[16] T. E. Gorochowski and T. O. Richardson, How behaviour and the environment influence transmission in mobile groups, in *Temporal Network Epidemiology*, edited by N. Masuda and P. Holme (Springer, Singapore, 2017), pp. 17–42.

[17] A. Li, S. P. Cornelius, Y.-Y. Liu, L. Wang, and A.-L. Barabási, The fundamental advantages of temporal networks, Science **358**, 1042 (2017).

[18] G. Miritello, E. Moro, and R. Lara, Dynamical strength of social ties in information spreading, Phys. Rev. E **83**, 045102(R) (2011).

[19] S. Liu, N. Perra, M. Karsai, and A. Vespignani, Controlling Contagion Processes in Activity Driven Networks, Phys. Rev. Lett. **112**, 118702 (2014).

[20] P. Holme and J. Saramäki, Temporal networks, Phys. Rep. **519**, 97 (2012).

[21] R. Pastor-Satorras, C. Castellano, P. Van Mieghem, and A. Vespignani, Epidemic processes in complex networks, Rev. Mod. Phys. **87**, 925 (2015).

[22] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D. U. Hwang, Complex networks: Structure and dynamics, Phys. Rep. **424**, 175 (2006).

[23] N. Perra, B. Gonçalves, R. Pastor-Satorras, and A. Vespignani, Activity driven modeling of time varying networks, Sci. Rep. **2**, 469 (2012).

[24] T. Takaguchi, N. Masuda, and P. Holme, Bursty communication patterns facilitate spreading in a threshold-based epidemic dynamics, PLoS One **8**, e68629 (2013).

[25] F. Karimi and P. Holme, Threshold model of cascades in empirical temporal networks, Physica A **392**, 3476 (2013).

[26] J. I. Perotti, H.-H. Jo, P. Holme, and J. Saramäki, Temporal network sparsity and the slowing down of spreading, arXiv:1411.5553.

[27] P. Holme, Probing empirical contact networks by simulation of spreading dynamics, in *Complex Spreading Phenomena in Social Systems* (Springer, Berlin, 2018), pp. 109–124.

[28] G. Petri and A. Barrat, Simplicial Activity Driven Model, Phys. Rev. Lett. **121**, 228301 (2018).

[29] M. De Domenico, C. Granell, M. A. Porter, and A. Arenas, The physics of spreading processes in multilayer networks, Nat. Phys. **12**, 901 (2016).

[30] G. Krings, M. Karsai, S. Bernhardsson, V. D. Blondel, and J. Saramäki, Effects of time window size and placement on the structure of an aggregated communication network, EPJ Data Sci. **1**, 4 (2012).

[31] A. Clauset and N. Eagle, Persistence and periodicity in a dynamic proximity network, in *DIMACS Workshop on Computational Methods for Dynamic Interaction Network* (Rutgers University, NJ, USA, 2007).

[32] B. Ribeiro, N. Perra, and A. Baronchelli, Quantifying the effect of temporal resolution on time-varying networks, Sci. Rep. **3**, 3006 (2013).

[33] R. K. Darst, C. Granell, A. Arenas, S. Gómez, J. Saramäki, and S. Fortunato, Detection of timescales in evolving complex systems, Sci. Rep. **6**, 39713 (2016).

[34] A. R. Benson, R. Abebe, M. T. Schaub, A. Jadbabaie, and J. Kleinberg, Simplicial closure and higher-order link prediction, Proc. Natl. Acad. Sci. USA **115**, E11221 (2018).

[35] D. Bisanzio, L. Bertolotti, L. Tomassone, G. Amore, C. Ragagli, A. Mannelli, M. Giacobini, and P. Provero, Modeling the spread of vector-borne diseases on bipartite networks, PLoS One **5**, e13796 (2010).

[36] L. Wen and J. Zhong, Global asymptotic stability and a property of the sis model on bipartite networks, Nonlinear Anal.: Real World Appl. **13**, 967 (2012).

[37] S. Han, M. Sun, B. C. Ampimah, and D. Han, Epidemic spread in bipartite network by considering risk awareness, Physica A **492**, 1909 (2018).

[38] G. A. Pavlopoulos, P. I. Kontou, A. Pavlopoulou, C. Bouyioukos, E. Markou, and P. G. Bagos, Bipartite graphs in systems biology and medicine: A survey of methods and applications, GigaScience **7**, giy014 (2018).

[39] L. Cao, X. Li, B. Wang, and K. Aihara, Rendezvous effects in the diffusion process on bipartite metapopulation networks, Phys. Rev. E **84**, 041936 (2011).

[40] S. Maslov and K. Sneppen, Specificity and stability in topology of protein networks, Science **296**, 910 (2002).

[41] S. Pilosof, M. A. Porter, M. Pascual, and S. Kéfi, The multilayer nature of ecological networks, Nat. Ecol. Evol. **1**, 1 (2017).

[42] K. Börner, J. T. Maru, and R. L. Goldstone, The simultaneous evolution of author and paper networks. Proc. Natl. Acad. Sci. USA **101**, 5266 (2004).

[43] S. Helleringer and H.-P. Kohler, Sexual network structure and the spread of HIV in Africa: Evidence from Likoma Island, Malawi, Aids **21**, 2323 (2007).

[44] L. E. C. Rocha, F. Liljeros, and P. Holme, Simulated epidemics in an empirical spatiotemporal network of 50,185 sexual contacts, PLoS Comput. Biol. **7**, e1001109 (2011).

[45] R. Guimera, B. Uzzi, J. Spiro, and L. A. N. Amaral, Team assembly mechanisms determine collaboration network structure and team performance, Science **308**, 697 (2005).

[46] J. Leskovec, A. Singh, and J. Kleinberg, Patterns of influence in a recommendation network, in *Pacific-Asia Conference on Knowledge Discovery and Data Mining, Delhi, India, 2006* (Springer, Berlin, 2006), pp. 380–389.

[47] T. Zhou, J. Ren, M. Medo, and Y. C. Zhang, Bipartite network projection and personal recommendation, Phys. Rev. E **76**, 046115 (2007).

[48] A. Godoy-Lorite, R. Guimerà, C. Moore, and M. Sales-Pardo, Accurate and scalable social recommendation using mixed-membership stochastic block models. Proc. Natl. Acad. Sci. USA **113**, 14207 (2016).

[49] M. T. Rivera, S. B. Soderstrom, and B. Uzzi, Dynamics of dyads in social networks: Assortative, relational, and proximity mechanisms, Annu. Rev. Sociol. **36**, 91 (2010).

[50] A. Clauset, E. Tucker, and M. Sainz, The Colorado Index of Complex Networks, https://icon.colorado.edu/ (2016).

[51] P. Holme, Network reachability of real-world contact sequences, Phys. Rev. E **71**, 046119 (2005).

[52] P. Shakarian, A. Bhatnagar, A. Aleali, E. Shaabani, and R. Guo, *Diffusion in Scial Networks* (Springer, Berlin, 2015), pp. 1–101.

[53] S. A. Myers, C. Zhu, and J. Leskovec, Information diffusion and external influence in networks, in *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Beijing, China, 2012* (ACM, New York, 2012), pp. 33–41.

[54] A. Nematzadeh, E. Ferrara, A. Flammini, and Y. Y. Ahn, Optimal Network Modularity for Information Diffusion, Phys. Rev. Lett. **113**, 088701 (2014).

[55] N. Masuda, M. A. Porter, and R. Lambiotte, Random walks and diffusion on networks, Phys. Rep. **716**, 1 (2017).

[56] E. Valdano, L. Ferreri, C. Poletto, and V. Colizza, Analytical Computation of the Epidemic Threshold on Temporal Networks, Phys. Rev. X **5**, 021005 (2015).

[57] E. Valdano, C. Poletto, A. Giovannini, D. Palma, L. Savini, and V. Colizza, Predicting epidemic risk from past temporal contact data, PLoS Comput. Biol. **11**, e1004152 (2015).

[58] J. P. Gleeson, K. P. O'Sullivan, R. A. Baños, and Y. Moreno, Effects of Network Structure, Competition and Memory Time on Social Spreading Phenomena, Phys. Rev. X **6**, 021019 (2016).

[59] C. H. Weiss, J. Poncela-Casasnovas, J. I. Glaser, A. R. Pah, S. D. Persell, D. W. Baker, R. G. Wunderink, and L. A. N. Amaral, Adoption of a High-Impact Innovation in a Homogeneous Population, Phys. Rev. X **4**, 041008 (2014).

[60] M. Wasserman, X. H. T Zeng, and L. A. N. Amaral, Cross-evaluation of metrics to estimate the significance of creative works, Proc. Natl. Acad. Sci. USA **112**, 1281 (2015).

[61] R. M. Anderson and R. M. May, *Infectious Diseases of Humans: Dynamics and Control* (Oxford University Press, Oxford, 1992), p. 768.

[62] P. S. Dodds and D. J. Watts, Universal Behavior in a Generalized Model of Contagion, Phys. Rev. Lett. **92**, 218701 (2004).

[63] P. S. Dodds and D. J. Watts, A generalized model of social and biological contagion, J. Theor. Biol. **232**, 587 (2005).

[64] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevE.103.022320 for data description, additional figures, supplementary text, and model algorithms.

[65] IMDb, https://imdb.com.

[66] R. Guimerà, M. Sales-Pardo, and L. A. N. Amaral, Module identification in bipartite and directed networks, Phys. Rev. E **76**, 036102 (2007).

[67] V. Colizza, A. Flammini, M. A. Serrano, and A. Vespignani, Detecting rich-club ordering in complex networks, Nat. Phys. **2**, 110 (2006).

[68] N. Masuda and P. Holme, *Temporal Network Epidemiology* (Springer, Berlin, 2017).

[69] M. J. Keeling and Ken T.D. Eames, Networks and epidemic models, J. R. Soc., Interface **2**, 295 (2005).

[70] G. F. de Arruda, G. Petri, F. A. Rodrigues, and Y. Moreno, Impact of the distribution of recovery rates on disease spreading in complex networks, Phys. Rev. Research **2**, 013046 (2020).

[71] H. Alt and M. Godau, Computing the Fréchet distance between two polygonal curves, Int. J. Comput. Geom. Appl. **05**, 75 (1995).

[72] M. Granovetter, Threshold models of collective behavior, Am. J. Sociol. **83**, 1420 (1978).

[73] J. Tang, J. Sun, C. Wang, and Z. Yang, Social influence analysis in large-scale networks, in *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Paris, France, 2009* (ACM, New York, 2009), pp. 807–816.

[74] J. K. Aggarwal and M. S. Ryoo, Human activity analysis: A review, ACM Comput. Surv. (CSUR) **43**, 1 (2011).

[75] W. Chen, Y. Yuan, and L. Zhang, Scalable influence maximization in social networks under the linear threshold model, *IEEE International Conference on Data Mining, Sydney, Australia, 2010* (IEEE, Piscataway, NJ, 2010), pp. 88–97.

[76] T. W. Valente, Network models of the diffusion of innovations, Comput. Math. Organ. Theory **2**, 163 (1996).

[77] T. Gross, C. J. D. D'Lima, and B. Blasius, Epidemic Dynamics on an Adaptive Network, Phys. Rev. Lett. **96**, 208701 (2006).

# Supplemental material: Spreader events and the limitations of projected networks for capturing dynamics on multipartite networks

Hyojun A. Lee,[1, *] Luiz G. A. Alves,[1, *] and Luís A. Nunes Amaral[1, 2, 3, †]

[1]*Department of Chemical and Biological Engineering,*
*Northwestern University, Evanston, Illinois, United States of America*
[2]*Department of Physics and Astronomy, Northwestern University, Evanston, Illinois, United States of America*
[3]*Northwestern Institute on Complex Systems, Northwestern University, Evanston, Illinois, United States of America*
(Dated: January 25, 2021)

## DATA

The shift schedules from pulmonary and critical care physicians medical intensive care unit (ICU) at Northwestern Memorial Hospital (NMH) are from July 2012 to August 2016 and comprise of 1348 teams (810 weekday teams and 538 weekend teams). There are two types of physicians, attendings and fellows, and they make the most important decisions regarding patient care in the ICU. At any given time, a team consists of one attending and one fellow. Each are on duty and assigned to treat different groups of patients; there are 3 teams on weekdays and 2 teams on weekends (Saturday and Sunday). Attendings are senior physicians who are also faculty members at the medical school, and fellows are junior physicians who have completed their residency but are still in training. The reason the active teams consist of both attending and fellow unlike other hospitals is because NMH is a teaching hospital tasked with training physicians. Weekday teams and weekend teams are independently scheduled. The members of a team do not change over the course of the week or weekend. Each team conducts daily rounds during which team members exchange information about each patient and make treatment decisions. From the data, we can infer a multipartite network with three groups of nodes: attendings, fellows, and teams. The two groups of physicians are only connected to each other through teams, teams are connected through physicians. The teams are the primary source of peer-to-peer interactions.

## MULTIPARTITE NETWORK PROJECTION ONTO UNIPARTITE NETWORK

A multipartite network is simplified onto a unipartite network by one-mode projecting. In the case of bipartite networks where there are two groups of nodes $X$ and $S$, one-mode projection onto $X$ means building a unipartite network containing only $X$ nodes, where edges are created for two $X$ nodes when they have at least one common neighboring $S$ node [1]. Temporal networks are simplified by first removing the timestamps of the events within a time period [2]. Then edges created during the time period are assumed to happen with equal probability [3]. In temporal multipartite network, timestamps from temporal nodes are removed, then the network is projected onto non-temporal nodes [4, 5].

We first aggregate the temporal nodes over a time interval of 7 days 1, 2, 3, and 12 months. All the edges that were created during the aggregated time periods are edges in the static graph. In the construction of a unipartite projected network, the weighting of edges is the key problem, and different suggestions have been investigated concerning the appropriate weighting of the edges. The first approach in determining weights when the number of teams that the physicians had together during the aggregated time periods (number of teams physician $i$ and $j$ shared together, $w_{ij}^n$). The second approach is to use the actual duration of interaction period, that is, the number of days that physician $i$ and $j$ had together $w_{ij}^d$). We investigate the models for both approaches.

———

* These authors contributed equally to this work
† amaral@northwestern.edu

# PSEUDO-CODES OF THE TRANSMISSION DYNAMICS ON MULTIPARTITE AND PROJECTED NETWORKS

ALGORITHM 1. Pseudo-code to update status of infection of nodes according to the GC model with infinite memory. For $D = d$, we recover the SI model.

---

**Input** $n_i$ and $n_j$ are two agent nodes in states $S_i$ an $S_j$, respectively, with $S_i$ taking values in [0,1].

**Input** $p$ is the transmission probability from infectious to susceptible/exposed.

**Input** $d$ is the dose shared at each successful transmission event.

**Input** $D$ is the cumulative dose needed for an exposed agent node to transition to the infected state $S = 1$.

**Input** $x$ is a random variate with a uniform distribution in the interval [0, 1]

**if** $S_i \geq D$ **and** $S_j < D$ **then**:
  **if** $x \leq p$ **then**:
    $S_j = \min(1, S_j + w * d)$
  **end if**
**else**:
  **if** $S_i < D$ **and** $S_j \geq D$ **then**:
    **if** $x \leq p$ **then**:
      $S_i = \min(1, S_i + w * d)$
    **end if**
  **end if**
**end if**

---

ALGORITHM 2. Pseudo-code to simulate the transmission dynamics in the temporal unipartite networks in the asynchronous case (random walk case).

---

**Input** N is a set of agent nodes who are in states $S_i$ for $i \in \{0, 1, \ldots, L-1\}$.

**Input** $G = \{G_0(N, E_0, \Delta t_0), \ldots, G_n(N, E_T, \Delta t_T)\}$ is a temporal unipartite network projected on the set of agent nodes at a temporal aggregation level of $\Delta t$.

Draw a node $n_1$ randomly from $N$.
Draw a node $n_2$ randomly from $N - \{n_1\}$.
$S_1 = S_2 = 1$

**for** $G_k$ **in** $\{G_0, G_1, ..., G_T\}$ **do**:
  $E_k$ is the set of edges obtained when projecting event nodes from multipartite network on agent nodes.
  $I_k$ is the number of days in interval $\Delta t_n$.


  **for** $m$ **in** $\{0, 1, \ldots, I_k * |E_k|\}$ **do**
    Draw an edge $e$ randomly from $E_k$.
    $n_i$ and $n_j$ are the terminus nodes of edge $e$
    Update status of nodes $n_i$ and $n_j$ using **Algorithm 1**
  **end for**
  Count number of nodes with $S > D$
**end for**

---

ALGORITHM 3. Pseudo-code to simulate the transmission dynamics in the temporal unipartite networks in the synchronous case.

---

**Input** N is a set of agent nodes who are in states $S_i$ for $i \in \{0, 1, \ldots, L-1\}$.

**Input** $G = \{G_0(N, E_0, \Delta t_0), \ldots, G_n(N, E_T, \Delta t_T)\}$ is a temporal unipartite network projected on the set of agent nodes at a temporal aggregation level of $\Delta t$.

Draw a node $n_1$ randomly from $N$.
Draw a node $n_2$ randomly from $N - \{n_1\}$.
$S_1 = S_2 = 1$

**for** $G_k$ **in** $\{G_0, G_1, \ldots, G_T\}$ **do**:
    $E_k$ is the set of edges obtained when projecting event nodes from multipartite network on agent nodes.
    $I_k$ is the number of days in interval $\Delta t_n$.

    **for** $m$ **in** $\{0, 1, \ldots, I_k\}$ **do**
        Update status of all nodes using **Algorithm 1** simultaneously.
    **end for**
    Count number of nodes with $S > D$
**end for**

---

ALGORITHM 4. Pseudo-code to simulate the transmission dynamics in the temporal multipartite networks

---

**Input** N is a set of agent nodes who are in states $S_i$ for $i \in \{0, 1, \ldots, L-1\}$.

**Input** $V = \{V_0, V_1, \ldots, V_M\}$ is a set of temporally ordered event nodes involving subsets of nodes from $N$ and having a defined duration in time.

Draw a node $n_1$ randomly from $N$.
Draw a node $n_2$ randomly from $N - \{n_1\}$.
$S_1 = S_2 = 1$

**for** $l$ **in** $\{0, 1, \ldots, M\}$ **do**:
    $N_V$ is the list of nodes connected to event $V_l$.
    $E_l$ is a **shuffled** list of edges of the complete graph of the nodes $N_V$.
    $I_l$ is the duration in days of event $V_l$.

    **for** $m$ **in** $\{0, 1, \ldots, I_l\}$ **do**
        **for** $e$ **in** $E_l$ **do**
            $n_i$ and $n_j$ are the terminus nodes of edge $e$
            Update status of nodes $n_i$ and $n_j$ using **Algorithm 1**
        **end for**
    **end for**
    Count number of nodes with $S > D$
**end for**

---

TRANSMISSION DYNAMICS COMPARISON



FIG. S1. **Comparison between the four different physicians multipartite networks.** Normalized trajectories of transmission dynamics for different parameters on the four real multipartite networks, i.e., one for each academic year. The mean trajectories and their 95% confidence interval are shown in red (2012), blue (2013), green (2014), and purple (2015).
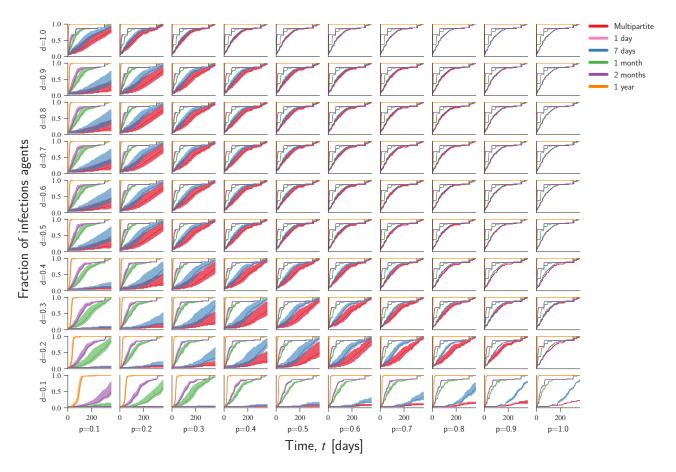
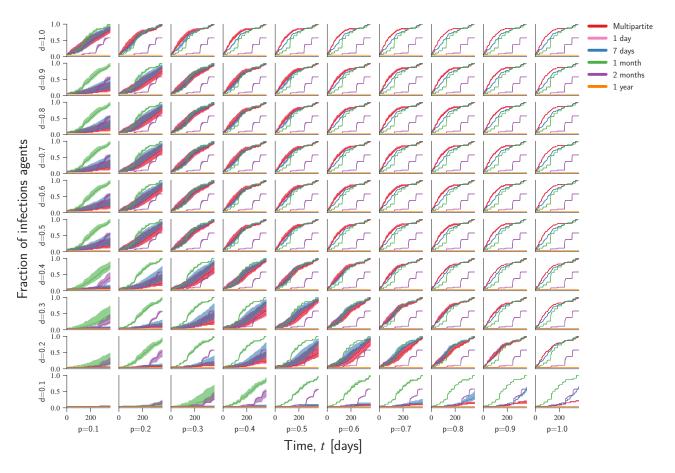# TRANSMISSION DYNAMICS COMPARISON: ASYNCHRONOUS CASE



FIG. S2. **Asynchronous dynamics on unweighted unipartite network projections.** Normalized trajectories of transmission dynamics for different parameters on different time-aggregated unweighted unipartite network projections compared with multipartite projection. These trajectories were obtained using the data from 2014. The mean trajectories and their 95% confidence interval are shown in red (multipartite), pink (unipartite, 1 day), blue (unipartite, aggregated 7 days), green (unipartite, aggregated 1 month), purple (unipartite, aggregated 2 months), and orange (unipartite, aggregated 1 year).

FIG. S3. **Asynchronous dynamics on weighted (by number of days) unipartite network projections.** Normalized trajectories of transmission dynamics for different parameters on different time-aggregated unweighted unipartite network projections compared with multipartite projection. These trajectories were obtained using the data from 2014. The mean trajectories and their 95% confidence interval are shown in red (multipartite), pink (unipartite, 1 day), blue (unipartite, aggregated 7 days), green (unipartite, aggregated 1 month), purple (unipartite, aggregated 2 months), and orange (unipartite, aggregated 1 year).
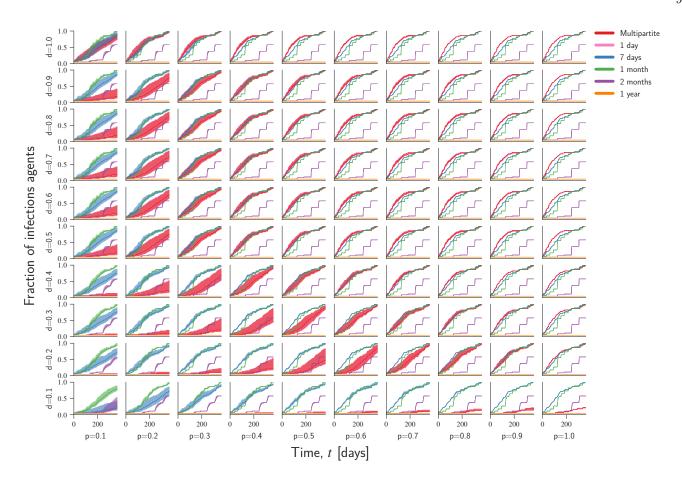
FIG. S4. **Asynchronous dynamics on weighted (by number of shifts) unipartite network projections.** Normalized trajectories of transmission dynamics for different parameters on different time-aggregated unweighted unipartite network projections compared with multipartite projection. These trajectories were obtained using the data from 2014. The mean trajectories and their 95% confidence interval are shown in red (multipartite), pink (unipartite, 1 day), blue (unipartite, aggregated 7 days), green (unipartite, aggregated 1 month), purple (unipartite, aggregated 2 months), and orange (unipartite, aggregated 1 year).
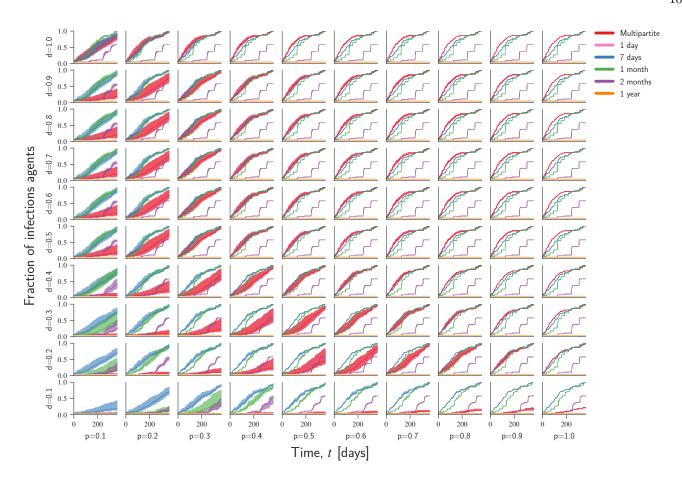
TRANSMISSION DYNAMICS COMPARISON: SYNCHRONOUS CASE



FIG. S5. **Synchronous dynamics on unweighted unipartite network projections.** Normalized trajectories of transmission dynamics for different parameters on different time-aggregated unweighted unipartite network projections compared with multipartite projection. These trajectories were obtained using the data from 2014. The mean trajectories and their 95% confidence interval are shown in red (multipartite), pink (unipartite, 1 day), blue (unipartite, aggregated 7 days), green (unipartite, aggregated 1 month), purple (unipartite, aggregated 2 months), and orange (unipartite, aggregated 1 year).

FIG. S6. **Synchronous dynamics on weighted (by number of days) unipartite network projections.** Normalized trajectories of transmission dynamics for different parameters on different time-aggregated unweighted unipartite network projections compared with multipartite projection. These trajectories were obtained using the data from 2014. The mean trajectories and their 95% confidence interval are shown in red (multipartite), pink (unipartite, 1 day), blue (unipartite, aggregated 7 days), green (unipartite, aggregated 1 month), purple (unipartite, aggregated 2 months), and orange (unipartite, aggregated 1 year).

FIG. S7. **Synchronous dynamics on weighted (by number of shifts) unipartite network projections.** Normalized trajectories of transmission dynamics for different parameters on different time-aggregated unweighted unipartite network projections compared with multipartite projection. These trajectories were obtained using the data from 2014. The mean trajectories and their 95% confidence interval are shown in red (multipartite), pink (unipartite, 1 day), blue (unipartite, aggregated 7 days), green (unipartite, aggregated 1 month), purple (unipartite, aggregated 2 months), and orange (unipartite, aggregated 1 year).

# GENERATING SYNTHETIC NETWORKS

We use information from the physicians schedules to construct nonparametric generative model for synthetic schedules. Physicians are required to participate a total of $M$ teams during period $T$ (e.g. $T$ is one year, 3 teams for 52 weeks = 156 teams). We will test three types of schedules depending on the amount of information. For all schedule types, we generate synthetic physician ids for attendings and fellows. The number of attendings and fellows is taken from the actual numbers of physicians in the data. For schedule type **N**, the number of teams per physician and the gap durations are randomly chosen. For schedule types **NT** and **NTG**, we assign teams to each physician sampled from the distribution of teams from the schedule data. This is recursively done until the sum of the teams are equal to the total number of required teams, $M$. For schedule type **NT**, the teams are then randomly assigned to any time with minimal rules:

1. no same physician is working on more than two teams during the same week,

2. physicians fulfill their required number of teams.

For the schedule type **NTG**, in addition to the two rules use for making type **NT** schedules, the team assignments have to consider the distribution of gap durations. Once the initial schedule is randomly assigned following the two above rules, we compare the distribution of gap durations between the synthetic schedule and the original schedule, using the two-sample Kolmogorov-Smirnov test. We use simulated annealing to shuffle physicians schedule until the difference between the two schedules' distribution of gap durations are statistically insignificant. Specifically, we measure the distribution of gap durations from the synthetic network and compare it to the distribution of gap durations of real network. Every time the teams are shuffled, we accept the new state if the distance from the original distribution of gap durations is decreased. The results displayed in Fig. S8 show that this algorithm leads to a good agreement with the target distribution.
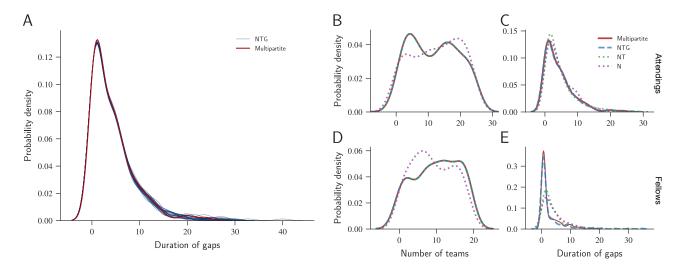


FIG. S8. **(A)** Convergence of the distribution of gap durations attendings, **NTG** model using simulated annealing. **(B-E)** kernel density plots of distributions of different properties of synthetic schedule, distribution of teams of attendings **(B)** and fellows **(C)**, distribution of gap durations of attendings **(D)** and fellows **(E)**. The $p$-values are obtained using the Kolomogorov-Smirnov 2 sample test. **NTG** model $p > 0.9$ for both team and gap durations distributions, **NT** model $p > 0.9$ for team distribution and $p < 0.001$ for the distribution of gap durations, and **N** model $p < 0.001$ for both teams and gap durations distributions
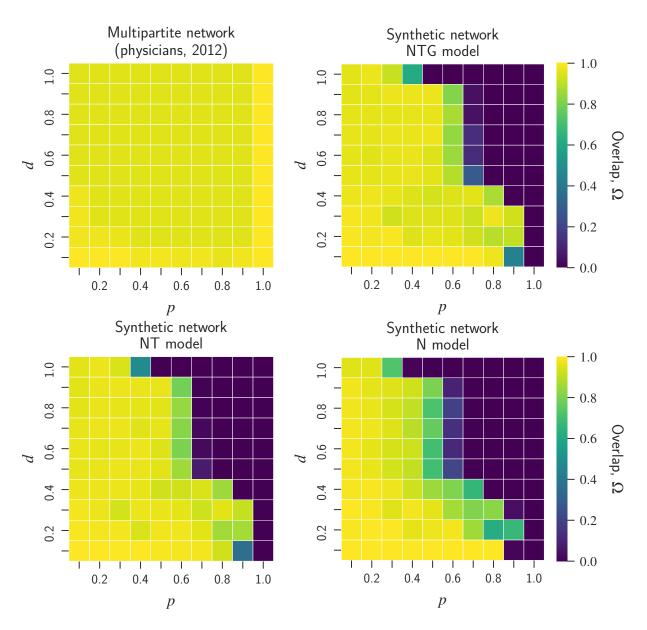
## DYNAMICAL OVERLAP



FIG. S9. **Similarity across realizations of the dynamics on multipartite networks.** Each cell shows $\Omega$, of the transmission dynamics with those obtained on the reference set $\mathcal{R} = \{\mathcal{R}_{2013}, \mathcal{R}_{2014}, \mathcal{R}_{2015}\}$ for multipartite network for 2012.
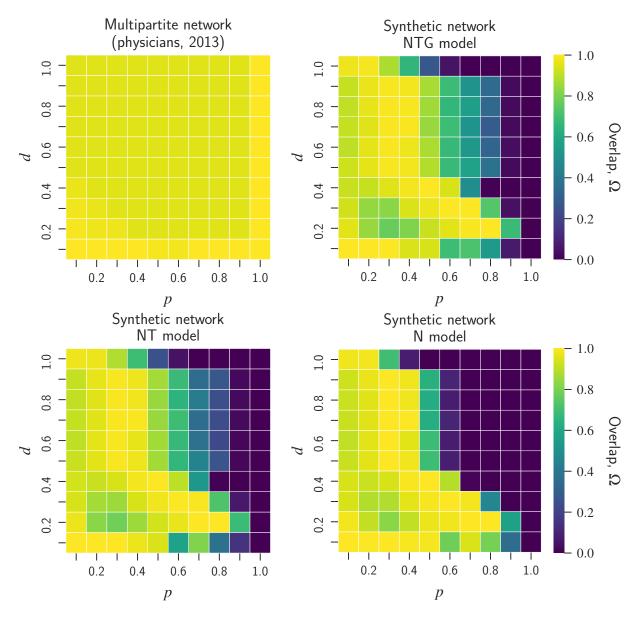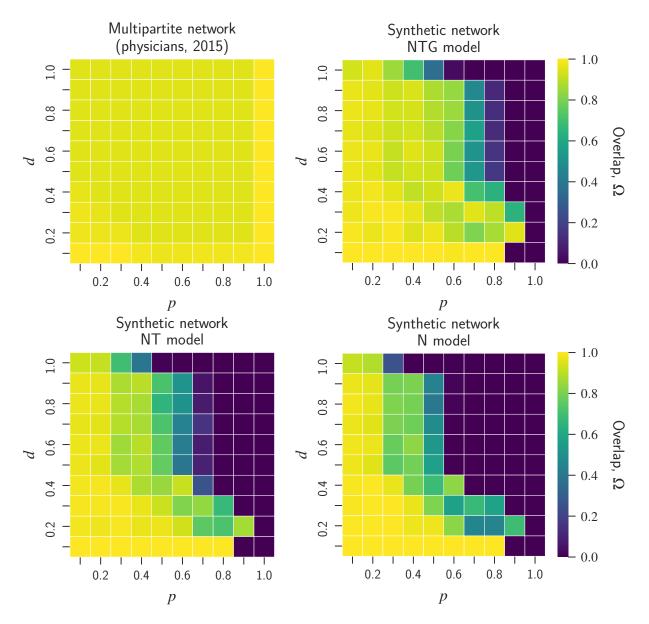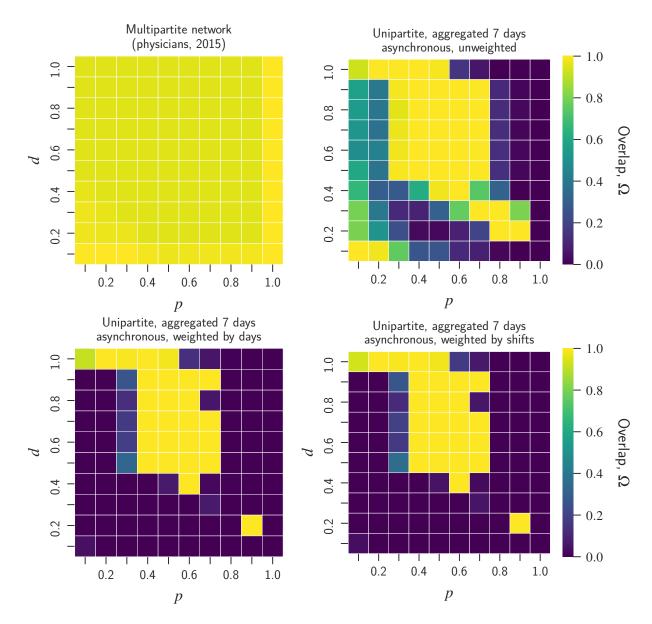
FIG. S10. **Similarity across realizations of the dynamics on multipartite networks.** Each cell shows $\Omega$, of the transmission dynamics with those obtained on the reference set $\mathcal{R} = \{\mathcal{R}_{2012}, \mathcal{R}_{2014}, \mathcal{R}_{2015}\}$ for multipartite network for 2013.

FIG. S11. **Similarity across realizations of the dynamics on multipartite networks.** Each cell shows $\Omega$, of the transmission dynamics with those obtained on the reference set $\mathcal{R} = \{\mathcal{R}_{2012}, \mathcal{R}_{2013}, \mathcal{R}_{2014}\}$ for multipartite network for 2015.
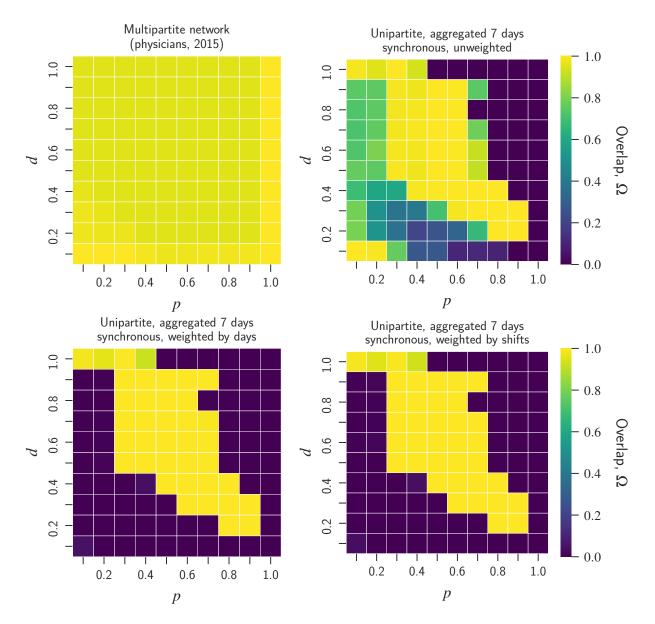
FIG. S12. **Similarity across realizations of the dynamics on multipartite and unipartite networks (aggregated 7 days) using asynchronous updates.** Each cell shows the dynamic range overlap, $\Omega$, of the transmission dynamics for the different projections using asynchronous update.
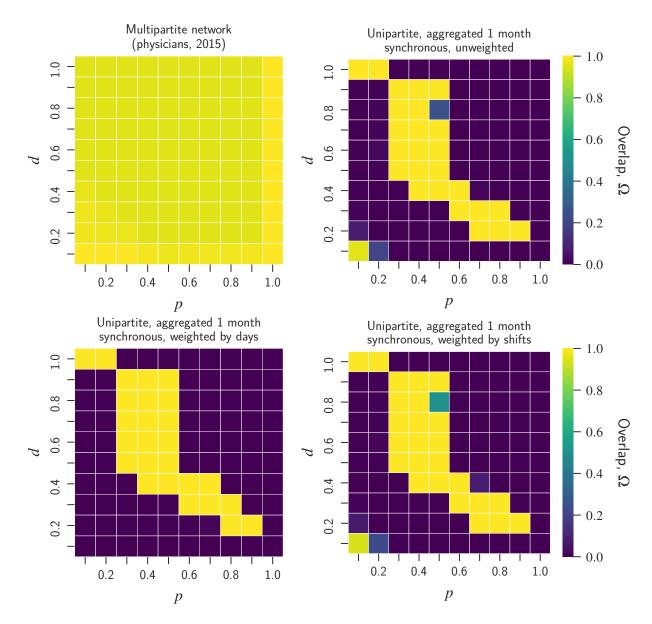
FIG. S13. **Similarity across realizations of the dynamics on multipartite and unipartite (aggregated 1 month) networks using asynchronous updates.** Each cell shows the dynamic range overlap, $\Omega$, of the transmission dynamics for the different projections using asynchronous update.
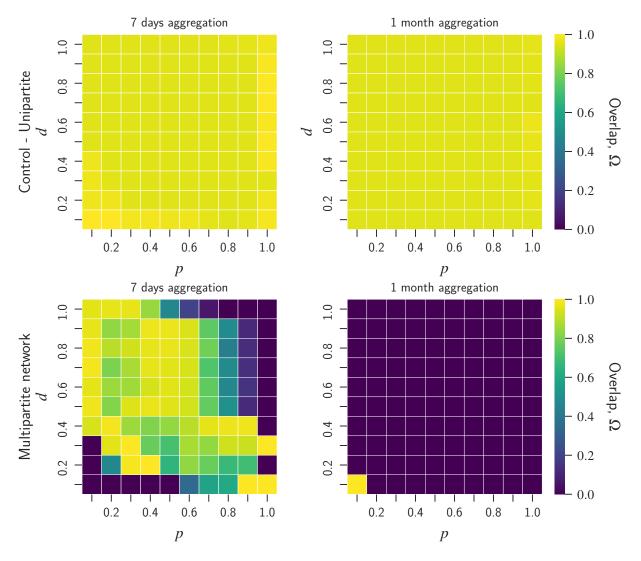
FIG. S14. **Similarity across realizations of the dynamics on multipartite and unipartite networks (aggregated 7 days) using asynchronous updates.** Each cell shows the dynamic range overlap, $\Omega$, of the transmission dynamics for the different projections using synchronous update.

FIG. S15. **Similarity across realizations of the dynamics on multipartite and unipartite networks (aggregated 1 month) using synchronous updates.** Each cell shows the dynamic range overlap, $\Omega$, of the transmission dynamics for the different projections using synchronous update.

FIG. S16. **Similarity across realizations of the dynamics on networks using projected unipartite networks as reference set (asynchronous case).** Each cell shows the dynamic range overlap, $\Omega$, of the transmission dynamics with those obtained for projected multipartite networks and multipartite networks using projected networks as reference set, $\mathcal{R}$, using asynchronous update dynamics.
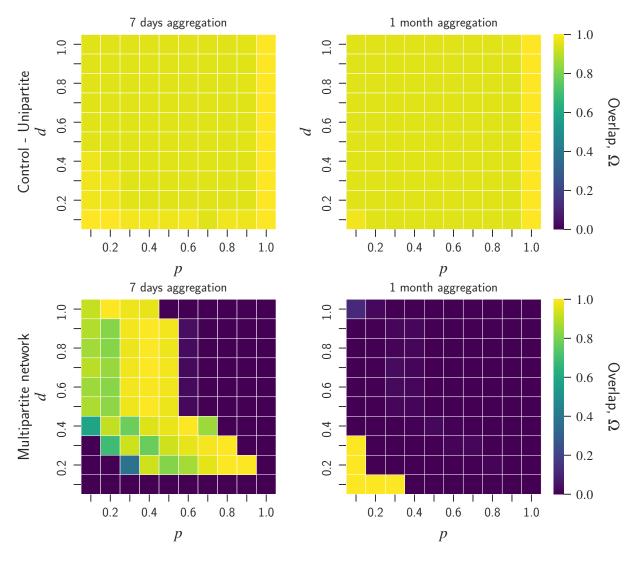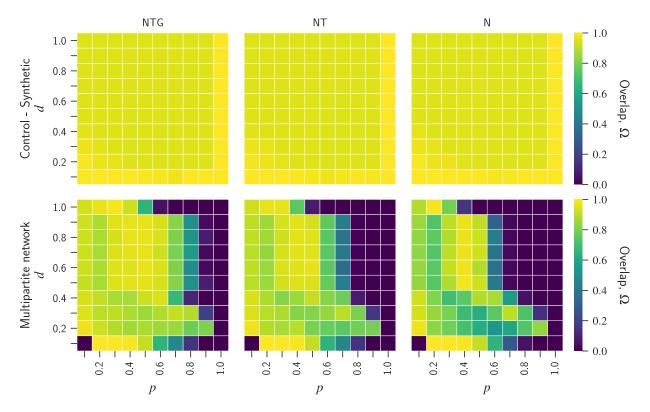
FIG. S17. **Similarity across realizations of the dynamics on networks using projected unipartite networks as reference set (synchronous case).** Each cell shows the dynamic range overlap, $\Omega$, of the transmission dynamics with those obtained for projected multipartite networks and multipartite networks using projected networks as reference set, $\mathcal{R}$, using synchronous update dynamics.

FIG. S18. **Similarity across realizations of the dynamics on networks using synthetic networks as reference set.** Each cell shows the dynamic range overlap, $\Omega$, of the transmission dynamics with those obtained for synthetic multipartite networks and multipartite networks using synthetic networks as reference set, $\mathcal{R}$
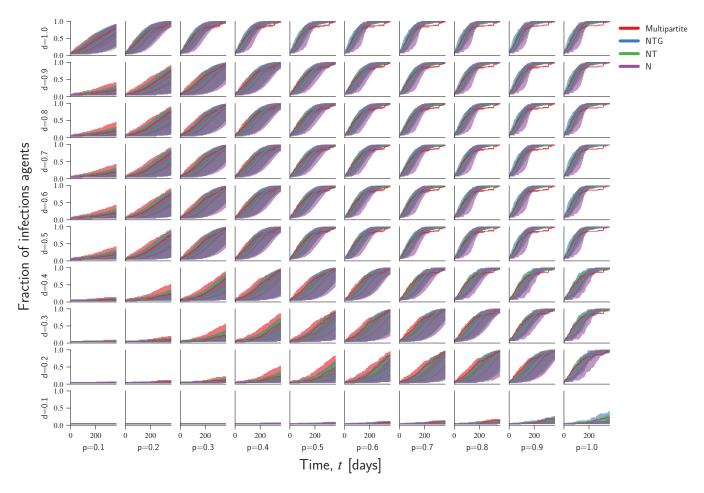
FIG. S19. **Contagion dynamics on synthetic networks**. Normalized trajectories of transmission dynamics for different parameters on different synthetic networks. The mean trajectories and their 95% confidence interval are shown for the contagion dynamics on the multipartite network (red), **NTG** (blue), **NT** (green), and **N** (purple) synthetic network models.
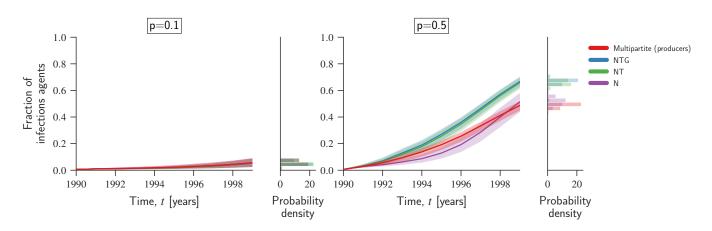
# MOVIE PRODUCER NETWORK



FIG. S20. **Spreading dynamics on multipartite networks and on their synthetic networks in producers network.** Mean fraction of adopters over time (lines) and 95% confidence interval for the dynamics (shaded regions) and the final fraction of adopters for the multipartite and synthetic multipartite networks using SI model $p = 0.1$ and $p = 0.5$
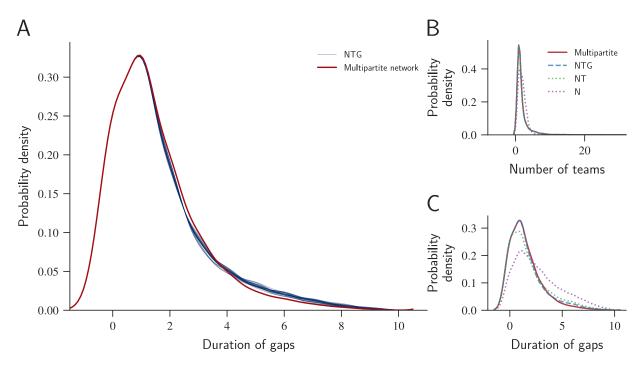


FIG. S21. **(A)** Convergence of the distribution of gap durations movie producers, **NTG** model using simulated annealing. **(B)** Kernel density plots of distributions number of teams of movie producers. Kernel density plots of distributions distribution of gap durations of movies by producers, i.e, years between each movie. The $p$-values are obtained using the Kolomogorov-Smirnov 2 sample test. **NTG** model $p > 0.1$ for both team and gap durations distributions, **NT** model $p > 0.9$ for team distribution and $p < 0.001$ for the distribution of gap durations, and **N** model $p < 0.001$ for both teams and gap durations distributions

[1] Tao Zhou, Jie Ren, Matúš Medo, and Yi Cheng Zhang, "Bipartite network projection and personal recommendation," Physical Review E **76**, 046115 (2007).
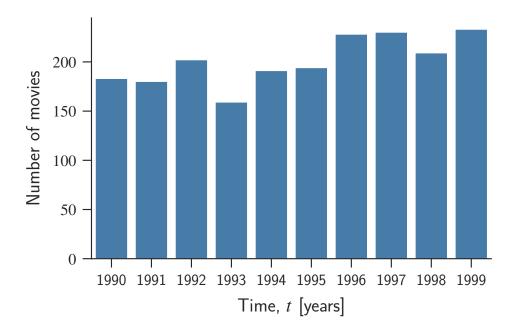[2] Petter Holme, "Network reachability of real-world contact sequences," Physical Review E **71**, 046119 (2005).

FIG. S22. Number of movies that are produced each year from 1990 to 1999 from IMDb.

[3] Luis E.C. Rocha, Fredrik Liljeros, and Petter Holme, "Simulated epidemics in an empirical spatiotemporal network of 50,185 sexual contacts," PLoS Computational Biology **7**, e1001109 (2011).

[4] Fredrik Liljeros, Christofer R. Edling, and Luís A. Nunes Amaral, "Sexual networks: Implication for the transmission of sexually transmitted infection," Microbes Infect. **5**, 189–196 (2003).

[5] Stéphane Helleringer and Hans-Peter Kohler, "Sexual network structure and the spread of HIV in Africa: evidence from Likoma Island, Malawi," Aids **21**, 2323–2332 (2007).