Submitted to the Annals of Applied Statistics

SUPPLEMENT A TO "ESTIMATING WITHIN-SCHOOL CONTACT NETWORKS TO UNDERSTAND INFLUENZA TRANSMISSION": DESCRIPTIVE ANALYSES OF SIMULATED CONTACT NETWORKS

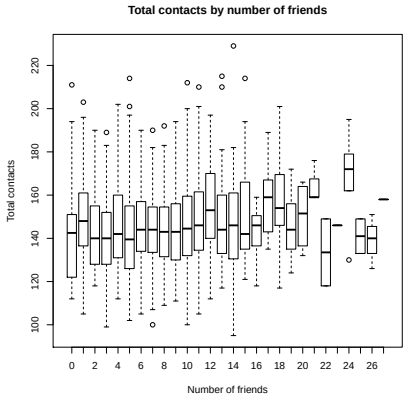
By Gail E. Potter, Mark S. Handcock, Ira M. Longini, Jr., and M. Elizabeth Halloran

In this supplement, we describe and assess properties of networks simulated from our model. We compare our fitted degree distribution to diary-based contact reports in an alternate data source, the POLYMOD study. Next, we compare properties of contact networks simulated from our model to reports in the epidemic survey (Xia et al., 2010).

1. Validation of contact degree distribution. We validated our fit of the degree distribution by comparing our fitted distribution to diary-based reports of contacts collected in a separate survey, the Belgian POLYMOD study (Mossong et al., 2008). In this study, respondents carried paper diaries for a 24-hour period and recorded details of every face-to-face contact they made: duration, frequency, location, age, sex, etc. This survey design may be less prone to measurement error than the epidemic survey since respondents are asked to enumerate each contact in a single day rather than estimate average patterns. The Belgian POLYMOD study includes 38 diaries of school contacts on non-holiday weekdays from respondents aged 13-18, the age group relevant for our model. Durations are reported as 0-5 minutes, 5-15, 15-60, 1-4 hours, or 4+ hours. To compare them to the contacts in our model, we count each contact coded as 0-15 minutes as 10 minutes of contact, those coded 15-60 minutes 40 minutes, those coded 1-4 hours as 2.5 hours, and those 4+ hours as 5 hours. This results in an average of 194 contacts per student per day, which is in line with our average of 148 per student per day. The duration recorded in POLYMOD is total duration of contact with the person over the entire 24-hour period in all locations, so includes contact outside of the school. Therefore, we expect the POLYMOD duration to be longer. In addition, since the data is from Belgium, we expect different contact patterns due to differences in cultural norms and school schedules. The POLYMOD survey does not include friendship data.

2. Descriptive analysis of a single contact network. Networks simulated from our model have an average density of 3.3%: that is, contact

FIG 1. *Total number of contacts by number of friends*



occurs in 3.3% of all possible pairs of people on a given day. The duration of contact ranges from 10 minutes to the maximum possible duration of 6 hours and 20 minutes, with a median of 30 minutes. On a given day, approximately 21% of dyads making contact are friends, but 56% of contacts occur to friends. This is because each pair can have multiple contacts, and friends tend to make more contacts per day in our model than non-friends. Contact networks simulated from our model are completely connected, which means that any two nodes in the network are connected by a path of edges. The friendship network is completely connected except for 26 isolates, i.e., students who have no friends.

The plots in this section show relationships from a single simulated network; we found these relationships to hold in other simulated networks. Under our model, the total number of contacts is slightly positively related to the number of friends, as shown in Figure 1. This relationship naturally arises from our use of number of friends as a predictor in modeling the degree distribution. The relationship is weak. The average number of contacts is 148; it is 144 for students with fewer than five friends and 156 for students with more than 15 friends. Students with more friends have higher percentages of contacts to friends, as shown in Figure 2 (left). This is in line with reports in the epidemic survey, in which the reported percentage of contacts to friends is positively related to the reported number of friends (Figure 2, right).

Students generally contact all or most of their friends in a given school day, as shown by Figure 3. The positive relationship between number of friends and number of friends contacted is reasonable. However, the total

FIG 2. *Percentage of contacts to friends by number of friends in simulated contact network (left), and in the epidemic survey (right)*

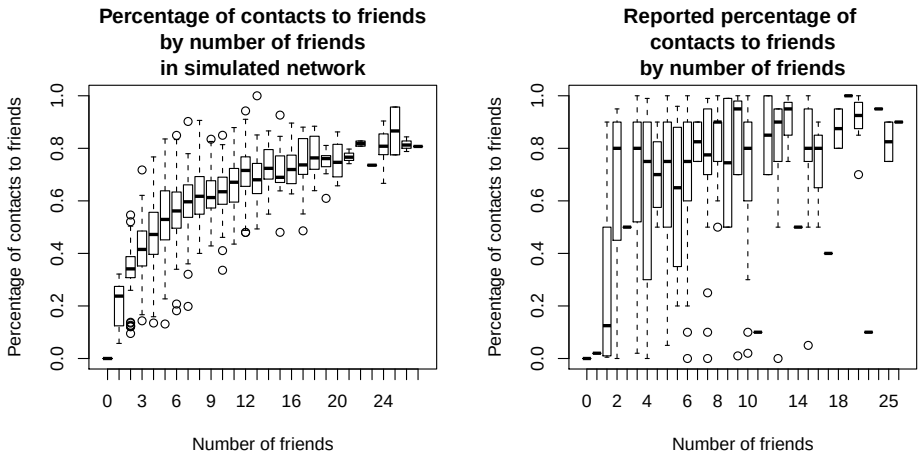


FIG 3. *Number of contacts to friends by number of friends*

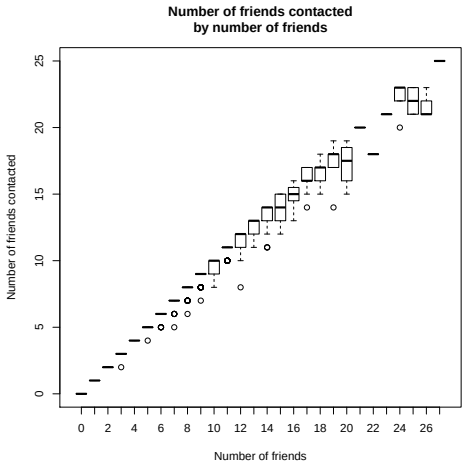
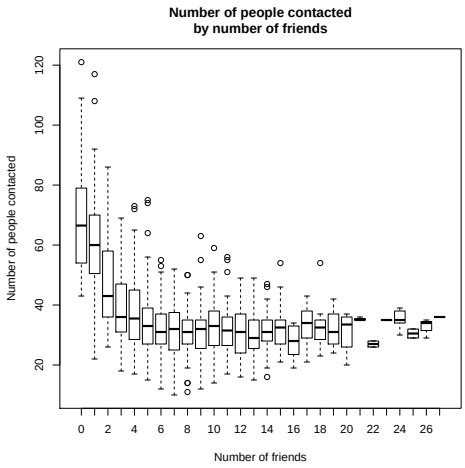


FIG 4. *Number of contacts to friends by number of friends*



number of people contacted (both friend and non-friend) is negatively related to the number of friends a student has, as shown in Figure 4. This is because popular students are likely to make large numbers of contacts to their friends, leaving only a few contacts to be distributed among non-friends. Students with fewer friends also make large numbers of contacts to their friends but have more contacts left over, which are then distributed among a large number of non-friends.

3. Descriptive analysis of network dynamics. The contact networks are dependent over time for two reasons. First, the class network is the same from one day to the next. Second, although the lunch and break networks are sampled independently each day, they rely on the same underlying friendship network, so are dependent.

We can quantify the changes in the contact network from day to day by considering the percentage of dyads whose contact status (contact vs. no contact) differs from one day to the next. On average and with little variation, this percentage is 3.1%. Since the density of the network is 3.3% and is similar from one day to the next, the change occurs because half of the people in contact on one day are not in contact the following day. Instead, a similar number of pairs who didn't contact each other on day 1, contact each other on day 2. These changing contacts occur during class breaks and at lunch, and nearly all (97%) of them have duration 10 minutes. Change is also manifested in the change in values (numbers of 10-minute contacts) on specific dyads that make contact on both days. On average and again with

very little variation, 66% of dyads which have contact on two consecutive days make the same duration of contact, 10% differ by 10 minutes, 6% by 20 minutes, 5% by 30 minutes, and the rest by 40 or more minutes.

We calibrated the random mixing model so that the expected number of people contacted per day is the same as in the friendship-based contact network model (36), and all contacts have equal duration: the average duration of contacts in the friendship-based contact model (41 minutes). Under random mixing, the percentage of dyads whose contact status (contact vs. no contact) differs from one day to the next is 6.2%. Since the density of the network is 3.3%, nearly all people in contact one day are not in contact the following day. Since all contacts are the same duration in this scenario, there is no change in duration of contact between two people who make contact on two consecutive days. While most of the changing contacts in the network scenario are only 10 minutes in duration, all of the changing contacts in the random mixing scenario are 41 minutes in duration.

References.

Mossong, J., Hens, N., Jit, M., Beutels, P., Auranen, K., Mikolajczyk, R., Massari, M., Salmaso, S., Tomba, G. S., Wallinga, J., Heijne, J., Sadkowska-Todys, M., Rosinska, M. and Edmunds, W. J. (2008). Social Contacts and Mixing Patterns Relevant to the Spread of Infectious Diseases. *PLoS Medicine* **5** 0381-0391.

Xia, H., Chen, J., Marathe, M. V. and Mortveit, H. S. (2010). Synthesis & Embedding of Subnetworks for Individual-based Epidemic Models. *NDSSL Technical Report 10-139.*