Influenza spread on context-specific social networks

Joshua Rubin Abrams[+], Anne Schwartz[o], Veronica Ciocanel[*], Alexandria Volkening[*], and Björn Sandstede[*]

[+]Department of Mathematics, University of Arizona, Tucson, AZ, USA
[o]Department of Mathematics, Connecticut College, Sudbury, MA, USA
[*]Division of Applied Mathematics, Brown University, Providence, RI, USA

October 13, 2016

Abstract

Studying the spread of infections is an important tool in limiting or preventing future outbreaks. A first step in understanding disease dynamics is constructing networks that reproduce features of real-world interactions. In this paper, we extend the interaction networks logged in an existing diary-based survey at the University of Warwick to larger populations. To preserve realistic structure in the extended network, a context-specific approach is used to lift the data; in particular, we propose different algorithms for producing larger home, work, and social networks. Simulating a discrete SIR model on the full network gives good agreement of the epidemic size (fraction of individuals infected) with data from mild and moderate seasons of influenza outbreak. Our approach allows us to explore how disease transmission and dynamic responses to infection differ depending on interaction context: our main finding is that, while social interactions may be the first to be reduced after influenza infection, curtailing work and school encounters is significantly more effective in controlling the overall severity of the epidemic.

Key Words: dynamic network; disease spread; SIR; influenza; social distance

1 Introduction

Identifying how social interactions shape the way disease spreads in a community is necessary for developing effective strategies to curtail future epidemic outbreaks. Network theory provides essential tools for understanding these human interaction patterns and their relationship to disease transmission, as reviewed in [1–5]. Most studies consider analytic and computational results for idealized networks, which attempt to simplify the complex processes that go into realistic network formation [1, 2, 4, 5]. Other work [6, 7] has focused on constructing more accurate networks of social encounters by making use of real-world data. Diary-based studies are one common way of elucidating network structure by tracking the daily interactions of individuals, but networks specified entirely from survey data are necessarily limited in size [8]. It is therefore important to understand how this data can be extended to produce large networks that preserve essential social structure.

In both frameworks, idealized and data-based network modeling, previous work has emphasized the importance of accounting for the dynamics of networks, as adaptive networks can better capture the capacity of social interactions to change during disease spread [7, 9–11]. However, significantly less attention has been given to exploring how the context of interaction relates to disease spread and dynamic response, despite agreement that distinguishing encounters by context would better elucidate the impact of social structure on disease dynamics [6,12]. In particular, the network structure in home, work, and social settings is intrinsically different: homes are small, fully clustered and distinct; work networks are made up of establishments of various sizes that sparsely connect households; and social networks are highly connected and, as such, serve to bridge between more isolated homes and work places. Extending networks from diary-based data separately for each interaction setting would more faithfully represent the real world, and, from a predictive and preventative perspective, it would be especially useful to investigate how dynamic responses in each context influence epidemic evolution.

In this work, our goal is to study the impact of different contexts of interaction on the spread of influenza in a realistic network. We extend data from the diary-based survey [6] of Read, Eames, and Edmunds to construct a larger population network; the study in [6] provided detailed information on the interactions of students and staff in a university setting and, most notably, categorized encounters by context, allowing for an in-depth study of how disease transmission in the university community depends on interaction setting. Our approach for extending this data differs from that presented in [6] in that their extended network did not preserve context-specific structure; for instance, the extension of the home network in [6] led to a connected network without distinct home clusters. In contrast, we propose a context-specific method for extending the diary-based data [6] to produce larger home, work and social networks that highlight differences in structure depending on setting and reproduce key features of the degree and frequency distributions from the original data.

We explore the roles different interaction contexts have on influenza transmission by simulating a model of disease spread on our data-based extended network. By investigating how dynamic changes in the network in response to infection impact the percentage of infected individuals, we are able to suggest strategies for reducing the spread of disease. Our main results indicate that

- disease spreads most frequently at work but remains localized unless other interactions are present (§4.2);
- social interactions are responsible for the wider spreading of an epidemic (§4.2);
- dynamic responses to infection can substantially reduce epidemic size (§4.4);
- individuals infected with the flu are most likely to reduce their social interactions (§4.5); and
- staying home from work or school has the strongest impact on reducing the severity of an influenza outbreak (§4.5).

In addition, in §4.1 we compare discrete SIR dynamics in the network setting to the classical ordinary differential equation framework, and, in §4.3, we explore how epidemic dynamics depend on the initial infected seed. Simulations of an SIR model on our network also serve to validate our extended network structure and show good agreement with data from mild to moderate seasons of influenza outbreak (discussed in §4.5).

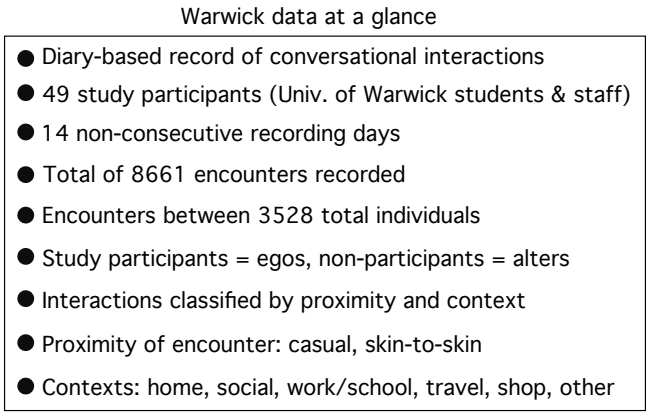


Figure 1: Key features of the diary-based data collected in [6].

2 Overview of the Warwick Study

We begin by discussing the key features of the diary-based data [6] (hereto referred to as the *Warwick data*) that serves as the basis for our work; these features are also summarized in Figure 1. The Warwick data, presented in [6], is a record of the person-to-person interactions of 49 volunteer participants over the course of 14 non-consecutive days. The volunteers, consisting of students and staff at the University of Warwick in Coventry, UK, were asked to keep a log of their conversational interactions on the specified days. These encounters were categorized based on proximity (casual or skin-skin) and context (home, social, work/school, shop, travel, or other). The resulting record contains a total of 8661 encounters among 3528 people, made up of the survey participants and the other individuals they interacted with. Following the terminology in [6], we will refer to the survey participants as egos and the secondary individuals they encountered as alters.

There are two main quantitative measurements obtained from the Warwick data that we rely on to construct our networks for each interaction context: first, for each of the 49 egos, we know the degree (average number of individuals encountered per day) in each interaction context:

$$\text{degree of individual } i \text{ in interaction context } c = \sum_{j=1, j\neq i}^{N} a_{ij}^{c},$$

where N is the number of nodes in the network (3528 in the case of the Warwick data), and the values of the adjacency matrix A are given by $a_{ij}^{c} = 1$ if individuals i and j interacted in context c on at least one of the survey days (that is, if i and j are neighbors) and 0 otherwise (see, for example, [1]). Note that encounters are categorized by interaction context to obtain separate home, social, work, shop, and travel degrees for each participant (thus, for example, two nodes may be neighbors at work, but not at home). These degree measurements are then used to calculate a separate degree distribution (fraction of nodes in the network with degree n [8]) for each interaction context.

Second, the survey data includes a record of repeat interactions over the 14 sample days, and this provides a measure of the strength, or frequency $1/14 \leq f_{ij} \leq 14/14$, of encounters between individuals i and j. We note that, because the Warwick data is recorded from the perspectives of the 49 egos only,

it is challenging to obtain accurate measurements of clustering (a widely-studied quantity related to how connected a graph is [13]), and it is for this reason that we focus mainly on degree and frequency distributions. In §3, we highlight key features of each context in the Warwick data and present our algorithms for extending it to larger networks with similar characteristics.

Read *et al.* [6] extend the diary-based data to a larger network by making multiple copies of the survey participants: in particular, each copy in the extended network has the same degree as an original ego, and is first represented as a node with unconnected edges or stubs emanating from it. Network formation then occurs by randomly connecting stubs with the same weight to create edges between nodes. Because this approach is not context-specific, key features that distinguish the structure in home, social, and work settings are lost; most notably, the extended home network that results from this approach is likely to be highly connected, yet real-world home networks are higly clustered and disconnected. Indeed, treating all interaction contexts in the same manner does not produce realistic networks where home and work sub-networks are made up of distinct units (households and workplaces), while social interactions serve to tie people together across groups. In §3, we highlight the key features of each context in the Warwick data and present our context-specific algorithms for extending it to larger networks that account for these essential differences in home, work/school, and social settings.

3 Model: Network Construction

Our goal is to build an extended network of 3000 individuals based on the Warwick data [6]. Because the intrinsic interaction structure in home, social, work/school, shop, and travel settings are very different, our network model consists of context-specific sub-networks. More precisely, we build an extended network in which each individual (or node) can interact with any other in one or more of three settings: home, social, and work/school (we found that including shop and travel interactions did not have a strong impact on our results, as discussed in §5, so we chose to neglect these contexts). This amounts to creating three separate sub-networks on the same 3000 nodes; taken together, these sub-networks for home, social, and work/school interactions make up our full extended network. It should be noted that we do not differentiate between proximity of interaction (casual or skin-to-skin); this simplifying assumption limits the types of diseases we can reliably simulate to those that do not require skin-to-skin contact for transmission, and so we focus on influenza dynamics (see §4.5).

Sub-network construction in each setting proceeds in two main steps: we first specify the form of the network by assigning edges between nodes, and we then assign a weight to each edge that reflects the frequency of the interaction between the two nodes: the weights, or frequency values, f_{ij} take values between 1/14 and 14/14, since the data [6] was collected on 14 days. The first step is context-specific, while the second step is implemented in each context by sampling from the appropriate frequency distribution generated from [6]. In the three contexts we consider, the core of each extended sub-network is built out of units: our home network consists of households or student dorms in which individuals live; our social network is based on friend groups; and our work/school network is made up of companies or classrooms. We now provide a detailed summary of network construction for the

home, social, and work/school contexts in §3.1, §3.2, and §3.3, respectively.

3.1 Home Network

Our home network is expanded from the Warwick data by adding one home unit at a time, with the size of the household determined by the degree distribution for Warwick home encounters [6]. In particular, suppose we sample from this distribution to obtain a target degree n. Then we generate $n+1$ new nodes and connect them all to each other, leading to a household where every member has degree n. We also define the average local clustering coefficient C as

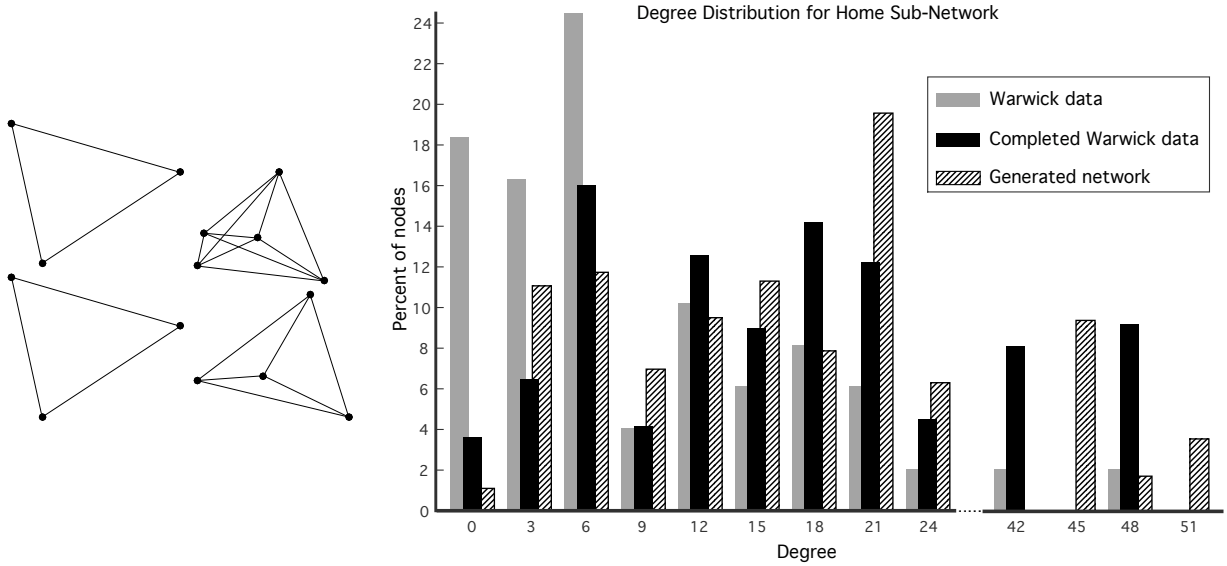
$$C = \frac{1}{N}\frac{\text{number of triangles connected to node } i}{\text{number of triples centered on node } i} = \frac{1}{N}\sum_{i=1}^{N}\frac{2\ell_i}{n_i(n_i-1)},$$

where N is the number of nodes in the network, a triangle is a set of three mutually connected individuals, a triple centered on node i consists of node i and any two nodes it is connected to (regardless of if they are connected to each other), n_i is the degree of node i, and ℓ_i is the number of edges in G_i (the subgraph of neighbors of node i) [8, 13]. Note that $C = 1$ for all the nodes in our home network because we prescribe all-to-all coupling in each household (that is, individuals interact with all members of their shared household). Thus, our approach captures the highly clustered, disconnected structure of the home network. In contrast, because Read *et al.* do not use a context-specific algorithm, the larger home network in [6] is unrealistically connected and does not reproduce this clustering feature. Example home units are shown in Figure 2a.

Once a home unit of size n is generated as discussed above, the next step is to determine the frequency or weight of each interaction in the household. Because survey participants consisted of students and staff at the University of Warwick, many of these individuals lived in residence halls or shared houses, and this leads to high degree and low frequency interactions in the home setting. Students in residence halls may encounter many individuals living in their building, but not necessarily see every housemate every day. In contrast, participants living in family homes might be more likely to display frequent interactions with a core group of fewer people at home. As shown in Figure 3a, having a high degree is indeed associated with lower average interaction frequency in the Warwick home data. To account for this feature, we separate the frequency distribution from [6] into two components: the distribution for survey participants with home degree ≤ 8 and the corresponding distribution when degree > 8.

For small households, we assign weights to each edge by sampling from the frequency distribution for home interactions with degree ≤ 8; and, for large homes, we assign each edge in the home unit a weight by sampling from the corresponding distribution for degree > 8. This completes the home unit: the result is a new, fully-clustered household of size $n+1$ added to the network, where the degree n of each household member was determined from the diary-based data [6], the frequency of interactions was sampled from the real frequency distribution [6], and the clustering coefficient is one for every node, since we assume all-to-all coupling. The full home network algorithm is summarized below:

1. Sample from the home degree distribution of the Warwick data to obtain a target degree n.



(a) (b)

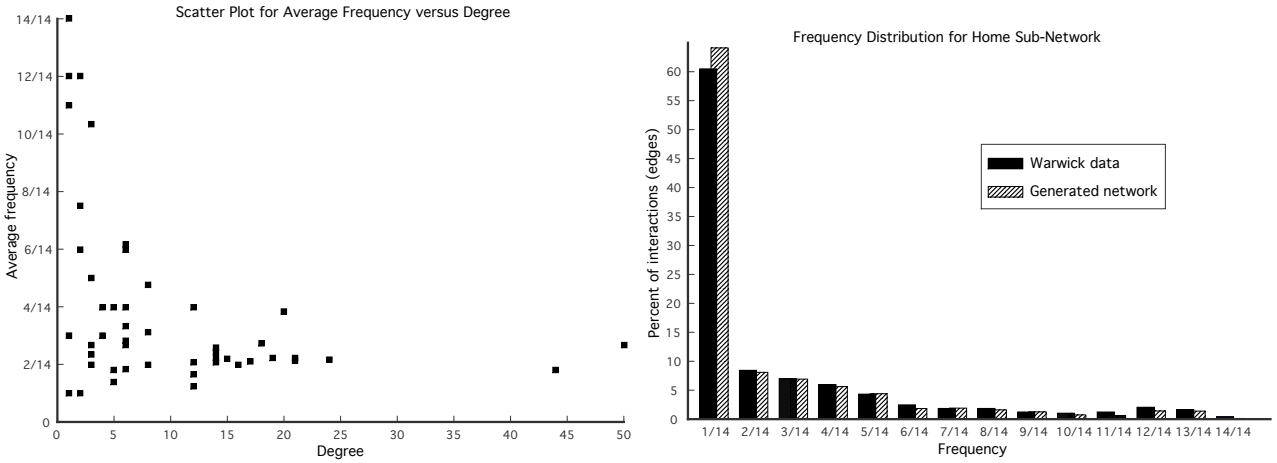
Figure 2: (a) Fully-clustered household units make up our home sub-network; 4 example units are shown. (b) The size of each home is determined by sampling from the home degree distribution associated with the Warwick data [6]. We also plot the completed Warwick data for direct comparison with our generated network (the Warwick data is completed by assuming each degree n corresponds to a fully clustered household of size $n + 1$).

2. Generate $n + 1$ new nodes and connect all to all to create a fully clustered home unit in which each node has degree n.
3. Assign an interaction weight to each edge in the home unit by sampling from the appropriate home frequency distribution of the Warwick data: the distribution for $n \leq 8$ or $n > 8$.
4. Add the new home unit to the existing network and repeat from Step 1 until 3000 nodes are generated (the size of the final home unit added is adjusted to reach the target number of nodes).

The degree and frequency distributions obtained from this algorithm are plotted in Figures 2b and 3b. We assume the 49 volunteers who participated in the diary-based study [6] are all from different households; this means we have to multiply the reported number of nodes with home degree n by $n+1$ to get the total number of alters and egos with that degree. For direct comparison with our network, we also plot this *completed* Warwick data in Figure 2b.

3.2 Social Network

In contrast to interactions in home settings, social encounters are much more widespread, and we seek to capture this more connected, less clustered character in our extended social network. This means we cannot build our social sub-network by specifying all-to-all coupling within social units as we did for the home sub-network, and it makes realistic network extension more challenging. To address this, we base our social network construction on a simplified distribution that captures features of the Warwick data [6]. As shown in Figure 4b, the social degree distribution appears to be roughly uniform until degree 36, with a few outliers who have many friends. This observation underlies the



(a) (b)

Figure 3: (a) Plot of home degree versus average frequency of encounters shows that individuals with high degree (inhabitants of large households) interact with other home members less frequently on average, while members of small households encounter a smaller core group more frequently. (b) The distributions for interaction frequencies show good agreement between our algorithm and the Warwick data.

construction of our extended social network.

We generate the structure of the social network in three steps: first, we construct social units (or friend groups) of size 38. Within each identical social unit, the nodes are assigned a degree from 0 to 36 in a uniform manner to account for the roughly uniform distribution on Warwick social degrees ≤ 36. (Note that an inductive argument shows that there is a unique way of specifying a uniform distribution on 38 nodes and it necessarily forces degree 18 to appears twice). Next, to capture the appearance of social outliers with many friends, we randomly select popular nodes and connect them to high degree individuals in other social units. Lastly, we shuffle some connections between nodes to bring the average local clustering coefficient down to values reported in [13, 14], in analogy to the small world model of Watts and Strogatz [15].

Interaction weights are assigned to each edge in the same way as for the home network: we separate the measured frequency distribution [6] for social interactions into two distributions, one for nodes with degree ≤ 18, the other for degree > 18, leading to frequency distributions for the extended network that are in good agreement with the Warwick data (see Figure 4c). As in the case of home networks, the choice to split the frequency distribution into two components is based on the observation that individuals with high degree appear to have fewer repeat interactions on average.

Figures 4c and 4d show a comparison of the frequency and degree distributions in our extended network to the Warwick data [6]. We summarize the full social network algorithm below:

1. Generate 79 identical social building blocks of size 38 so that the degree distribution within each social unit is approximately uniform from 0 to 36 (degree 18 necessarily appears twice).
2. For each social unit, choose α of the β most popular (highest degree) nodes in the unit. Connect each of the chosen nodes to γ high degree nodes randomly selected from other social units, where high degree means one of the top β highest degree nodes in a social unit.

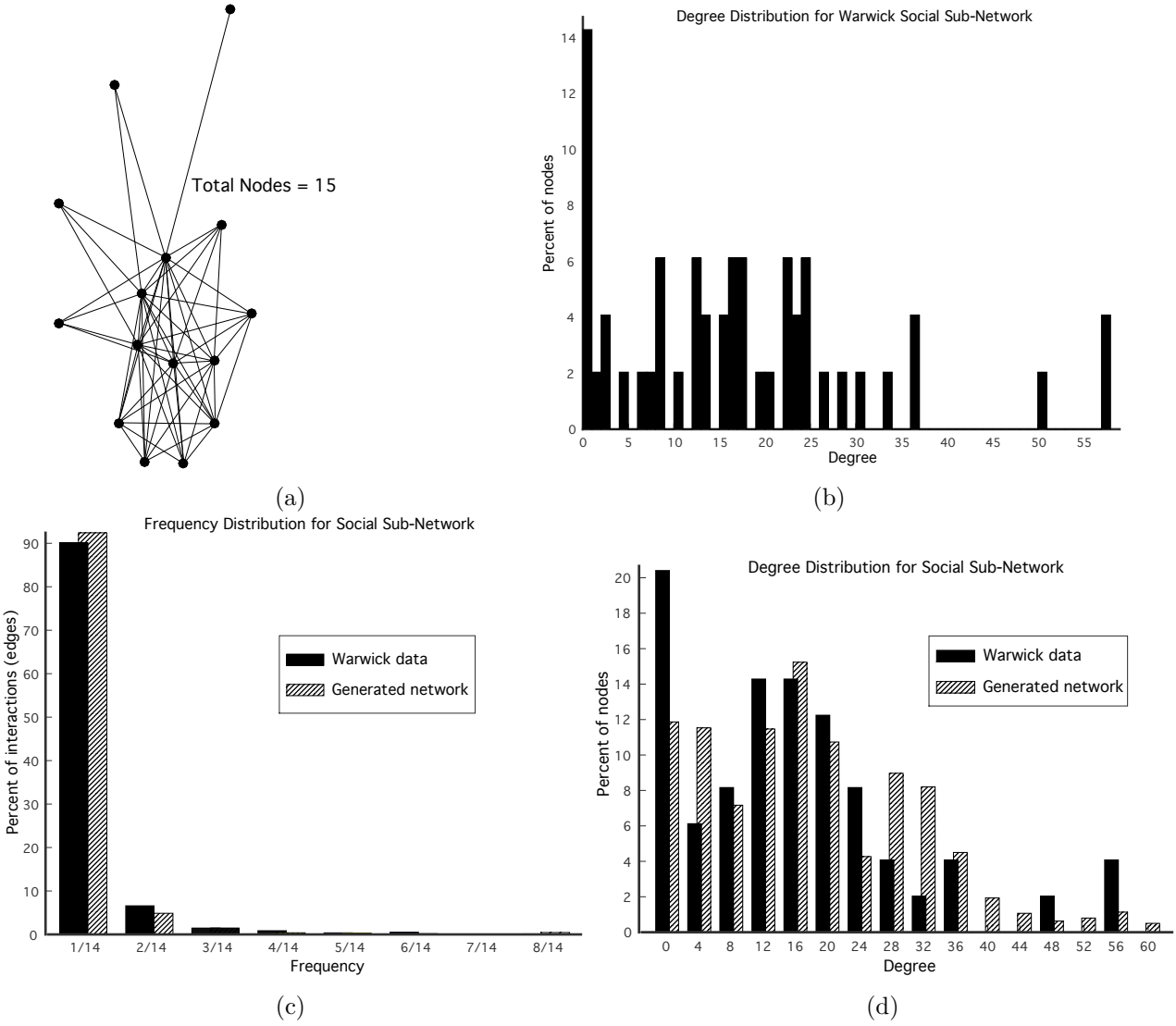


Figure 4: (a) Social units, each with a uniform degree distribution, serve as the basic building block of our extended social network. We show an example network of 15 nodes with a uniform distribution (our social units are each made up of 38 nodes, but a smaller network illustrates the structure more clearly as an illustrative example). (b) The social degree distribution for the Warwick data [6] is approximately uniform from degree 0 to 36. (c-d) Comparison of social frequency (c) and degree (d) distributions for the Warwick data [6] and our extended network.

3. Randomly select M edges and, for each such edge, disconnect one end and reconnect it to another node chosen at random.
4. Randomly select and remove 2 nodes to reduce the total network size to the target 3000 nodes.
5. Assign interactions weights to each edge by sampling from the social frequency distribution for the Warwick data [6].

Multiple values of α, β, and γ were tested, and those values that best fit the degree distribution of the Warwick data, namely $\alpha = 1$, $\beta = 5$, and $\gamma = 25$, were chosen. Similarly, $M = 13000$ was selected to reproduce the average local clustering coefficient $C = 0.16$ reported by Ahn *et al.* [14] for the Cyworld social network. (Note that we used reports [13, 14] of clustering in online communities to inform our network algorithm because it is difficult to formulate an accurate measure of clustering from [6]). We found that the fewer edges we randomly shuffled, the larger the clustering coefficient, with $C = 0.8$ for $M = 0$. Thus, tuning this parameter could allow for the generation of a social network with any given clustering coefficient.

In conclusion, we construct extended social networks in three steps to take into account the key features of the Warwick data and measurements [13, 14] of clustering in online social communities: the first step, building social units of uniform degree, captures the approximately uniform character of the Warwick social degree distribution. The second step, adding edges between randomly selected high degree members of social units, accounts for high-degree outliers in the Warwick data [6]. Lastly, reducing the level of structure in the network by breaking and reconnecting some edges at random brings the clustering coefficient down in agreement with empirically measured values.

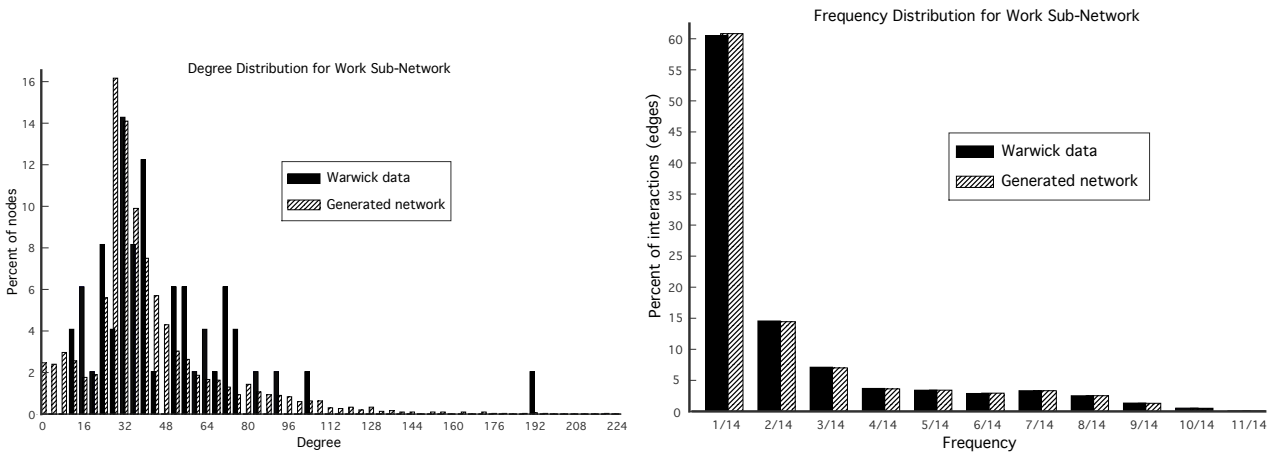
3.3 Work/School Network

Read *et al.* [6] found that the degree distribution for casual contacts across all contexts had a significantly longer right-tail than the corresponding distribution for skin-to-skin encounters. Because the majority (95.97% [6]) of encounters at work were casual, while most skin-to-skin interactions took place in home or social contexts, we expect that the work/school degree distribution [6] should display a longer right-tail character than the distributions for other contexts. Indeed, as shown in Figure 5a, the Warwick degree distribution for work/school encounters displays a long right-tail, a feature that is characteristic of power-law distributions and often captured using network growth models involving preferential attachment [16, 17].

Preferential attachment is a common means of generating networks with scale-free power-law distributions and was popularized by the work of Barabási and Albert [17]. According to the Barabási-Albert model, network growth occurs by starting with an initial network of m_0 nodes and then adding one node at a time. At each step, the new node is connected to $m \leq m_0$ other nodes, with the probability of connecting to node i given by

$$\Pi(k_i) = \frac{k_i}{\sum_{j=1}^{N} k_j}, \tag{1}$$

where k_i is the degree of the ith node and N is the total number of nodes in the network. This rule means that new nodes are most likely to connect to existing nodes of high degree, and the result is a



(a) (b)

Figure 5: (a-b) Degree and frequency distributions for the work network show good agreement between the Warwick data [6] and our extended model.

network structure in which many nodes are connected to a few very popular (high degree) individuals.

We chose to build our extended work/school network based on preferential attachment because of the structure of the work environment itself: one can think of a business scenario in which many employees interact with a common manager. Alternatively, in the school context, we would envision many students conversing with a few teaching assistants, and everyone interacting with a single course instructor. Thus, as we add nodes to the network, each individual will be more likely to connect to the instructor (node of high degree) than to any given student. The long right-tail of the real work/school degree distribution further supports our choice to base network extension on preferential attachment.

We tested three different implementations of network growth using the idea of preferential attachment. We began by building complete networks one node at a time according to the Barabási-Albert model [17], but this led to networks in which the degree distribution was too narrow. To remedy this problem, we also tried a variation of the Barabási-Albert model [17] that was motivated by the fitness model of Bianconi and Barabási [18]: to penalize high-degree nodes from receiving additional edges after they reach a given degree k_0, we modified the original probability in equation (1) to become

$$\Pi(k_i) = \frac{k_i \eta(k_i)}{\sum_{j=1}^{N} k_j}, \tag{2}$$

where $\eta(k) = \frac{1}{2}(1 - \tanh \frac{k-k_0}{\epsilon})$ is a cutoff function. Here, $\epsilon, k_0 > 0$ are selected to best fit the real data [6]. It should be noted that we have replaced η_i, a native fitness value for each node i that is chosen from a specified distribution in [18], with $\eta(k)$. While the original Barabási-Albert algorithm [17] and our altered version of the fitness model [18] are able to produce degree distributions with the observed long right-hand tail, neither method captures the high amount of clustering reported in the real data [6].

To raise the clustering coefficient, we returned to the idea of building networks out of units. Using 2011 data on establishment sizes from the United States Census Bureau (Statistics of US

Businesses) [19], we approximated appropriate unit sizes: a 3000-person network should have one company of size ≈ 625, five companies of size ≈ 300, and nine companies of size ≈ 60. The original Barabási-Albert algorithm [17] was then used to generate the degree distribution within each of the large businesses. To account for the remaining 335 nodes needed to make up a 3000-person network, we created small work/school units of size ≈ 20, and a roughly uniform distribution was specified within each unit. The choice to use a uniform degree distribution within the small businesses/classrooms was based on the idea that small settings allow for a more interactive structure than do larger ones; additionally, incorporating small work units of uniform degree into the network served to increase the amount of clustering. We summarize the details of the full work network algorithm below:

1. Generate one work unit made up of 625 nodes using the original Barabási-Albert model [17] with $m = m_0 = 30$. The value of m was chosen to best match the degree distribution of the real data [6].
2. Generate 5 work units, each of 300 nodes, according the Barabási-Albert model with $m = 30$.
3. Generate 9 work units, each of 60 nodes, according the Barabási-Albert model with $m = 30$.
4. Generate 17 small work units of 20 nodes each so that the degree distribution within each work/school unit is roughly uniform from 1 to 20 (degree 10 will appear twice).
5. Randomly select 5 nodes from the small work/school units to remove from the network, bringing the total network size to 3000 nodes.
6. Together the work/school units generated in Steps 1-5 represent the network. Assign an interaction weight to each edge in this network by sampling from the work frequency distribution for the Warwick data [6].

By combining the concept of preferential attachment in large businesses (or classrooms) with uniform degree distributions in small work units, we were able to generate an extended work/school network with degree and frequency distributions that capture many of the features of the Warwick data [6], as shown in Figures 5a and 5b.

4 Results

We now turn to a study of the spread of epidemics on our extended network generated as described in §3 from the Warwick data [6]. We assume a disease that gives long-term immunity after recovery, so that the susceptible-infected-recovered (SIR) model framework is appropriate [20]. Individuals can therefore be susceptible (S), infected (I) or recovered (R), and the infected individuals are assumed to be capable of infecting other susceptibles. The i^{th} susceptible node is denoted by S_i, and the j^{th} infected node is denoted by I_j. Each infected individual is assumed to recover from the disease after a time drawn from an exponential distribution with mean T, where T is the average duration of infectiousness.

To model disease transmission, we define the probability for a susceptible individual to become

infected per unit time to be

$$P(S_i \text{ becomes infected per unit time}) = \sum_{\text{infected neighbors } I_j} P(I_j \text{ infects } S_i \text{ per unit time})$$
$$= \frac{R_0}{T} \sum_{\text{infected neighbors } I_j} \frac{f_{ij}}{\bar{f}}\,, \tag{3}$$

where f_{ij} is the frequency of the interaction between S_i and I_j, $\bar{f}$ is the average frequency of pairwise interactions across the network (the average weighted network degree), and R_0 is the basic reproduction number of the disease (the average number of people infected by one infectious person over the course of the infection period T). As mentioned in §3, the frequency f_{ij} is a weight assigned to each edge in the network. This approach allows us to account for the frequency of interactions between individuals and their neighbors in various contexts. Disease spread is then simulated stochastically and the state (S, I, R) of each individual is updated at every time step, which we choose to be one day. In the following, we will refer to the fraction of infected individuals as a function of time as the epidemic size over time, defined as epidemic size $= \frac{I(t)}{N}$.

Since influenza offers long-term immunity and can be distributed through casual interactions [6], we test the spread of a flu epidemic using the discrete SIR model on our extended network. We consider a mean infection time of $T = 4.5$ days as given in [21, 22], and use $R_0 = 1.2515$, corresponding to the average of five seasons of flu surveillance data [23]; this value for the reproduction number is also consistent with estimates in [7, 22].

Unless noted otherwise, 0.0034% (10 nodes) of the 3000 individuals in the extended network are initialized as infected and provide the seed for disease spread, while the remaining individuals start as susceptibles. The 10 initially infected individuals consist of a chosen node and its neighbors; if the target of 10 infected individuals is not reached, we select the remaining nodes by sampling randomly from individuals connected to the neighbors of the originally infected seed. We discuss in §4.3 and Figure 9 how the choice of the initially infected node affects the spread of the disease.

4.1 Comparing discrete and continuous SIR models reveals the importance of accounting for context-specific network topology

The discrete SIR approach allows for direct comparison with the classical continuous SIR model [20]

$$\begin{aligned} \frac{dS}{dt} &= -\beta SI \\ \frac{dI}{dt} &= \beta SI - \gamma I \\ \frac{dR}{dt} &= \gamma I \end{aligned} \tag{4}$$

with the same parameters and initial conditions as for the discrete model. We assume a total population of constant size N, where $S(t)$, $I(t)$ and $R(t)$ have the same meaning as in the discrete model and correspond to the sizes of the susceptible, infectious and recovered populations, respectively, at time

t. Here, $\beta := \frac{R_0}{NT}$ is the number of new disease cases per unit time; and $\gamma := \frac{1}{T}$ represents the rate at which infected individuals recover from the disease [24].

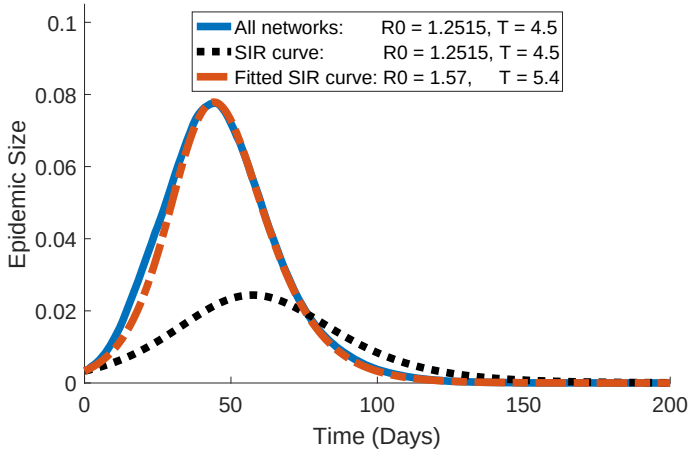


Figure 6: Epidemic size over time predicted by the discrete SIR model (3) on all the context networks considered (home, social, and work — solid line), compared with results of the SIR model (4) with the same R_0 and T parameters (dotted line) and with fitted parameters (dashed line).

We compare our model results with simulations of equation (4) for 200 days in Figure 6. This timescale allows the epidemic to peak as well as fully return to the equilibrium (a population with no infected individuals). The differential equation model for influenza leads to a considerably smaller peak epidemic size (size of the infected population) and a later onset of the disease compared to the discrete SIR model on networks. This means that the model (4) cannot account for the effects of complex home, social and work interactions on the progression of the disease. On the other hand, our discrete SIR model approach allows for testing the impact of different networks structures and context interactions of disease spread. It should be noted that the basic reproduction number R_0 and the mean infection time T in the continuous model can be chosen and fit so that the epidemic size over time closely resembles our discrete SIR model prediction (Figure 6). However, these parameters are different from those considered in our influenza simulations, suggesting that the classical SIR model may yield erroneous parameter estimates for R_0 and T when fitting realistic epidemic data.

4.2 Interaction context has a high impact on disease progression

Since the frequencies of interaction are key in the transmission probability formula (3), we investigate the contribution of different contexts to disease spread in Figure 7. The sub-networks detailed in §3 allow us to identify the impact of context-specific interactions. The horizontal lines in Figure 7 correspond to percent contributions of each context and are calculated as $\bar{f}_k/\bar{f}$, where k stands for the context (home, work/school, and social), and $\bar{f}_k$ is the average frequency of interaction in context k across the network. This measure thus depends on the network structure and frequency of interactions only. The scatter plots are obtained by simulating the discrete SIR model for influenza on the networks, and are calculated using the percent contribution infected individuals in each context make to the probability of infecting susceptibles at each time step. Figure 7 shows that the network predictions

and discrete SIR model contributions to infection agree fairly well. As expected, the comparison is no longer useful for analysis in the second half of the 200 days simulated, when the epidemics dies off (see Figure 6).

The work context has the highest contribution to disease spread: Figure 7 reveals that work network interactions influence disease spread the most. This is expected since the interactions in this sub-network are more frequent and likely to last longer than those in the social context. This result is also consistent with the fact that our method of extending the work sub-network in §3.3 renders the work units more clustered than the social ones.

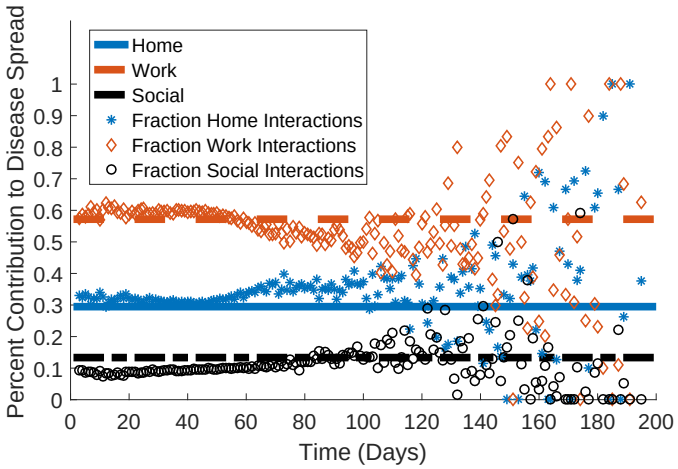


Figure 7: Percent contributions of each context to disease spread: horizontal lines are expected contributions given the network structure of frequency of interaction, and scatter plots are contributions to epidemics spread observed by simulating the discrete SIR model (3).

Excluding social interactions has the highest impact on disease transmission: We also simulate the spread of influenza on networks where we exclude certain context interactions. Figure 8 shows how the size of the infected population and the onset of disease are affected by eliminating social, work/school, or home networks, respectively. Removing connections through the social network influences disease dynamics most strongly, as it prevents the spread of the epidemic and considerably shortens its duration. This is not surprising given that social interactions are the only connections between the more clustered home and work networks. However, this result is clearly not reflected in Figure 7, where the social context has the lowest percent contribution among the networks considered. While there are fewer social interactions compared to work and home encounters, the social context enables disease spread across loosely connected clusters in the network, thus facilitating the epidemic.

4.3 Initial infection seed does not considerably impact disease spread

As mentioned in §4.1, we simulate the spread of the disease by starting with an infected node together with its infected neighbors. The choice of the initially infected individual is explored in Figure 9. As shown, assigning a random, maximum, or minimum degree node in the network as the initial seed does not significantly impact the epidemic size over time. Infecting a randomly-chosen maximum degree individual causes the disease to spread sooner, but does not alter the duration or the severeness of the

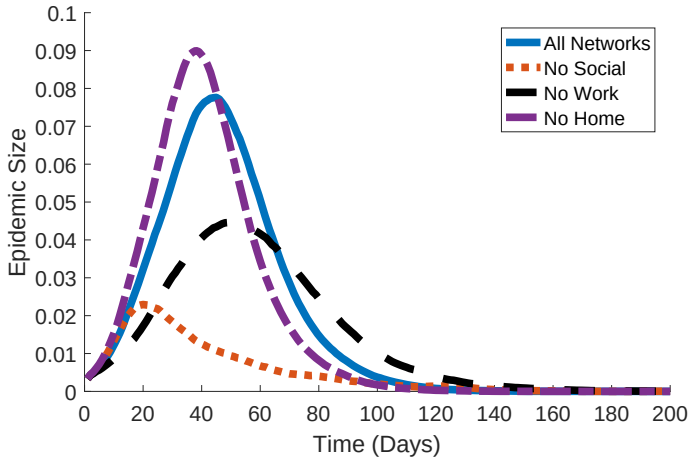


Figure 8: Epidemic size over time predicted by the discrete SIR model (3) using all the context networks considered (solid line), compared with removing social interactions (dotted line), removing work interactions (dashed line), and removing home interactions (dash-dotted line).

epidemic considerably. Therefore, the rest of the simulations of the discrete SIR model are carried out with a randomly chosen initial infected individual.

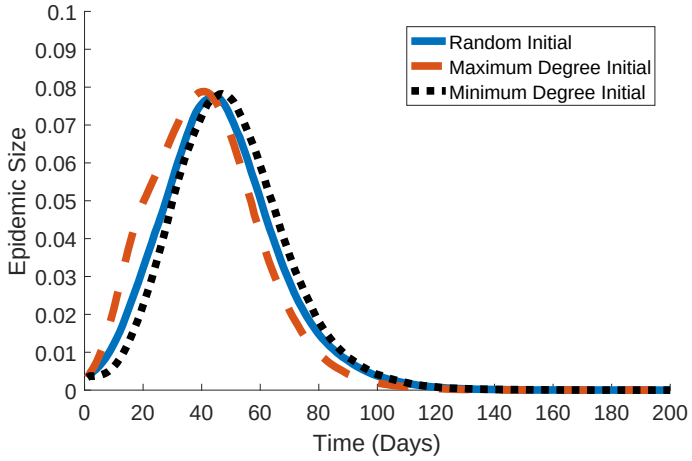


Figure 9: Epidemic size over time predicted by the discrete SIR model (3) using all the context networks starting with an initially infected seed chosen randomly (solid line), with maximum degree (dashed line), and with minimum degree (dotted line).

4.4 Dynamic responses to disease substantially reduce epidemic size

The network-based discrete SIR model allows us to test dynamic responses of the population after the onset of the disease. We consider a few realistic reactions to the onset of influenza, such as scenarios where home interactions become more frequent, while social and work interactions are reduced following infection. We model the disease responses by lowering the frequencies of interactions of individuals 1-2 days after they become infected.

Considerably reducing interactions at work leads to smaller epidemic size and duration of infection: The effect of reducing interactions through different networks is shown in Figure 10. Reducing social interactions to a large extent decreases the size of the epidemic, but predicts a similar or slightly increased epidemic duration (dash-dotted and starred curves, Figure 10). On the other hand, decreasing the frequency of work context interactions to a significant degree yields a smaller epidemic size and duration of infection (dashed and dotted curves, Figure 10). This is expected given the network structure considered: reducing frequent work interactions does not allow the spread of the infection inside work clusters; this in turn limits the extension to other network clusters through occasional social encounters.

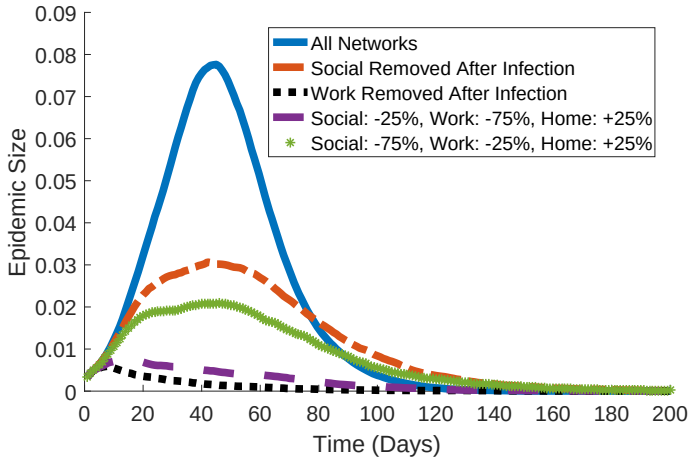


Figure 10: Epidemic size over time predicted by the discrete SIR model (3) using all the context networks and incorporating various dynamic responses 1-2 days after disease onset.

This observation is also suggested by Figure 11, where we plot the percent contributions to infection through each context in the situation where work interactions are removed completely after infection onset. Compared to Figure 7, the work and home percent contributions switch, with home clusters becoming the most influential in disease spread. The social contribution increases to a small extent, but only a few of these interactions are likely to spread the disease, as individuals do not interact in work clusters and can thus no longer spread the infection through social interactions as well. The high degree of variance of context contributions in this figure is due to the small epidemic size over time under this response scenario (dotted curve, Figure 10).

4.5 Model results capture trends in CDC seasonal epidemic data

We compare the epidemic size over time curves predicted by our discrete SIR model with data from the Center for Disease Control and Prevention (CDC) [25]. Figures 12 and 13 overlay information on patient visits for influenza-like illnesses with curves of the epidemic dynamics given different responses to the epidemic. Since this seasonal data [25] varies in shape every year, we choose sample representative curves for the percentages of flu-related visits for plotting. It should be noted that all curves are shifted on the time axis so that the weeks when the epidemic size peaks are aligned. We also shift the

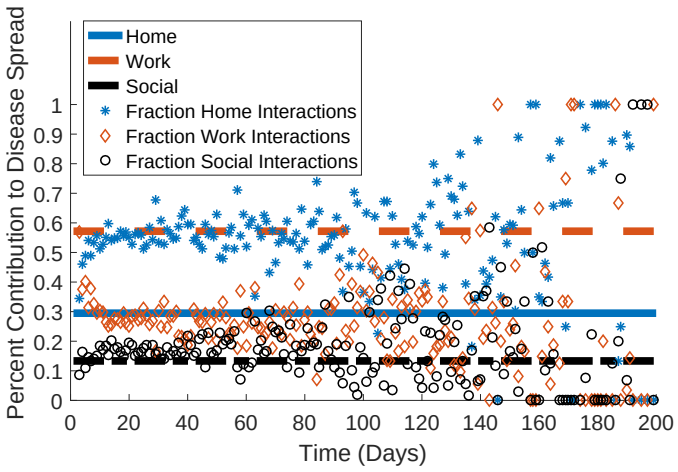


Figure 11: Percent contributions of each context to disease spread: horizontal lines are as in Figure 7, and scatter plots are contributions to epidemic spread observed by simulating the discrete SIR model (3) under the scenario that individuals remove work interactions after infection.

discrete SIR model results (solid, dashed, and dotted curves in Figure 12 and solid & dotted curves in Figure 13) up on the epidemic size axis by a constant 0.01: this is because the SIR model has a stable equilibrium at $I = 0$ (no infected individuals), while realistic outbreak distribution data for different influenza seasons shows a background epidemic size of roughly 1% of the population even outside the peak epidemic weeks. These shifts do not affect our comparisons, since we are primarily interested in results on the peak epidemic size and the epidemic duration.

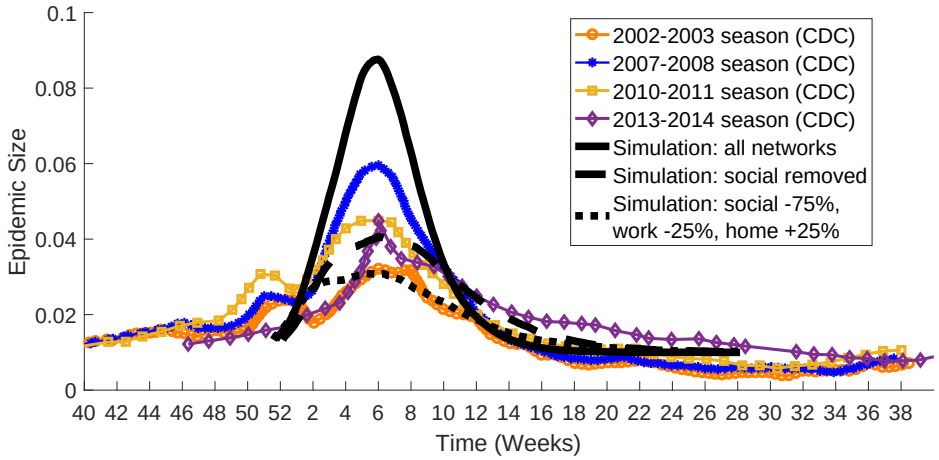


Figure 12: Epidemic size over time predicted by the discrete SIR model (3) using various dynamic responses 1-2 days after disease onset (solid, dashed, and dotted lines), and flu visit data reported by the Center for Disease Control and Prevention [25] for different moderate to moderately severe influenza seasons.

Reductions in social encounters and work interactions yield epidemic dynamics that compare well with moderate and mild flu season data, respectively: Figure 12 shows that our simulation of influenza spread across our full extended network with no dynamic responses to disease onset included may be

unrealistic. The peak epidemic size predicted in this situation is larger than the moderately severe flu season 2007-2008 [26]. On the other hand, incorporating substantial changes in the social context interactions predicts epidemic size over time curves that closely resemble the moderate seasons 2010-2011 and 2002-2003 [25]. Similarly, including large changes in the pattern of work interactions leads to a good prediction of the start of the epidemics during mild seasons (see Figure 13). Our model therefore suggests that strategies involving reductions of work interactions after onset of the flu may have the largest impact in avoiding severe influenza seasons. The results also indicate that, as expected, individuals are likely to change their social and work interactions shortly after they contract the flu.

We note that we do not expect to recreate the double-peak curves observed in some of the seasons represented on Figures 12 and 13, since the dynamic responses in our model are not influenced by factors such as cognition of the epidemic spread (as in [7]). In that study, the authors use a heterogeneous graph modeling approach to describe flu virus transmission in a population of hospital patients, and an agent-based model incorporating many unknown parameters to model the dynamic change in individuals' interactions as a reaction to the epidemic. Our SIR model approach on extended networks distinguishes itself from [7] by being less computationally costly and requiring no parameter estimations as well as fewer assumptions on individuals' cognitive and protective behavior during an epidemic. This minimal model is able to reproduce key features of the first peak in real influenza epidemics, such as size and duration of the outbreak.

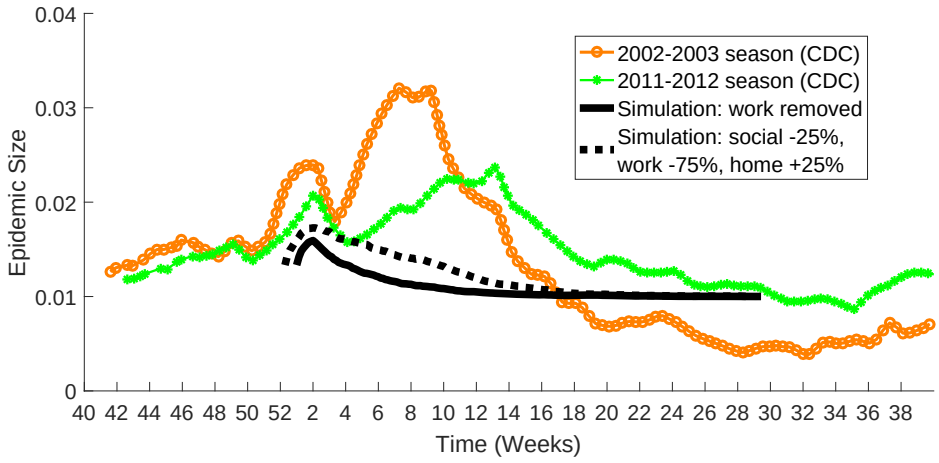


Figure 13: Epidemic size over time predicted by the discrete SIR model (3) using various dynamic responses 1-2 days after disease onset (solid and dotted lines), and flu visit data reported by the Center for Disease Control and Prevention [25] for different mild to moderate influenza seasons.

5 Discussion

Exploring human network structure is essential to understanding how epidemics propagate and how they can be contained before further spreading to other communities. However, knowledge of human interactions at the population level is difficult to obtain given the challenges imposed by large-scale data collection [1]. In this work, we proposed a method for extending data from the diary-based study [6] to

construct larger networks for the interactions of individuals in home, social and work/school settings. The methods detailed in §3 are based on building context-specific networks that take into account the intrinsic differences in the structure of interactions that occur in these settings. The extensions reflect the specific degree and cluster distributions revealed in [6] and use the interaction frequencies as weights for the network edges.

The spread of influenza was then tested on the extended network using the discrete SIR model (3) in §4. Our results show that the classical differential equation SIR system with the same chosen influenza parameters is unable to reproduce the epidemic size results of the discrete SIR model (Figure 6); this suggests that accounting for network structure and frequency of interactions is crucial in understanding real-world disease transmission. Our network model also predicts that the home, social, and work sub-networks have significantly different effects on the epidemic dynamics (Figures 7 and 8). In particular, we find that while social interactions are less frequent and account for a small percent of infections, they can greatly facilitate disease spread by providing connections between work and home clusters.

Realistically, individuals often choose to reduce various interactions after contracting an infectious disease. Accounting for such dynamic responses in our discrete SIR model yields predictions of epidemic size over time that agree with the peak of the epidemic data [25] for moderate or mild flu seasons (Figures 12 and 13). One limitation of our model is that it cannot recover the double-peak epidemic size displayed in several flu seasons as this would likely require knowledge of how the dynamic response to the disease varies with time and epidemic size [7]. On the other hand, the advantage of the minimal model proposed here is that it requires no costly parameter fitting or assumptions about dynamical responses to disease onset, yet it can still reproduce the size and duration of the first peak in the epidemic.

The discrete epidemic spread model can also be applied to diseases that do not confer long-term immunity (such as bacterial meningitis). The spread of these diseases is simulated using the susceptible-infectious-susceptible (SIS) model, and the probability of transmission is defined as in (3). Our results for meningitis show that the epidemic size reaches an equilibrium after about 150-250 simulated days, and that the dynamic evolution compares well to meningitis outbreak data from the World Health Organization [27]. Similar to the influenza prediction, changes in the interaction behavior of individuals leads to significant reductions in the peak epidemic size (results not shown).

A small percentage of the interactions recorded in the Warwick data [6] occurred in the context of shopping and travel. We also tested the discrete SIR model on networks that include these interactions, which are more likely to take place during weekends as suggested in the Warwick data. The extension of these sub-networks to a larger population is based on sampling from the travel and shop degree distributions, as well as specifying all-to-all coupling of nodes in clusters (based on the idea that groups of individuals traveling or shopping together are small and fully clustered). The epidemic size over time given these complete networks is almost identical to the full network results in Figure 6 (results not shown), so we chose not to include these contexts in our main results. However, it would be interesting to consider these sub-networks in future simulations of disease spread across several communities generated as in §3. This approach could be used to study the speed of disease spread across cities, as well as to identify and test strategies for isolating an epidemic. Furthermore, the model

proposed could be extended by differentiating between proximity of interactions in the Warwick data [6] (casual or skin-to-skin). This would allow for a comparison of diseases that spread through casual interactions with those that require closer contact between individuals, and would provide insight into the variability in transmission dynamics for different infections.

Competing Interests We have no competing interests.

Authors' Contributions JA, AS, VC, AV, and BS constructed the model and analyzed results; JA and AS carried out simulations. VC and AV drafted the manuscript. All authors gave final approval for publication.

Acknowledgments Abrams, Schwartz, and Volkening were supported by the NSF through grant DMS-1148284. Ciocanel and Sandstede were supported by the NSF under grant DMS-1408742. We are very grateful to Prof. Edmunds for providing us with the anonymized survey data published in [6].

References

[1] Keeling MJ, Eames KT. Networks and epidemic models. Journal of the Royal Society Interface. 2005;2(4):295–307. doi:10.1098/rsif.2005.0051.

[2] Zheng X, Zhong Y, Zeng D, Wang FY. Social influence and spread dynamics in social networks. Frontiers of Computer Science. 2012;6(5):611–620. doi:10.1007/s11704-012-1176-1.

[3] House T, Ross JV, Sirl D. How big is an outbreak likely to be? Methods for epidemic final-size calculation. In: Proceedings of the Royal Society of London A. vol. 469. The Royal Society; 2013. p. 20120436. doi:10.1098/rspa.2012.0436.

[4] House T. Modelling epidemics on networks. Contemporary Physics. 2012;53(3):213–225. doi:http://dx.doi.org/10.1080/00107514.2011.644443.

[5] Tao Z, Zhongqian F, Binghong W. Epidemic dynamics on complex networks. Progress in Natural Science. 2006;16(5):452–457. doi:10.1080/10020070612330019.

[6] Read JM, Eames KT, Edmunds WJ. Dynamic social networks and the implications for the spread of infectious disease. Journal of The Royal Society Interface. 2008;5(26):1001–1007. doi:10.1098/rsif.2008.0013.

[7] Guo D, Li KC, Peters TR, Snively BM, Poehling KA, Zhou X. Multi-scale modeling for the transmission of influenza and the evaluation of interventions toward it. Scientific Reports. 2015;5. doi:doi:10.1038/srep08980.

[8] Newman M. The Structure and Function of Complex Networks. SIAM Review. 2003;45(2):167–256. doi:10.1137/S003614450342480.

[9] Gross T, D'Lima CJD, Blasius B. Epidemic dynamics on an adaptive network. Physical Review Letters. 2006;96(20):208701. doi:https://doi.org/10.1103/PhysRevLett.96.208701.

[10] Böhme GA. Emergence and persistence of diversity in complex networks. The European Physical Journal Special Topics. 2013;222(12):3089–3169. doi:10.1140/epjst/e2013-02078-7.

[11] Siettos CI, Russo L. Mathematical modeling of infectious disease dynamics. Virulence. 2013;4(4):295–306. doi:10.4161/viru.24041.

[12] Riley S. Large-scale spatial-transmission models of infectious disease. Science. 2007;316(5829):1298–1301. doi:10.1126/science.1134695.

[13] Hardiman SJ, Katzir L. Estimating clustering coefficients and size of social networks via random walk. In: Proceedings of the 22nd international conference on World Wide Web. International World Wide Web Conferences Steering Committee; 2013. p. 539–550. doi:10.1145/2790304.

[14] Ahn YY, Han S, Kwak H, Moon S, Jeong H. Analysis of topological characteristics of huge online social networking services. In: Proceedings of the 16th international conference on World Wide Web. ACM; 2007. p. 835–844. doi:10.1145/1242572.1242685.

[15] Watts DJ, Strogatz SH. Collective dynamics of 'small-world' networks. Nature. 1998;393(6684):440–442. doi:10.1038/30918.

[16] Albert R, Barabási AL. Statistical mechanics of complex networks. Reviews of Modern Physics. 2002;74(1):47. doi:https://doi.org/10.1103/RevModPhys.74.47.

[17] Barabási AL, Albert R. Emergence of scaling in random networks. Science. 1999;286(5439):509–512. doi:10.1126/science.286.5439.509.

[18] Bianconi G, Barabási AL. Competition and multiscaling in evolving networks. EPL (Europhysics Letters). 2001;54(4):436. doi:http://dx.doi.org/10.1209/epl/i2001-00260-6.

[19] United States Census Bureau. Number of Firms, Number of Establishments, Employment, and Annual Payroll by Large Enterprise Employment Sizes for the United States and States, NAICS Sectors: 2011;. Http://www.census.gov/data/tables/2011/econ/susb/2011-susb-annual.html.

[20] Kermack WO, McKendrick AG. A contribution to the mathematical theory of epidemics. In: Proceedings of the Royal Society of London A: Mathematical, physical and engineering sciences. vol. 115. The Royal Society; 1927. p. 700–721. doi:10.1098/rspa.1927.0118.

[21] Longini IM, Halloran ME, Nizam A, Yang Y. Containing pandemic influenza with antiviral agents. American Journal of Epidemiology. 2004;159(7):623–633. doi:10.1093/aje/kwh092.

[22] Tuite AR, Greer AL, Whelan M, Winter AL, Lee B, Yan P, et al. Estimated epidemiologic parameters and morbidity associated with pandemic H1N1 influenza. Canadian Medical Association Journal. 2010;182(2):131–136. doi:10.1503/cmaj.091807.

[23] Zhang S. Estimating transmissibility of seasonal influenza virus by surveillance data. Journal of Data Science. 2011;9:55–64. doi:10.6339/JDS.2011.09(1).664b.

[24] Ellner SP, Guckenheimer J. Dynamic models in biology. Princeton University Press; 2011.

[25] Centers for Disease Control, Prevention (CDC). Percentage of visits for Influenza-like Illness (ILI), National Summary;. Reported by: U.S. WHO/NREVSS Collaborating Laboratories and ILINet, http://gis.cdc.gov/grasp/fluview/fluportaldashboard.html.

[26] Centers for Disease Control, Prevention (CDC). Percentage of visits for Influenza-like Illness (ILI);. Reported by: The U.S. Outpatient Influenza-like Illness Surveillance Network (ILINet), Weekly National Summary, 2012-13 and Selected Previous Seasons, http://www.cdc.gov/flu/weekly/weeklyarchives2012-2013/picILI52.htm.

[27] World for Health Organization (WHO). WHO-Multi-Disease Surveillance Centre Ouagadougou, Regional Meningitis Surveillance;. Http://www.who.int/csr/disease/meningococcal/epidemiological/en/.