

# 1 An algorithm to build synthetic temporal contact 2 networks based on close-proximity interactions data

3 **Short title: An algorithm to build synthetic contact networks from contact data**

4 **Audrey Duval<sup>1,2,3+</sup>, Quentin Leclerc<sup>1,2,4+</sup>, Didier Guillemot<sup>1,2</sup>, Laura Temime<sup>4,5#</sup>, Lulla  
5 Opatowski<sup>1,2#</sup>**

6 + these authors contributed equally

7 # these authors contributed equally

8 <sup>1</sup>Epidemiology and Modelling of Bacterial Escape to Antimicrobials, Department of Global  
9 Health (EMEA), Université Paris Cité, Institut Pasteur, Paris, France

10 <sup>2</sup>Echappement aux Anti-infectieux et Pharmacoépidémiologie U1018, CESP, INSERM,  
11 Université Paris-Saclay, Université de Versailles St-Quentin-en-Yvelines, France

12 <sup>3</sup>Imagine Institute, Data Science Platform, INSERM UMR 1163, Université de Paris, Paris,  
13 France

14 <sup>4</sup>Laboratoire Modélisation, Epidémiologie et Surveillance des Risques Sanitaires (MESuRS),  
15 Conservatoire National des Arts et Métiers, Paris, France

16 <sup>5</sup>Unité PACRI, Institut Pasteur, Conservatoire National des Arts et Métiers, Paris, France

17

18 **Keywords:** long-term care facility, contact network, close-proximity interactions, sensors,  
19 network reconstruction

20

## 21 **Acknowledgments**

22 AD, LT and LO received funding from the French National Research Agency (SPHINX-17-CE36-  
23 0008-01).

## 24 **Abstract**

25 Small populations (e.g., hospitals, schools or workplaces) are characterised by high contact  
26 heterogeneity and stochasticity affecting pathogen transmission dynamics. The increased  
27 availability of empirical individual contact data provides unprecedented information to  
28 characterize such heterogeneity. However, these detailed data are usually collected over a  
29 limited period, and can suffer from observation bias. We propose an algorithm to  
30 stochastically reconstruct realistic temporal networks from individual contact data in health  
31 care settings (HCS) and test this approach using real data previously collected in a long-term  
32 care facility (LTCF).

33 Our algorithm generates full networks from recorded close-proximity interactions, using  
34 hourly inter-individual contact rates and information on individuals' wards, the categories of  
35 staff involved in contacts, and the frequency of recurring contacts. It also provides data  
36 augmentation by reconstructing contacts for days when some individuals are present in the  
37 HCS without having contacts recorded in the empirical data. Recoding bias is formalized  
38 through an observation model, to allow direct comparison between the augmented and  
39 observed networks.

40 The algorithm successfully reconstructed unobserved contacts, and was substantially more  
41 accurate to reproduce network characteristics than random graphs. The reconstructed  
42 networks reproduced well the assortativity by ward (first–third quartiles observed: 0.54–0.64;  
43 synthetic: 0.52–0.64) and the hourly staff and patient contact patterns. Importantly, the  
44 observed temporal correlation was also well reproduced (0.39–0.50 vs 0.37–0.44), indicating  
45 that our algorithm could recreate a realistic temporal structure. The daily degree (10.8–14.7  
46 vs 14.7–18.0), density (0.07–0.08 vs 0.08–0.10) and global efficiency (0.39–0.43; 0.46–0.49)  
47 were slightly overestimated.

48 To conclude, we propose an approach to generate realistic temporal contact networks and  
49 reconstruct unobserved contacts from summary statistics computed using individual-level  
50 interaction networks. This could be applied and extended to generate contact networks to  
51 other settings using limited empirical data, to subsequently inform individual-based epidemic  
52 models.

## 53 Introduction

54 Limiting the public health burden of infectious diseases requires a good understanding of how  
55 they spread. For diseases transmitted mostly via close-proximity interactions, the rate at  
56 which individuals come into contact with each other is strongly correlated with the expected  
57 spread of the disease across the population [1]. In large populations such as cities or countries,  
58 contact structures can be approximated by grouping individuals into relatively broad  
59 categories (neighbourhood, age...), and assuming that contact rates are heterogeneous  
60 between categories, but homogeneous within [2,3]. In small populations such as healthcare  
61 institutions, schools, or workplaces however, disease transmission is affected by high contact  
62 heterogeneity and stochasticity [4]. Capturing these characteristics requires a detailed,  
63 individual-level description of contacts.

64 Contact networks are increasingly used to fully capture the interactions between individuals  
65 in small populations [5,6]. These networks explicitly represent the links between all individuals  
66 in such populations, as opposed to contact matrices which only capture average contact rates  
67 between groups of individuals [7,8]. Temporal contact networks further capture the time-  
68 changing nature of contacts, therefore representing individual interactions more accurately  
69 than static networks [9–13]. Contact networks can be coupled with individual-based  
70 mathematical models to help design effective interventions against the spread of infectious  
71 diseases, since they enable the identification of highly connected individuals who can be  
72 targeted to lead to the greatest impact on transmission [8]. Recently, empirical data collected  
73 to build inter-individual temporal networks has become increasingly available. For example,  
74 previous studies have used sensors to record close-proximity interactions between individuals  
75 [14,15], and contact tracing programs have relied on the integrated Bluetooth technology in  
76 mobile phones [16].

77 However, the detailed empirical data required to build temporal contact networks remain  
78 subject to several limitations [17,18]. Even studies designed to collect such data are usually  
79 limited in time, and may be subject to observation bias; sensors might not be properly placed  
80 to register contacts [19], or individuals may disable Bluetooth on their mobile phones at  
81 different times [16]. Due to the resulting missed contacts, the networks derived from these  
82 data may only be partially observed. Transmission rates estimated using these partially

83 observed networks would be overestimated compared to reality due to the lower number of  
84 contacts, which could lead to an incorrect evaluation of the impact of interventions [20–22].  
85 By comparison, although they do not provide individual-level information, contact matrices  
86 and summary statistics such as contact rates between individual groups are more readily  
87 available, as they can be inferred using simple cross-sectional survey data [23–25].

88 Here, we propose an algorithm to stochastically reconstruct realistic contact networks from  
89 partially observed contact data. To validate this approach, we use close-proximity data  
90 collected in a long-term care facility (LTCF) during the i-Bird study [1]. We compute summary  
91 contact parameters from these data to generate reconstructed contact networks and compare  
92 these synthetic contact networks with the observed i-Bird data.

## 93 **Methods**

### 94 **Data description**

95 The data used here were collected during the Individual-Based Investigation of Resistance  
96 Dissemination (i-Bird) study [1]. This study took place in a rehabilitation and long-term care  
97 facility (LTCF) from the beginning of July to the end of October 2009. Over this period, each  
98 participant (patient or hospital staff) was wearing an RFID sensor that recorded close-  
99 proximity interactions (CPIs, at less than 1.5m) every 30 seconds. Here, we only used contacts  
100 recorded between 27 July to 23 August 2009 (included). This period corresponds to the weeks  
101 between two sensor battery replacements and hence avoids interference due to loss of  
102 contact. A temporal network of proximities was therefore available over 28 days with  
103 information on individual ID and ward of affectation.

104 The hospital was structured into five wards: three neurological wards, one nutritional care  
105 ward and one geriatric ward. Patients were systematically linked to a ward, whilst some staff  
106 were mobile and not linked to a specific ward. For the purpose of this work, we consider here  
107 that mobile staff belong to an “artificial” 6<sup>th</sup> ward, to compute contact rates according to the  
108 algorithm detailed below. Staff were divided into 13 professions: administrative,  
109 animation/hairdresser, logistic, hospital service agent, porter, occupational therapist,  
110 physiotherapist, other re-education, nurse, head nurse, care assistant, student/intern, and  
111 physician. A total of 200 patients and 213 hospital staff were included and had contacts  
112 recorded during the 28 days of study.

113 We used hospital staff schedules to determine the hourly presence of each staff. We  
114 compared these schedules, as well as admission and discharge dates of patients, to the dates  
115 and times when individuals had any contact recorded. Through this, we estimated that there  
116 was on average no contact data recorded for 37% (standard deviation: 30%) of a patient’s  
117 presence days, and 42% (sd: 38%) for staff (Supplementary Figure 1). The raw i-Bird network,  
118 measured directly from the sensors, is therefore an incomplete representation of the real  
119 inter-individual proximity network over the period and underestimates the number of  
120 contacts in the hospital. Interestingly, at the population level, there was no correlation  
121 between the proportion of presence time during which contact data were recorded for a given

122 individual and their average number of contacts on presence days where data were available  
123 (Supplementary Figure 2).

124

## 125 **Building synthetic contacts**

### 126 ***Algorithm outline***

127 We built an algorithm to stochastically reconstruct a realistic full temporal network of inter-  
128 individual close-proximity interactions in the hospital using parameters estimated from the  
129 observed i-Bird data. This algorithm generates a new synthetic network which notably  
130 reconstructs contacts over days when individuals were known to be present in the hospital  
131 but had no contact data recorded, which we consider to be a recording bias. The synthetic  
132 network hence includes both the observed and unobserved parts of the empiric network. This  
133 approach first involves the calculation of contact rates and durations between individuals,  
134 stratified by the individuals' ward, category (patient, or staff profession), type of day (weekday  
135 or weekend) and hour. The algorithm then reconstructs a new network, taking as input these  
136 summary statistics as well as data on presence days for each individual in the facility. Each CPI  
137 is generated stochastically, with individuals chosen in order to promote recurring contacts,  
138 based on a probability estimated from the data.

139

### 140 ***Estimation of contact rates from the i-Bird data***

141 Contact rates per hour ( $h$  from 00h to 23h), category of individual ( $C_i$ , i.e. patient, or hospital  
142 staff profession) and ward  $W_i$  are estimated from the data as:

$$143 \quad T_{h,c_1w_1 \rightarrow c_2w_2} = \frac{\sum_{i \in C_1 W_1} \sum_{j \in C_2 W_2} \sum_{k=1}^{N_{h,i}} V_{i,j,k}}{\sum_{l=1}^{N_h} N_{C_1 W_1, l}} \quad (1)$$

144 where  $T_{h,c_1w_1 \rightarrow c_2w_2}$  is the average per-person contact rate at the hour  $h$  between individuals  
145 from category  $C_1$  belonging to the ward  $W_1$  and individuals from category  $C_2$  belonging to the  
146 ward  $W_2$ . For given hour  $h$  and individual  $i$ ,  $N_{h,i}$  is the number of instances of the hour  $h$  where  
147 at least one contact was recorded for that individual. For example, if  $i$  had a contact recorded  
148 on Tuesday 11<sup>th</sup> August at 10h, and on Tuesday 18<sup>th</sup> August at 10h,  $N_{10,i}$  would be equal to 2.  
149 For two individuals  $i$  from  $C_1 W_1$  and  $j$  from  $C_2 W_2$ ,  $V_{i,j,k}$  indicates whether contacts have been  
150 recorded on occurrence  $k$ : it equals 1 if  $i$  and  $j$  had at least one contact recorded at the instance

151  $k$  of the hour  $h$ , and 0 otherwise. Finally,  $N_h$  is the total number of instances of the hour  $h$  in  
152 the full dataset and, for a given instance  $l$  of the hour  $h$ ,  $N_{C_1W_1,l}$  is the number of individuals  
153 from  $C_1W_1$  that had any contact recorded during that hour.

154

155 This estimation was conducted separately for contacts during weekdays and contacts during  
156 weekends.

157

### 158 **Estimation of recurring contacts**

159 For each individual  $i$ , we calculate the probability of recurring contact for each day  $d$  between  
160 the first ( $d_0$ ) and last ( $d_{max}$ ) days where a contact was recorded for  $i$ , according to

$$161 \quad p_{i,d} = \frac{|U_{i,d} \cap U_{i,[d_0,d]}|}{|U_{i,d}|} \quad (2)$$

162 Where  $U_{i,d}$  is the set of unique individuals with whom  $i$  had a contact on day  $d$ , and  $U_{i,[d_0,d]}$  is  
163 the set of unique individuals with whom  $i$  had at least one contact on any day between the  
164 first day  $d_0$  and the current day  $d$ . For example, if  $i$  had a contact with four unique individuals  
165 on day  $d$ , and previously had a contact with two of those on any day between  $d_0$  and  $d$ , the  
166 probability of recurring contact for day  $p_{i,d}$  would be  $2/4 = 0.5$ .

167

168 We then calculated the mean probability of recurring contacts for individual  $i$  across all days  
169 as

$$170 \quad p_i = \frac{\sum_{d=d_0}^{d_{max}} p_{i,d}}{1+(d_{max}-d_0)} \quad (3)$$

171 Finally, we calculated the mean probability of recurring contacts by individual category  $c$   
172 (patient or staff) as

$$173 \quad p_c = \frac{\sum_{i \in C} p_i}{|C|} \quad (4)$$

174 Where  $C$  represents the set of individuals belonging to category  $c$ .

175

### 176 **Generation of synthetic CPIs: number and individuals in contacts**

177 For each hour of our period of interest, we estimate the number of contacts between  
178 individuals present in the hospital during that hour, determined using the admission data and  
179 staff schedule collected during the i-Bird study. We generate the number of individuals  $n$  from  
180 category  $C_2S_2$  in contact with an individual  $i$  from category  $C_1S_1$  during an hour  $h$  by sampling

181 from a Poisson distribution with the mean being the contact rate as described above. Before  
182 selecting these  $n$  individuals, since contacts are generated dynamically, we check if  $i$  is already  
183 included in the contacts of individuals from  $C_2S_2$  during  $h$ . If  $n'$  individuals from  $C_2S_2$  have  
184 already had a contact with  $i$  during  $h$ , we only select  $n = n - n'$  new individuals from those  
185 available, in order to avoid double counting.

186

187 These  $n$  individuals are selected by favouring contacts between individuals who have already  
188 met at any other time previous to  $h$ . Let  $p_c$  be the probability of a recurring contact for  
189 category  $c$  (patient or staff) of the individual  $i$ . To determine the identity of the  $n$  individuals  
190 in contact with  $i$ , we draw a random number  $r \sim Uniform(0,1)$

191 • If  $r \leq p_c$ , a recurring contact is generated:  $j$  is chosen among  $S$ , the subset of  $C_2S_2$   
192 individuals who previously met  $i$ , according to probability  $p_{i \rightarrow j}$ :

193 
$$p_{i \rightarrow j} = \frac{N_{i \rightarrow j}}{\sum_{k \in S} N_{i \rightarrow k}} \quad (5)$$

194 Where  $N_{i \rightarrow j}$  is the number of previous contacts between  $i$  and  $j$  before hour  $h$ , and  
195  $N_{i \rightarrow k}$  is the number of previous contacts between  $i$  and each individual  $k$  belonging to  
196  $S$ .

197 • Otherwise, the contact is not recurring: the individual  $j$  in contact is randomly and  
198 uniformly chosen among  $S'$ , the subset of  $C_2S_2$  individuals who have not yet met  $i$ .

199

## 200 **Generation of contact durations**

201 We generated the duration of contact between two individuals ( $i$  from  $C_1S_1$  and  $j$  from  $C_2S_2$ )  
202 by sampling from a log-normal distribution. This distribution was calibrated using the mean  
203 and variance of the duration of contact at the hour  $h$  between two individuals from  $C_1S_1$  and  
204  $C_2S_2$ , estimated from the data.

205

## 206 **Observation bias process**

207 As mentioned earlier, in any real-life data, it can be assumed that there are periods of non-  
208 recording of CPIs (bias in collection). For each individual in the observed network, we  
209 identified the hours when they had no contact recorded. We then removed those individuals  
210 on those hours before proceeding with the algorithm described above. Hence, the  
211 reconstructed biased network and the observed network suffer from the same bias and are  
212 comparable.



213

## 214 **Simulations and analysis**

215 From the observed network, we generated 100 full reconstructed networks, and 100  
216 reconstructed networks with observation bias. For comparison, we also generated 100  
217 pseudo-random contact networks with observation bias, and 100 without. The latter networks  
218 simulate contacts without taking into account the ward, staff category, and probability of  
219 recurring contact in the calculation of contact rates and durations. At each contact, the  
220 individual encountered is therefore chosen randomly from all those present in the hospital at  
221 that time, regardless of whether or not the individual was previously encountered.

222

223 We implemented the algorithm in C++ with the repast HPC 2.3.0 library. All simulations were  
224 performed on the Maestro cluster hosted by the Institut Pasteur. The networks were analysed  
225 in R [27], using the igraph package [28]. The relevant contact networks and analysis code are  
226 available in the following GitHub repository: [https://github.com/gleclerc/network\\_algorithm](https://github.com/gleclerc/network_algorithm).

227

## 228 **Validation of the full reconstructed networks**

229 For validation, we also applied the algorithm to each of the 100 reconstructed networks with  
230 bias, to generate 100 new full reconstructed networks and confirm these “re-simulated  
231 networks” were similar to the full reconstructed networks generated from the observed data.

232

## 233 Results

### 234 Description of the observed network, application to the i-Bird dataset

235 In this section, the contact data are aggregated at the daily level, so that if two individuals  
236 have two separate contacts with each other at different times of the day, this is only counted  
237 once. The contact network is considered undirected, since contacts are assumed to be  
238 reciprocal. Daily-averaged contact matrices built from these data are described in a previous  
239 work [1].

240

241 We first summarise the observed temporal network, comparing the total network and  
242 subgraphs with only patient-patient, staff-staff, or patient-staff contacts (Figure 1a-d). Table  
243 1 provides the degree, global efficiency, density, transitivity, assortativity and temporal  
244 correlation of these four networks. The mean degree of the total network per day is 12.99  
245 (standard deviation: 3.53), which corresponds to the average number of unique contacts per  
246 individual per day. In the subgraphs, the degree is highest in the patient-staff subgraph (8.09;  
247 sd: 1.89), although we still note a relatively important number of patient-patient contacts,  
248 with a degree of 5.25 (sd: 1.87) in the corresponding subgraph. The distribution of individual  
249 degrees for all individuals and all days across the total network is heterogeneous, with a  
250 squared coefficient of variation equal to 0.44 (Figure 1e). The global efficiency of the total  
251 network is 0.40 (sd: 0.05), meaning that on average the shortest path between any two  
252 individuals has a distance of 2.5 (whereby the shortest path between two individuals in direct  
253 contact would be of distance 1). As expected, the efficiencies are lower in the subgraphs, since  
254 we remove individuals and hence increase the distance between those remaining (patient-  
255 patient: 0.25 (sd: 0.08), staff-staff: 0.32 (sd: 0.10), patient-staff: 0.31 (sd: 0.05)). Densities in  
256 the total network and subgraphs are relatively low ( $< 0.1$ ), indicating that less than 10% of all  
257 possible connections between individuals in the network are actual observed connections.

258

259 Transitivity in the total network is high (0.37; sd: 0.02), meaning that for any two individuals  $a$   
260 and  $b$  both in contact with the same third individual  $c$ , the probability that  $a$  and  $b$  are also in  
261 contact is 0.37. Transitivity is also high in the patient-patient and staff-staff subgraphs, but  
262 this metric is not relevant for the patient-staff subgraph – it is impossible for a triangle of

263 contacts to occur in this subgraph as it excludes staff-staff and patient-patient contacts by  
264 design. Assortativity by degree is negative in the total network (-0.13; sd: 0.10), indicating that  
265 highly connected individuals are more likely to be in contact with less connected individuals.  
266 It is also strongly negative in the patient-staff subgraph (-0.42; sd: 0.14), reflecting the  
267 expected disassortivity of healthcare contacts, where each staff member is in contact with  
268 multiple patients, whilst each patient is contact with relatively few staff members. In the  
269 patient-patient and staff-staff subgraphs, assortativity by degree is positive, as frequently  
270 seen in social networks.

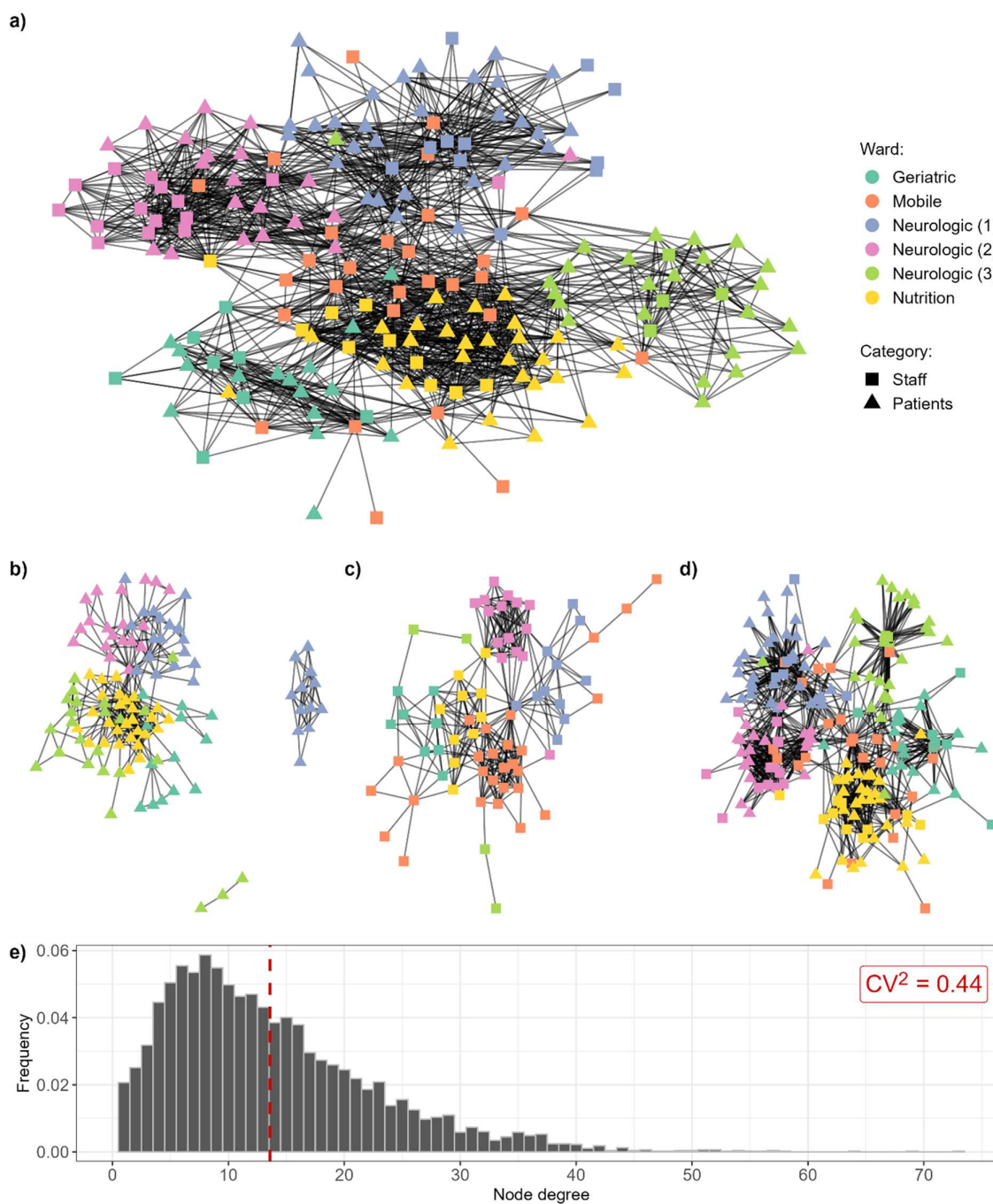
271

272 **Table 1: Summary of network characteristics for the observed total network, patient-patient**  
273 **subgraph, staff-staff subgraph, and patient-staff subgraph.** Values were estimated for each  
274 day of the 28-days period and summarised here with the mean and standard deviation (sd).  
275 Transitivity is not calculated for the patient-staff subgraph as triangles of contacts cannot  
276 occur in this network.

	<b>Total</b>	<b>Patient-patient</b>	<b>Staff-staff</b>	<b>Patient-staff</b>
<b>Degree (sd)</b>	12.99 (3.53)	5.25 (1.87)	5.82 (1.87)	8.09 (1.89)
<b>Global efficiency (sd)</b>	0.40 (0.05)	0.25 (0.08)	0.32 (0.10)	0.31 (0.05)
<b>Density (sd)</b>	0.07 (0.01)	0.05 (0.01)	0.09 (0.01)	0.05 (0.00)
<b>Transitivity (sd)</b>	0.37 (0.02)	0.41 (0.05)	0.56 (0.07)	NA
<b>Assortativity (sd)</b>				
<i>By degree</i>	-0.13 (0.10)	0.22 (0.10)	0.14 (0.14)	-0.42 (0.14)
<i>By ward</i>	0.59 (0.08)	0.77 (0.11)	0.72 (0.09)	0.47 (0.09)
<b>Temporal correlation</b>	0.47 (0.11)	0.65 (0.07)	0.35 (0.16)	0.41 (0.12)

277

278



279

280 **Figure 1: Representation of the observed (a) total network, and (b) patient-patient, (c) staff-**  
281 **staff and (d) patient-staff subgraphs on a single day.** The date of 28<sup>th</sup> of July 2009 was chosen  
282 arbitrarily. The layout was calculated using the Kamada-Kawai algorithm, with no weights  
283 applied to edges. **e) Distribution of individual degrees for the total network per person per**  
284 **day, across the entire study period.** The dashed red line indicates the mean degree (13.59).  
285 CV: coefficient of variation (standard deviation/mean).

286

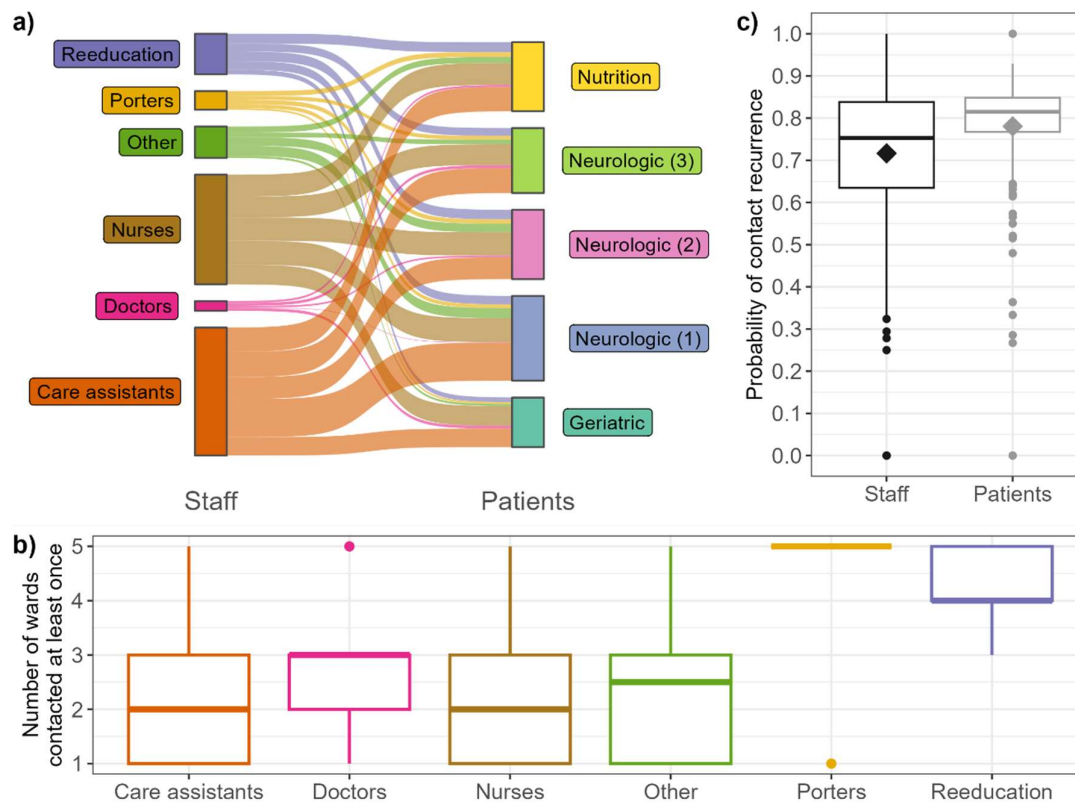
287 Visually, we observe that contacts are naturally clustered by ward (Figure 1a-d). This is  
288 reflected in the assortativity by ward, which is systematically high ( $> 0.45$ ) and indicates that  
289 individuals in a ward are always more likely to have contacts with other individuals in the same  
290 ward than with individuals in other wards (Table 1). We also observe that contacts exist  
291 between all grouped staff professions and patients in different wards, although the  
292 distribution is heterogeneous (Figure 2a-b). For example, the median number of wards with  
293 which a care assistant (orange) is in contact with is two, while almost all porters (yellow) have  
294 contacts with patients from all five wards (Figure 2b).

295

296 Overall, contacts are relatively well maintained over time, as shown by the temporal  
297 correlation coefficient of 0.47 (sd: 0.11, Table 1). This corresponds to the average probability  
298 that, between two subsequent days, an individual maintains the same number of unique  
299 contacts, with the same individuals. This metric is highest in the patient-patient subgraph  
300 (0.65, sd: 0.07) and lowest in the patient-staff subgraph (0.35, sd: 0.16), indicating that  
301 patients tend to have the same contacts with each other every day, whilst contacts amongst  
302 healthcare workers often vary between subsequent days. This consistency over time is  
303 reflected in the probability of recurring contacts (mean probability: 0.78 for patients, 0.71 for  
304 staff), although we note more variability amongst staff than patients (Figure 2c).

305

306 All the characteristics described above differ between weekdays and weekends in the network  
307 and indicate that there are fewer contacts during weekends (Supplementary Table 1). This  
308 difference is reflected in the temporal correlation, which tends to be high when comparing  
309 Sunday to Saturday, but low when comparing Saturday to Friday and Monday to Sunday,  
310 indicating that the structure of the network changes the most between these timepoints  
311 (Supplementary Figure 3).



312

313 **Figure 2: Description of contact heterogeneity and recurrence across the facility. a)**

314 **Repartition of contacts between grouped staff professions and patient wards. A link**

315 **between one staff category and one patient ward indicates that, at any point during the**

316 **investigation period, a staff member from that category had a contact with a patient from that**

317 **ward. For ease of visualisation, occupational therapists, physiotherapists, and other re-**

318 **education staff are grouped into “Reeducation”; administrative, animation/hairdresser,**

319 **logistic, and hospital service agents are grouped into “Other”; and nurses, head nurses, and**

320 **students/interns are grouped into “Nurses”. Porters, doctors and care assistants are not**

321 **grouped. b) Distribution of number of wards with which each staff member has had at least**

322 **one contact with during the study period. c) Distribution of probabilities of recurring**

323 **contacts. Each observation is calculated over the entire period, and corresponds to the**

324 **average probability for one staff or one patient to form a new contact with a previously-met**

325 **individual (staff or patient) rather than a new individual. Diamonds indicate the mean values.**

326

327

328

329

## 330 **Comparison of synthetic networks**

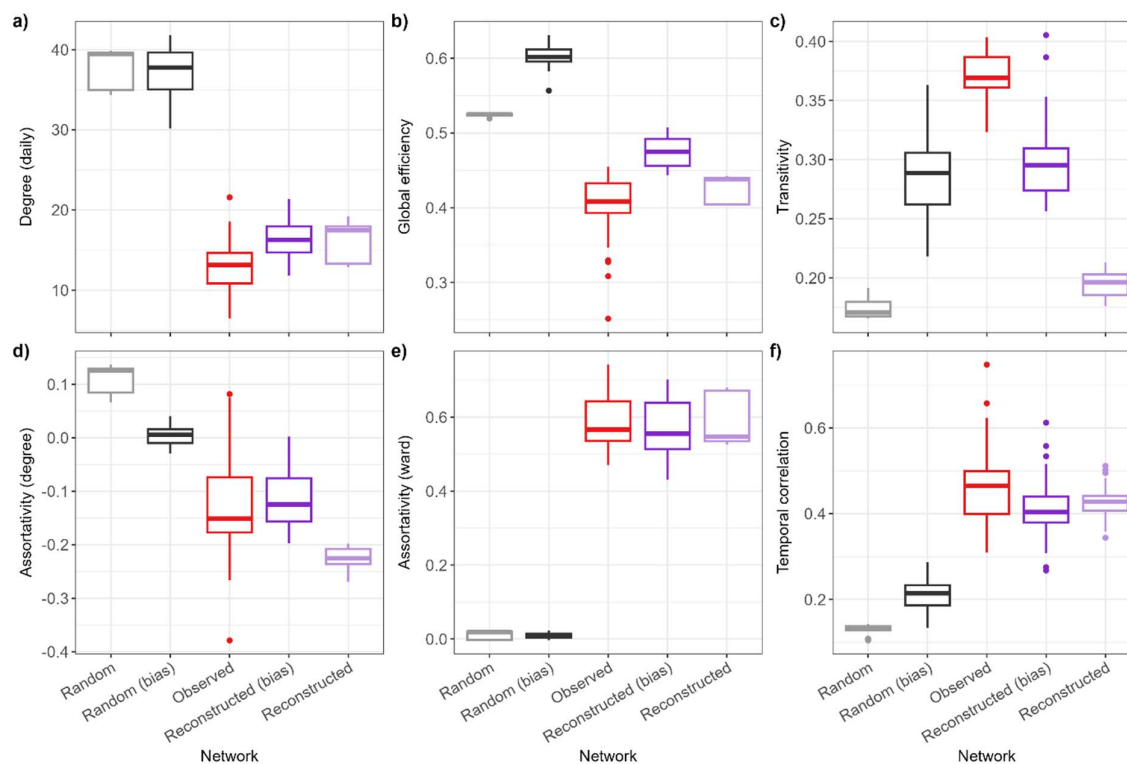
331 We applied our algorithm to the network described above to stochastically construct four  
332 types of synthetic networks: 100 full reconstructed networks, 100 reconstructed networks  
333 with observation bias, 100 full pseudo-random networks, and 100 pseudo-random networks  
334 incorporating observation bias. We expected that the characteristics of the reconstructed  
335 networks with observation bias would be broadly similar to those of the observed i-Bird  
336 network. Summary network characteristics are reported in Figure 3 and Supplementary Figure  
337 4.

338

339 The daily degrees in the reconstructed networks were slightly higher than the observed  
340 network (Figure 3a). Global efficiency was similar between the observed and reconstructed  
341 networks, but slightly higher in the reconstructed network with bias (Figure 3b). This is  
342 because the algorithm with bias removed individuals from the network at times when they  
343 did not wear their sensor during the study, hence reducing the average distance between  
344 remaining individuals. For the same reason, the density of the reconstructed network with  
345 bias was slightly higher than the observed (Supplementary Figure 4). Transitivity was slightly  
346 higher for the reconstructed network with observation bias than without, but lower than the  
347 observed network in any case (Figure 3c), as expected since the algorithm did not take into  
348 account any element of transitivity when constructing synthetic networks. Finally,  
349 assortativity by degree and by ward, as well as temporal correlation, were all well preserved  
350 in the reconstructed networks (Figure 3d-f). As a comparison, the random networks with or  
351 without bias either substantially over- or under-estimated the values for all metrics compared  
352 to the observed network (Figure 3a-f), although we note that transitivity was similar to the  
353 other synthetic networks (Figure 3c).

354





355

356 **Figure 3: Comparison of network characteristics.** The reconstructed networks with  
357 observation bias exclude individuals from the network at times when they were known to not  
358 wear their sensors. The random networks did not take into account the ward-level structure  
359 of the contacts or the probability of recurring contacts. Boxplots for the observed network  
360 show the distribution of values calculated for each day. Boxplots for all reconstructed and  
361 random networks show the distribution of the median values calculated for each day across  
362 100 networks.

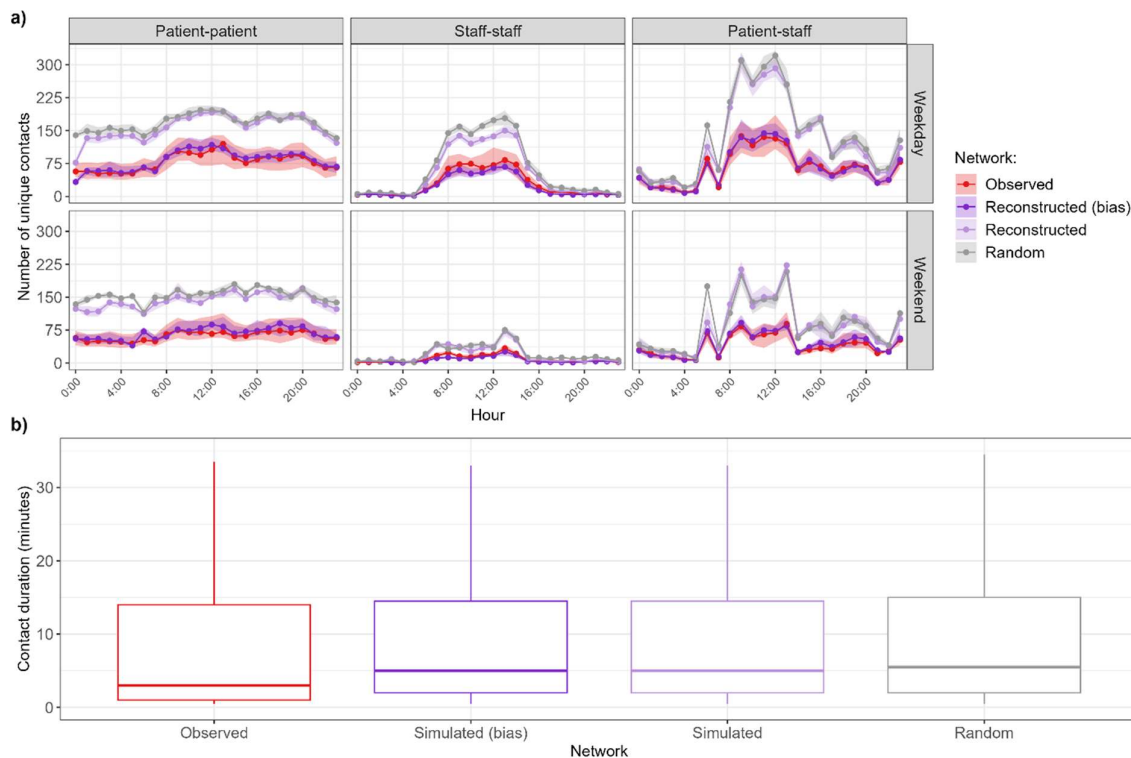
363

364 The hourly distributions of numbers of unique patient-patient, staff-patient and staff-staff  
365 contacts in the reconstructed network with bias align with those in the observed network  
366 (Figure 4a). Whilst these two networks are only partially observed since individuals in the i-  
367 Bird study did not have contacts recorded during all their presence days, those unobserved  
368 contacts are present in the reconstructed network without bias, leading to approximately  
369 twice as many contacts in that network (Figure 4a). Similarly, the random network without  
370 bias which is only informed by the hourly distribution of patient-patient, staff-staff and  
371 patient-staff contact rates is aligned with the reconstructed network (Figure 4a).

372



373 Finally, the distributions of contact durations in the synthetic networks were similar to the  
374 distribution in the observed network, although there are slightly less shorter contacts (Figure  
375 4b). This is because all networks sample their contact durations from a lognormal distribution  
376 parameterised by the mean and variance estimated from the data, which puts less emphasis  
377 on very short contacts of less than one minute (Supplementary Figure 5).  
378



379 **Figure 4: Comparison of network contact number and duration. a) Distribution of number**  
380 **of unique contacts per hour, separated by type of day (weekday or weekend).** Points  
381 **correspond to the median, and the shaded areas correspond to the interquartile range. b)**  
382 **Distribution of contact durations.** For ease of visualisation, outliers are not shown on the  
383 **graph.**  
384

385  
386 Finally, in supplementary analyses, we assessed the robustness of our algorithm by  
387 quantifying the variability of network characteristics across 100 reconstructed networks  
388 without bias (Supplementary Figure 6). The variability across reconstructed networks was not  
389 statistically significant for any metric (Kruskal-Wallis test,  $p$  value  $> 0.05$ ) except for  
390 assortativity by degree ( $p < 0.001$ ). We also aimed to validate our approach by generating “re-  
391 simulated” networks informed by summary statistics derived from the reconstructed

392 networks with bias. These re-simulated networks are similar to the full reconstructed  
393 networks, indicating that our algorithm consistently recreates realistic networks and  
394 reconstructs unobserved contacts (Supplementary Figure 7). However, the number of patient-  
395 patient contacts in the re-simulated networks is slightly higher than in the reconstructed  
396 networks (Supplementary Figure 7).

## 397 **Discussion**

### 398 **Summary of findings**

399 In this article, we present an approach to construct stochastic synthetic temporal contact  
400 networks from partially observed contact data. To test our algorithm, we used 28 days of CPI  
401 data from the i-Bird study, which recorded contacts of staff and patients in a long-term care  
402 facility. The observed i-Bird network was heterogeneous, with notably a strong assortativity  
403 by ward, varying contact rates between different staff categories and patients, and different  
404 contact structures on weekends compared to weekdays. Importantly, we observed temporal  
405 correlation between subsequent days in the network, and we estimated that individuals were  
406 generally more likely to have contacts with other individuals they previously met rather than  
407 new individuals. We therefore informed our reconstruction algorithm with both contact rates  
408 by hour, type of day (weekday or weekend) and staff category, and probabilities of recurring  
409 contacts estimated for patients and staff using the i-Bird data. The resulting reconstructed  
410 networks reproduced well the characteristics of the observed network, as well as the specific  
411 distribution of unique contacts per hour.

412

### 413 **Similarities between the observed and reconstructed networks**

414 The i-Bird contact network was only partially observed, since contact data was missing on  
415 average for 37% of each patient's presence days (42% for staff). This could have occurred for  
416 a number of reasons which we cannot distinguish, including depleted batteries, sensor  
417 malfunction, or imperfect sensor-wearing compliance. However, since the average contact  
418 rates of individuals did not correlate with the proportion of their presence time during which  
419 no contact data were recorded (Supplementary Figure 2), it can be assumed that contact  
420 patterns on unobserved days were similar to those on observed days. With that assumption,  
421 we were able to reconstruct contacts at those times when individuals were present but had  
422 no reported contact data. The resulting full reconstructed network is a valuable  
423 representation of individual interactions, as it represents the "true" contact network,  
424 compared to the i-Bird empirical network which was only partially observed. Although we  
425 were inherently limited in our ability to validate this network since the real, fully observed  
426 network was not available, we compared it to a re-simulated network which used the

427 reconstructed network with observation bias as input. The reconstructed and re-simulated  
428 networks without bias were almost identical with regards to all the network metrics we  
429 considered (Supplementary Figure 7), demonstrating the consistency of our algorithm to  
430 reconstruct contacts.

431

432 To generate a reconstructed network directly comparable to the observed i-Bird data, we  
433 included an observation process to only simulate contacts for individuals at hours when they  
434 had contact data reported. This reconstructed network with bias and the observed network  
435 had similar positive assortativity by ward, as expected since the input data captured the  
436 contact structure by ward. The negative assortativity by degree was also similar, however we  
437 noted variability between different networks generated independently by the algorithm  
438 (Supplementary Figure 6). Since the algorithm did not directly account for assortativity when  
439 simulating networks, this similarity stems from our use of a recurring contact probability  
440 coupled with the contact rates estimated by staff categories, resulting in a non-random  
441 contact structure with regards to this metric. The hourly contact distribution of patient-  
442 patient, staff-staff, and patient-staff contacts was also successfully reproduced by our  
443 algorithm.

444

445 A key metric of interest here is temporal correlation, which indicates how conserved the  
446 network structure is over time. This type of metric is useful to determine the efficiency of  
447 disease spread across temporal networks over time [29,30]. Since our algorithm took into  
448 consideration the probability of recurring contacts between individuals, our reconstructed  
449 networks displayed similar temporal correlation as observed, whilst random networks  
450 substantially underestimated this. This aspect is therefore an important strength of our  
451 approach, compared to only using estimated average contact rates to construct synthetic  
452 contacts.

453

#### 454 **Limitations of the algorithm**

455 Density and global efficiency in the reconstructed network with bias were slightly higher than  
456 in the observed network. This is a likely consequence of our observation process which forcibly  
457 removed individuals from the network at times when they had no contacts recorded, hence

458 reducing the number of nodes available in the network. Simultaneously, there was still a need  
459 at those times to generate some novel contacts between individuals who never previously  
460 met, since the probability of recurring contacts was less than 1. Combined, these elements  
461 increased the overall connectivity amongst all individuals in the reconstructed network with  
462 bias. Although this could facilitate disease transmission across these reconstructed networks  
463 if they are used for such purpose [31], the high assortativity by ward may counter this effect  
464 by slowing down transmission across the entire healthcare facility.

465

466 Our algorithm did not specifically account for transitivity when recreating contacts. This is  
467 likely why the resulting transitivity was similar to that of the random network and  
468 underestimated the observed value (Figure 3). Similarly to density and global efficiency  
469 mentioned above, any transitivity in the reconstructed network was likely an indirect  
470 consequence of assortativity by ward, restricting the pool of available individuals to generate  
471 contacts and leading to interconnectivity between individuals present in the same ward. In  
472 any case, this may not substantially affect disease transmission simulated across these  
473 networks, since previous work has shown that transitivity is a poor predictor of the total  
474 number of individuals who would be infected across the network [31]. Whilst we could extend  
475 our algorithm to consider transitivity when choosing the individuals to put in contact, we  
476 decided not to do this here to maximise the generalisability of our approach by not requiring  
477 such highly detailed contact data.

478

479 Information about patient temporary releases from the hospital (e.g., for weekends with their  
480 families) was not available in the i-Bird data, hence such events were not accounted for here,  
481 although they may occur frequently in a LTCF. Consequently, the duration of hospital stays  
482 may have been overestimated, leading to an overestimation of some contacts among  
483 patients. Although this is negligible when comparing the observed and reconstructed network  
484 with bias, this is likely why the re-simulated networks slightly overestimated the number of  
485 patient-patient contacts compared to the full reconstructed network (Supplementary Figure  
486 7). We expect that this overestimation would be absent in settings with more complete  
487 information on individual presence, such as schools, workplaces, or acute care facilities.

488

489 **Future work**

490 In this study, we restricted our detailed analysis of the accuracy of our algorithm to a long-  
491 term care facility setting. However, contact structures are known to vary depending on the  
492 setting investigated, even between different healthcare settings such as long-term or acute.  
493 In addition, even though we had substantial data covering four weeks, it's unclear how  
494 representative this contact structure is for other time periods. Notably, our study period falls  
495 in the middle of the summer, which is a holiday period in France and may have affected  
496 contact patterns. A first important next step would therefore be to repeat this analysis using  
497 data collected in a different setting such as acute care, over a different time period. Although  
498 we do not expect that our algorithm will perform differently, the strengths and limitations we  
499 have highlighted above may be more or less relevant in these different settings. For example,  
500 in a setting with low transitivity, the fact that our algorithm underestimates this metric would  
501 be less problematic.

502

503 Here we directly re-used patient admission and discharge data as well as staff schedules to  
504 identify which individuals were present in the facility at each hour, and hence whom the  
505 algorithm had to build contacts for. While this choice was coherent since our aim was to  
506 compare the observed and reconstructed networks, a second possible extension of our work  
507 would be to simulate the presence of individuals over time. This could be implemented by  
508 extracting admission and discharge rates for each category of staff and patients and using  
509 these values to recreate new presence times for individuals by sampling from relevant  
510 probability distributions. This would allow us to further account for possible variability in the  
511 structure of the population in the facility, and hence add further stochasticity in our algorithm.

512

513 Since contact data may only be available for short periods of time, a third question of interest  
514 would be to understand the volume of data required to generate realistic temporal contact  
515 networks using our algorithm. In our main analysis, we used the entire four weeks available  
516 to both derive contact parameters and compare the reconstructed and observed networks.  
517 For sensitivity, we also considered smaller timer periods to calculate the summary contact  
518 parameters required by the algorithm (Supplementary Figure 8). As expected, this led to  
519 variability amongst the reconstructed networks depending on the length of the period used.  
520 In any case, the main risk of using only a short period of time is to miss out some contacts  
521 between categories. For example, during a single week, by chance there may not be any

522 observed contact between patients from one ward  $w_1$  and a nurse from another ward  $w_2$ ,  
523 while in reality over a longer period of time we may observe a few of such contacts. In that  
524 case, the algorithm will systematically assume that such contacts never occur during the entire  
525 period over which the reconstructed networks are generated and will therefore construct an  
526 incomplete network. Therefore, it is essential for users to be confident that the data they use  
527 include contact rates for all relevant categories in their setting and for typical representative  
528 days.

529

530 As discussed above, taking into account the probability for contacts to be recurring instead of  
531 assuming a uniform distribution is a key element of our approach. Here, we estimate the  
532 average probabilities of recurring contacts as 0.71 for staff and 0.78 for patients, but we note  
533 some individual variation in this value (Figure 2). In addition, our estimation here is made using  
534 the entire observed contact networks, but this may be difficult in instances where only limited  
535 data are available. For sensitivity, we investigated the impact of manually setting the  
536 probabilities to 0.1, 0.5 and 0.9 for both staff and patients (Supplementary Figure 9). This led  
537 to important variations in assortativity by degree and temporal correlation compared to using  
538 the estimated probability. A greater understanding of this recurring contact probability in  
539 various settings would be helpful to improve the generalisability of our algorithm and will also  
540 be useful more broadly to better understand contact heterogeneity. In healthcare settings,  
541 this probability could likely be estimated without requiring complete contact data, using  
542 information on staff schedules and patient ward allocation instead.

543

544 Deep learning algorithms such as graph convolutional networks (GCN) have become  
545 increasingly popular to study contact networks, particularly in the context of infectious  
546 disease transmission [32–36]. It would be interesting to compare the performance of these  
547 approaches with our algorithm to estimate network characteristics and reconstruct  
548 unobserved contacts. However, traditional GCN approaches do not account for temporal  
549 dependencies between contacts such as the ones we observed in the i-Bird network where  
550 the probability of recurring contacts plays a key role [37,38]. On the other hand, temporal  
551 graph networks can capture this temporal dependency [39,40], but require substantial  
552 computational resources to be applied to a network such as i-Bird, with hundreds of  
553 interactions recorded every 30 seconds during several weeks. Finally, deep learning methods

554 require large amounts of training data. Democratising their use would therefore first require  
555 new studies to collect close-proximity interaction data in different settings and time periods,  
556 presenting further logistical challenges.

557

## 558 **Implications**

559 Our algorithm relied on computing summary statistics from an observed network, then using  
560 these statistics to stochastically reconstruct contact networks. Such statistics could be derived  
561 directly from other observed networks, as we have done here to validate our approach. In  
562 that case, our approach to generate multiple reconstructed networks could be useful to  
563 evaluate the impact of stochasticity on the contact structure in a given setting, instead of only  
564 relying on a single observed network. Our approach could also be used to infer information  
565 on potentially unobserved contacts and to predict realistic temporal dynamics over longer  
566 time periods than the data collection.

567

568 Alternatively, summary contact statistics could be more simply collected from cross-sectional  
569 surveys or even derived exclusively from individual schedules, which would not require a  
570 detailed and costly follow-up using sensors. In this scenario, the only other data required  
571 would be individual presence times, which should either be routinely available (e.g. in  
572 healthcare settings or schools) or relatively easy to collect (e.g. in workplaces). Although as  
573 mentioned in the Limitations the amount of data our algorithm requires to generate realistic  
574 networks is still unclear, our approach could ultimately be used to generate contact networks  
575 from contact matrices. This would substantially facilitate research on the impact of contact  
576 heterogeneity in various populations and settings.

577

578 To the best of our knowledge, this is the first proposed algorithm to generate temporal contact  
579 networks by taking into consideration empirically measured contact rates based on close-  
580 proximity sensors, while most available packages only construct static networks and rely on  
581 hyperparameters [41]. These temporal networks could then be used within mathematical  
582 models used to evaluate the potential impact of interventions against disease transmission  
583 networks [9–12]. In particular, this could improve the wider applicability of individual-based



584 model which can take into account this detailed contact heterogeneity to test the effect of  
585 highly targeted interventions.

## 586 References

- 587 1. Keeling MJ, Rohani P. Modeling Infectious Diseases in Humans and Animals. Modeling  
588 Infectious Diseases in Humans and Animals. Princeton University Press; 2008.  
589 doi:10.1515/9781400841035
- 590 2. Anderson RM, May RM. Infectious diseases of humans: dynamics and control.  
591 Reprinted. Oxford: Oxford Univ. Press; 2010.
- 592 3. Diekmann O, Heesterbeek JAP. Mathematical Epidemiology of Infectious Diseases:  
593 Model Building, Analysis and Interpretation. John Wiley & Sons; 2000.
- 594 4. Großmann G, Backenköhler M, Wolf V. Heterogeneity matters: Contact structure and  
595 individual variation shape epidemic dynamics. PLoS One. 2021;16: e0250050.  
596 doi:10.1371/journal.pone.0250050
- 597 5. Kiss IZ, Miller JC, Simon PL, others. Mathematics of epidemics on networks. Cham:  
598 Springer. 2017;598: 31.
- 599 6. Danon L, Ford AP, House T, Jewell CP, Keeling MJ, Roberts GO, et al. Networks and the  
600 Epidemiology of Infectious Disease. Interdiscip Perspect Infect Dis. 2011;2011: 284909.  
601 doi:10.1155/2011/284909
- 602 7. Mossong J, Hens N, Jit M, Beutels P, Auranen K, Mikolajczyk R, et al. Social Contacts and  
603 Mixing Patterns Relevant to the Spread of Infectious Diseases. PLOS Medicine. 2008;5:  
604 e74. doi:10.1371/journal.pmed.0050074
- 605 8. Keeling MJ, Eames KTD. Networks and epidemic models. Journal of The Royal Society  
606 Interface. 2005;2: 295–307. doi:10.1098/rsif.2005.0051
- 607 9. Bansal S, Read J, Pourbohloul B, Meyers LA. The dynamic nature of contact networks in  
608 infectious disease epidemiology. Journal of Biological Dynamics. 2010;4: 478–489.  
609 doi:10.1080/17513758.2010.503376
- 610 10. Masuda N, Holme P. Predicting and controlling infectious disease epidemics using  
611 temporal networks. F1000Prime Rep. 2013;5: 6. doi:10.12703/P5-6
- 612 11. Gross T, D’Lima CJD, Blasius B. Epidemic Dynamics on an Adaptive Network. Phys Rev  
613 Lett. 2006;96: 208701. doi:10.1103/PhysRevLett.96.208701
- 614 12. Valdano E, Poletto C, Giovannini A, Palma D, Savini L, Colizza V. Predicting Epidemic Risk  
615 from Past Temporal Contact Data. PLOS Computational Biology. 2015;11: e1004152.  
616 doi:10.1371/journal.pcbi.1004152
- 617 13. Holme P, Saramäki J. Temporal networks. Physics Reports. 2012;519: 97–125.  
618 doi:10.1016/j.physrep.2012.03.001
- 619 14. Hornbeck T, Naylor D, Segre AM, Thomas G, Herman T, Polgreen PM. Using Sensor  
620 Networks to Study the Effect of Peripatetic Healthcare Workers on the Spread of

- 621 Hospital-Associated Infections. *Journal of Infectious Diseases*. 2012;206: 1549–1557.  
622 doi:10.1093/infdis/jis542
- 623 15. Obadia T, Silhol R, Opatowski L, Temime L, Legrand J, Thiébaud ACM, et al. Detailed  
624 Contact Data and the Dissemination of *Staphylococcus aureus* in Hospitals. Salathé M,  
625 editor. *PLoS Comput Biol*. 2015;11: e1004170. doi:10.1371/journal.pcbi.1004170
- 626 16. Min-Allah N, Alahmed BA, Albreek EM, Alghamdi LS, Alawad DA, Alharbi AS, et al. A  
627 survey of COVID-19 contact-tracing apps. *Computers in Biology and Medicine*.  
628 2021;137: 104787. doi:10.1016/j.compbiomed.2021.104787
- 629 17. Eames K, Bansal S, Frost S, Riley S. Six challenges in measuring contact networks for use  
630 in modelling. *Epidemics*. 2015;10: 72–77. doi:10.1016/j.epidem.2014.08.006
- 631 18. Read JM, Edmunds WJ, Riley S, Lessler J, Cummings D a. T. Close encounters of the  
632 infectious kind: methods to measure social mixing behaviour. *Epidemiology & Infection*.  
633 2012;140: 2117–2130. doi:10.1017/S0950268812000842
- 634 19. Smieszek T, Castell S, Barrat A, Cattuto C, White PJ, Krause G. Contact diaries versus  
635 wearable proximity sensors in measuring contact patterns at a conference: method  
636 comparison and participants’ attitudes. *BMC Infectious Diseases*. 2016;16: 341.  
637 doi:10.1186/s12879-016-1676-y
- 638 20. Watts DJ, Strogatz SH. Collective dynamics of ‘small-world’ networks. *Nature*. 1998;393:  
639 440–442. doi:10.1038/30918
- 640 21. Almutiry W, Deardon R. Contact network uncertainty in individual level models of  
641 infectious disease transmission. *Stat Commun Infect Dis*. 2021;13: 20190012.  
642 doi:10.1515/scid-2019-0012
- 643 22. Shirley MDF, Rushton SP. The impacts of network topology on disease spread.  
644 *Ecological Complexity*. 2005;2: 287–299. doi:10.1016/j.ecocom.2005.04.005
- 645 23. Gimma A, Munday JD, Wong KLM, Coletti P, van Zandvoort K, Prem K, et al. Changes in  
646 social contacts in England during the COVID-19 pandemic between March 2020 and  
647 March 2021 as measured by the CoMix survey: A repeated cross-sectional study. *PLoS*  
648 *Med*. 2022;19: e1003907. doi:10.1371/journal.pmed.1003907
- 649 24. Mousa A, Winskill P, Watson OJ, Ratmann O, Monod M, Ajelli M, et al. Social contact  
650 patterns and implications for infectious disease transmission – a systematic review and  
651 meta-analysis of contact surveys. Rodriguez-Barraquer I, Serwadda DM, editors. *eLife*.  
652 2021;10: e70294. doi:10.7554/eLife.70294
- 653 25. Hoang T, Coletti P, Melegaro A, Wallinga J, Grijalva CG, Edmunds JW, et al. A Systematic  
654 Review of Social Contact Surveys to Inform Transmission Models of Close-contact  
655 Infections. *Epidemiology*. 2019;30: 723–736. doi:10.1097/EDE.0000000000001047

- 656 26. Duval A, Obadia T, Martinet L, Boëlle P-Y, Fleury E, Guillemot D, et al. Measuring  
657 dynamic social contacts in a rehabilitation hospital: effect of wards, patient and staff  
658 characteristics. *Scientific Reports*. 2018;8: 1686. doi:10.1038/s41598-018-20008-w
- 659 27. R Core Team. R: A Language and Environment for Statistical Computing. Vienna,  
660 Austria: R Foundation for Statistical Computing; 2022. Available: [https://www.R-](https://www.R-project.org/)  
661 [project.org/](https://www.R-project.org/)
- 662 28. Csardi G, Nepusz T. The igraph software package for complex network research.  
663 *InterJournal*. 2006;Complex Systems: 1695.
- 664 29. Tang J, Scellato S, Musolesi M, Mascolo C, Latora V. Small-world behavior in time-  
665 varying graphs. *Phys Rev E*. 2010;81: 055101. doi:10.1103/PhysRevE.81.055101
- 666 30. Kretzschmar M, Morris M. Measures of concurrency in networks and the spread of  
667 infectious disease. *Mathematical Biosciences*. 1996;133: 165–195. doi:10.1016/0025-  
668 5564(95)00093-3
- 669 31. Pérez-Ortiz M, Manescu P, Caccioli F, Fernández-Reyes D, Nachev P, Shawe-Taylor J.  
670 Network topological determinants of pathogen spread. *Sci Rep*. 2022;12: 7692.  
671 doi:10.1038/s41598-022-11786-5
- 672 32. Fritz C, Dorigatti E, Rügamer D. Combining graph neural networks and spatio-temporal  
673 disease models to improve the prediction of weekly COVID-19 cases in Germany. *Sci*  
674 *Rep*. 2022;12: 3930. doi:10.1038/s41598-022-07757-5
- 675 33. Gao J, Sharma R, Qian C, Glass LM, Spaeder J, Romberg J, et al. STAN: spatio-temporal  
676 attention network for pandemic prediction using real-world evidence. *Journal of the*  
677 *American Medical Informatics Association*. 2021;28: 733–743.  
678 doi:10.1093/jamia/ocaa322
- 679 34. Panagopoulos G, Nikolentzos G, Vazirgiannis M. Transfer Graph Neural Networks for  
680 Pandemic Forecasting. *Proceedings of the AAAI Conference on Artificial Intelligence*.  
681 2021;35: 4838–4845. doi:10.1609/aaai.v35i6.16616
- 682 35. Kapoor A, Ben X, Liu L, Perozzi B, Barnes M, Blais M, et al. Examining COVID-19  
683 Forecasting using Spatio-Temporal Graph Neural Networks. *arXiv*; 2020.  
684 doi:10.48550/arXiv.2007.03113
- 685 36. Zhao G, Jia P, Zhou A, Zhang B. InfGCN: Identifying influential nodes in complex  
686 networks with graph convolutional networks. *Neurocomputing*. 2020;414: 18–26.  
687 doi:10.1016/j.neucom.2020.07.028
- 688 37. Li L, Zhou J, Jiang Y, Huang B. Propagation source identification of infectious diseases  
689 with graph convolutional networks. *Journal of Biomedical Informatics*. 2021;116:  
690 103720. doi:10.1016/j.jbi.2021.103720

- 691 38. Ni Q, Wu X, Chen H, Jin R, Wang H. Spatial-temporal deep learning model based rumor  
692 source identification in social networks. *J Comb Optim.* 2023;45: 86.  
693 doi:10.1007/s10878-023-01018-5
- 694 39. Holme P. Modern temporal network theory: a colloquium. *Eur Phys J B.* 2015;88: 234.  
695 doi:10.1140/epjb/e2015-60657-4
- 696 40. Tang J, Leontiadis I, Scellato S, Nicosia V, Mascolo C, Musolesi M, et al. Applications of  
697 Temporal Graph Metrics to Real-World Networks. In: Holme P, Saramäki J, editors.  
698 Temporal Networks. Berlin, Heidelberg: Springer; 2013. pp. 135–159. doi:10.1007/978-  
699 3-642-36461-7\_7
- 700 41. Prettejohn B, Berryman M, McDonnell M. Methods for Generating Complex Networks  
701 with Selected Structural Properties for Simulations: A Review and Tutorial for  
702 Neuroscientists. *Frontiers in Computational Neuroscience.* 2011;5. Available:  
703 <https://www.frontiersin.org/articles/10.3389/fncom.2011.00011>
- 704