Data on face-to-face contacts in an office building suggest a low-cost vaccination strategy based on community linkers

-

Supplementary Material

Mathieu Génois, Christian L. Vestergaard, Julie Fournet
Aix Marseille Université, Université de Toulon, CNRS, CPT, UMR 7332, 13288
Marseille, France;
André Panisson
Data Science Laboratory, ISI Foundation, Torino, Italy;
Isabelle Bonmarin
Département des maladies infectieuses, Institut de veille sanitaire, Saint-Maurice, France;
Alain Barrat
Aix Marseille Université, Université de Toulon, CNRS, CPT, UMR 7332, 13288
Marseille, France
Data Science Laboratory, ISI Foundation, Torino, Italy
(*e-mail:* genois@cpt.univ-mrs.fr)

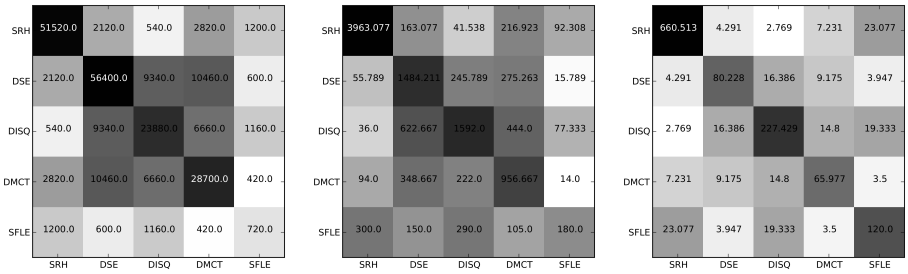
1 Contact matrices

Each element of the contact matrix shown in the main text represents the total time of contact between nodes from two departments. Since the department populations do not have the same size, it is also of interest to compute normalized contact matrices. We show in Fig. 1 two different normalizations: in the first one, we divide the elements of each row by the number of people in the corresponding department. Each element thus represents the mean time each node from the row department has spent with nodes from the column department (in this case, the matrix is no longer symmetrical). In the second normalization procedure, we divide each element of the original matrix by the number of potential links between the two departments. Each element gives then the mean contact duration between individuals of the two departments.

For completeness, we therefore also compute the similarities between daily contact matrices normalized in different ways (Fig. 2 gives the weekly and daily original – symmetric – contact matrices). The values, given in Table 1, depend on the specific normalization procedure but remain large.

If the ratio of the number of links over the total number of possible links is considered instead of the contact durations, contact matrices represent the connectivity of the contact network (Fig. 3).

In order to test the effect of spatial organization on the shape of the contact matrix, we build a null model where only the timeline of presence of each node is taken into account,



a) Non normalised **b)** Number of nodes **c)** Number of potential links

Fig. 1. Original and normalized contact matrices for the whole data set.

Table 1. **Daily contact matrices similarities.** *For each day, we compute the cosine similarity between the corresponding contact matrix and the ones of the other days, both with and without diagonals, and for three different normalizations: no normalization, normalization of the rows by the number of nodes in each department, normalization of each element by the number of possible links between the two departments. We list in this table the mean value and the standard deviation of the similarities.*

	No normalization		Number of nodes		Number of links	
Day	full	w/o diagonal	full	w/o diagonal	full	w/o diagonal
06/24	0.753 ± 0.103	0.563 ± 0.222	0.649 ± 0.098	0.437 ± 0.258	0.691 ± 0.277	0.494 ± 0.197
06/25	0.843 ± 0.069	0.481 ± 0.087	0.748 ± 0.084	0.297 ± 0.225	0.707 ± 0.276	0.301 ± 0.199
06/26	0.837 ± 0.046	0.400 ± 0.246	0.637 ± 0.072	0.207 ± 0.157	0.421 ± 0.220	0.329 ± 0.151
06/27	0.870 ± 0.052	0.534 ± 0.108	0.819 ± 0.051	0.380 ± 0.091	0.762 ± 0.297	0.370 ± 0.184
06/28	0.871 ± 0.075	0.426 ± 0.134	0.804 ± 0.110	0.341 ± 0.220	0.798 ± 0.171	0.324 ± 0.086
07/01	0.821 ± 0.058	0.592 ± 0.152	0.788 ± 0.083	0.437 ± 0.234	0.763 ± 0.302	0.431 ± 0.199
07/02	0.850 ± 0.087	0.579 ± 0.180	0.827 ± 0.110	0.463 ± 0.231	0.772 ± 0.305	0.510 ± 0.185
07/03	0.858 ± 0.072	0.488 ± 0.262	0.770 ± 0.096	0.376 ± 0.225	0.273 ± 0.275	0.418 ± 0.224
07/04	0.767 ± 0.058	0.317 ± 0.131	0.704 ± 0.089	0.219 ± 0.111	0.679 ± 0.277	0.255 ± 0.138
07/05	0.795 ± 0.123	0.398 ± 0.199	0.762 ± 0.148	0.271 ± 0.177	0.725 ± 0.294	0.387 ± 0.176

and interactions are assumed to take place at random between individuals who are in the same location. The timelines of presence are built from the empirical contact data: a node is present at a given location at a time t if it takes part in a contact recorded here at this time. A node that is present at t is moreover assumed to be present during the interval $[t-\Delta, t+\Delta]$. At each time t, all nodes that are present in a given location have a constant probability to be in contact. The probability is chosen such that the total cumulative duration of contacts is equal to its empirical value. The contact matrix obtained, shown in Fig. 4, is significantly different from the empirical one. This indicates that the empirical contact matrix structure is not explained by random encounters of individuals with different presence timelines.

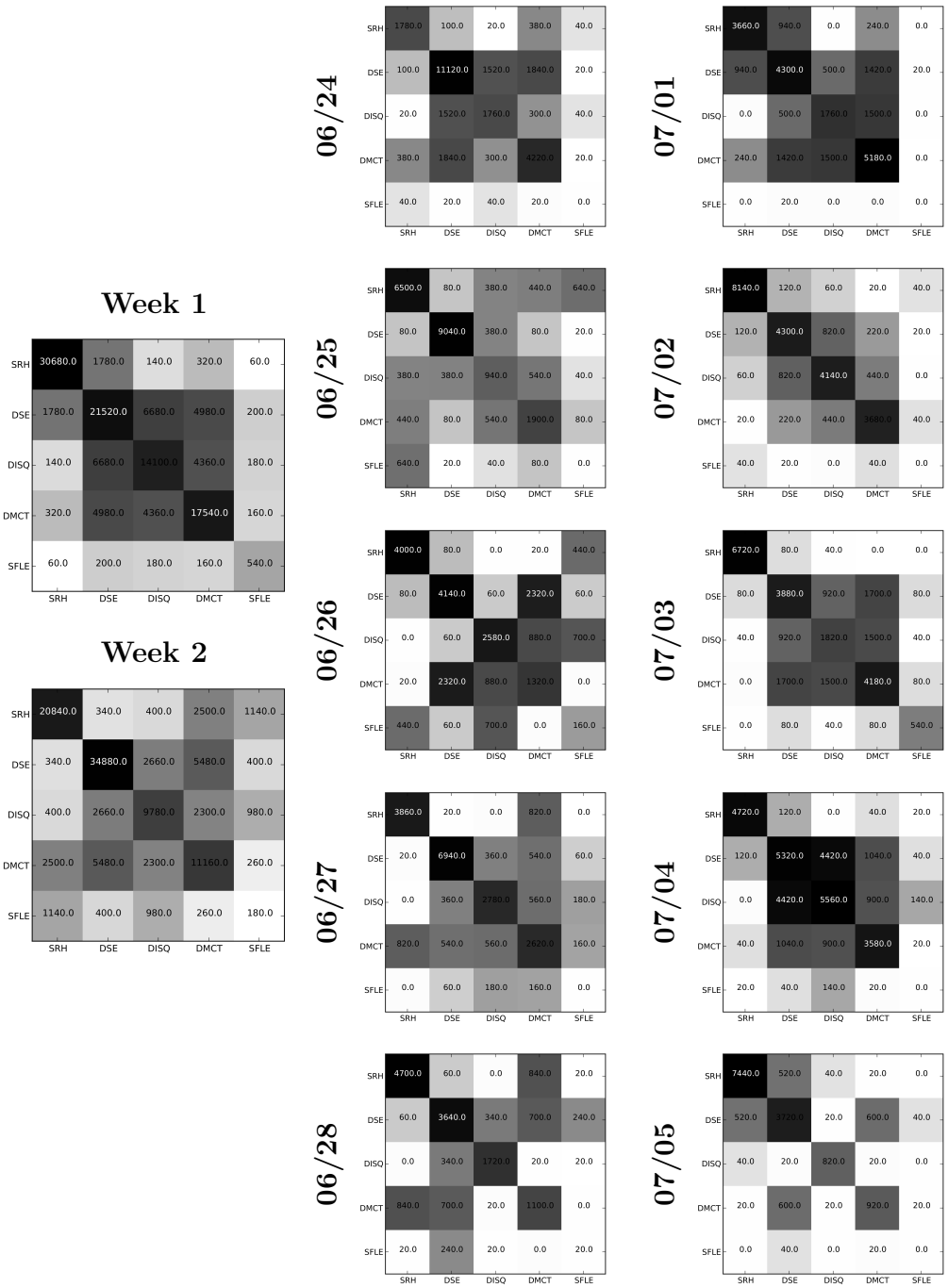


Fig. 2. Weekly and daily contact matrices.

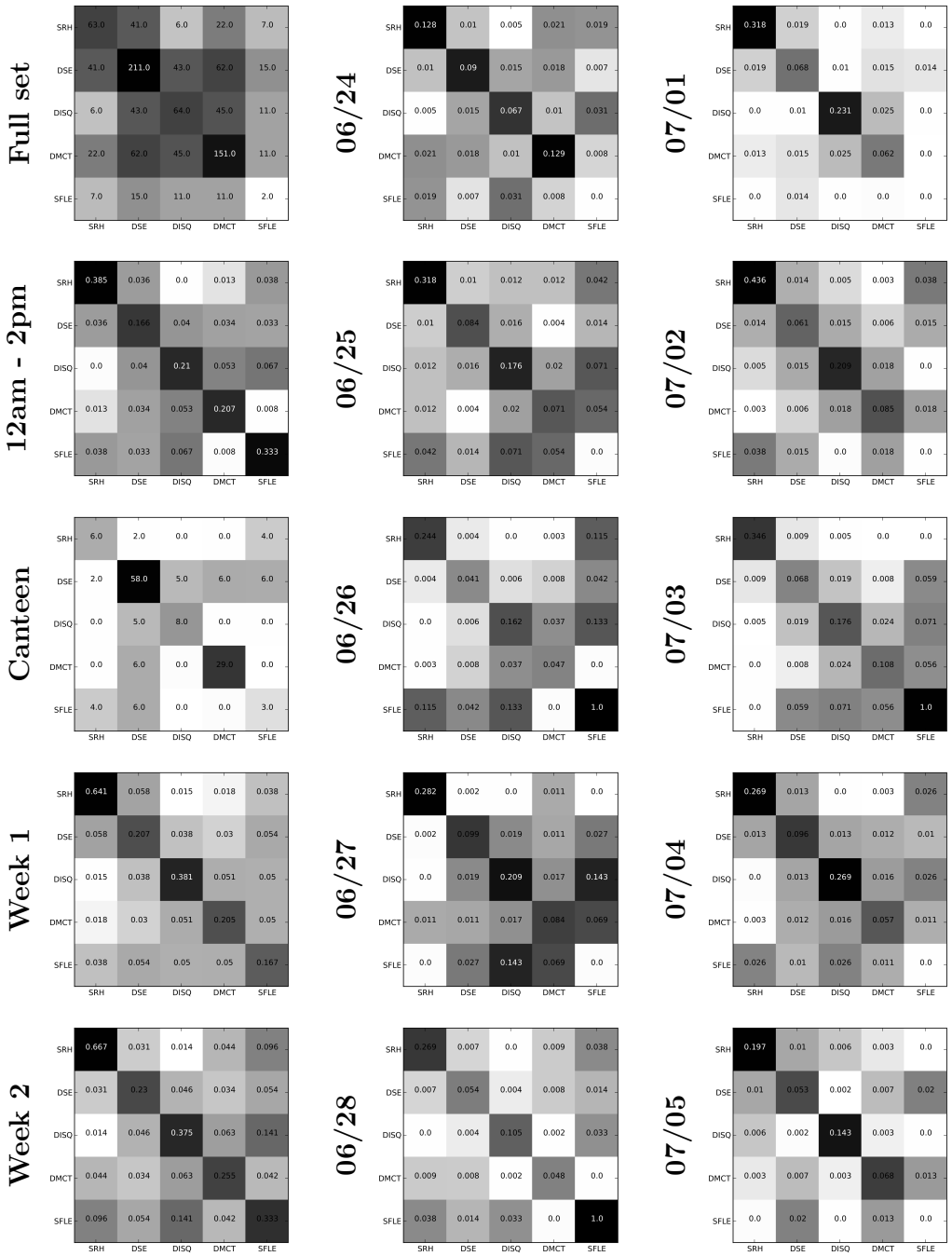


Fig. 3. Link density contact matrices.

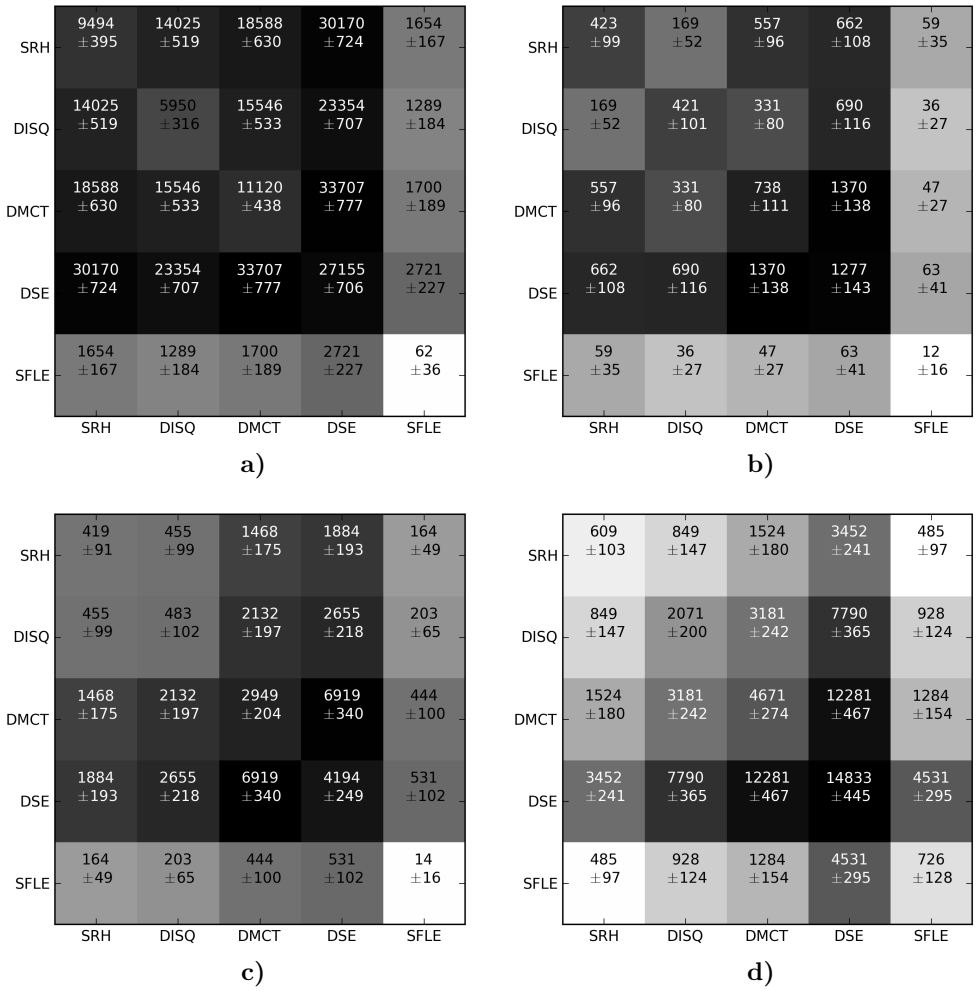


Fig. 4. Contact matrices: null model with constant contact probability and empirical presence timelines, for $\Delta = 30$ min, averaged over 100 different realizations. Each matrix element (at row X and column Y) gives the total time of contact (mean $\pm$ s.e.m.) between individuals from departments X and Y during the two weeks of the study, in different locations, according to the null model. **a)** Entire building. **b)** Conference room. **c)** Cafeteria, restricted to the interval between 12am and 2pm for each day. **d)** Canteen. This place is in a different building and was not taken into account in a).

2 Effect of data representation on epidemic spreading.

The collected contact data consist in a temporal network at a very high temporal resolution. These data can be aggregated and represented in different ways, both along the temporal and the organizational dimensions, in order e.g. to build models for the spread of epidemics in the population. As discussed e.g. in (Smieszek *et al.*, 2009; Read *et al.*, 2008; Eames *et al.*, 2009; Stehlé *et al.*, 2011; Machens *et al.*, 2013; Blower & Go, 2011), the level of detail of the data representation that is taken into account in the model can influence the outcome of the simulations. We consider this issue with the data at hand, by using five different representations for the contact dynamics, from highly detailed to simplistic, and by using each representation as the support of an SIR model for epidemic spread (Machens *et al.*, 2013):

- **Full data.** We use the temporal network built from the empirical data at the highest temporal resolution (20 s).
- **Heterogeneous static network.** We use the contact network aggregated over the whole data collection period: in this network, nodes representing individuals who have been in contact at least once are connected by a link whose weight is given by the total contact time of these individuals, normalized by the total duration of the data set.
- **Global contact matrix.** We consider that all nodes are connected to each other, and that the weight of a link connecting two nodes depends only on their respective departments: it is given by the average contact time of all pairs of individuals belonging to these departments. In other words, the total contact time between each pair of departments is equally redistributed among all pairs of individuals of these departments.
- **Daily contact matrices.** We consider a contact matrix representation, using for each day the corresponding daily contact matrix to compute the weights of the links between individuals.
- **Homogeneous mixing.** We consider a fully connected contact network with homogeneous weights, computed as the average contact time between any two individuals (independently of their department).

In each representation, we moreover take into account inactivity periods (nights and weekends) by assuming that all nodes are isolated during these periods.

The results of the numerical simulations of an SIR model are shown in Fig. 5 for two values of β/μ. For $\beta/\mu = 100$, no matter which representation is used, most of the epidemics do not reach a large fraction of the population. The tails of the distribution of epidemic sizes however become broader when using contact matrices or a homogeneous mixing assumption, as the sparsity of the contacts is then not correctly considered (Machens *et al.*, 2013).

This effect is seen most clearly for $\beta/\mu = 1000$. With complete contact information, i.e., if the spread is simulated on the time-resolved contact network, the distribution depends on the value of β, as discussed in the main text. For small β (slow epidemics), a second mode develops at large values of the epidemic size. For faster epidemics, this second mode is suppressed: the recovery time becomes small enough for the temporal contact patterns to have an impact on the spread. Much less infection paths are available during the infectious

period of each node, and thus the epidemics does not spread as much as when β and μ are small (in which case a node remains infectious for a longer time implying that it has more contacts during its infectious period and the disease has more occasions to spread). When static representations are used, and in particular when using contact matrices or a homogeneous mixing hypothesis, the second mode is strongly overestimated. The second mode moreover is not suppressed when β increases, and even shows the opposite tendency with respect to the time-resolved network: as the epidemic spreads faster, it unfolds over a smaller number of inactivity periods (nights and week-ends) and therefore tends to reach larger sizes; on the other hand, the spread during the days does not depend on β, at fixed β/μ.

Overall, these results are similar to the ones obtained in (Machens *et al.*, 2013) in a different context and show the importance of using a data representation that includes enough information on the sparsity and heterogeneity of contact networks.

References

Blower, Sally, & Go, Myong-Hyun. (2011). The importance of including dynamic social networks when modeling epidemics of airborne infections: does increasing complexity increase accuracy? *Bmc medicine*, **9**(1), 88.

Eames, Ken T.D., Read, Jonathan M., & Edmunds, W. John. (2009). Epidemic prediction and control in weighted networks. *Epidemics*, **1**(1), 70 – 76.

Machens, Anna, Gesualdo, Francesco, Rizzo, Caterina, Tozzi, Alberto, Barrat, Alain, & Cattuto, Ciro. (2013). An infectious disease model on empirical networks of human contact: bridging the gap between dynamic network data and contact matrices. *Bmc infectious diseases*, **13**(1), 185.

Read, Jonathan M, Eames, Ken T.D, & Edmunds, W. John. (2008). Dynamic social networks and the implications for the spread of infectious disease. *Journal of the royal society interface*, **5**(26), 1001–1007.

Smieszek, Timo, Fiebig, Lena, & Scholz, Roland. (2009). Models of epidemics: when contact repetition and clustering should be included. *Theoretical biology and medical modelling*, **6**(1), 11.

Stehlé, Juliette, Voirin, Nicolas, Barrat, Alain, Cattuto, Ciro, Colizza, Vittoria, Isella, Lorenzo, Régis, Corinne, Pinton, Jean-François, Khanafer, Nagham, Van den Broeck, Wouter, & Vanhems, Philippe. (2011). Simulation of an SEIR infectious disease model on the dynamic contact network of conference attendees. *Bmc medicine*, **9**(1), 87.

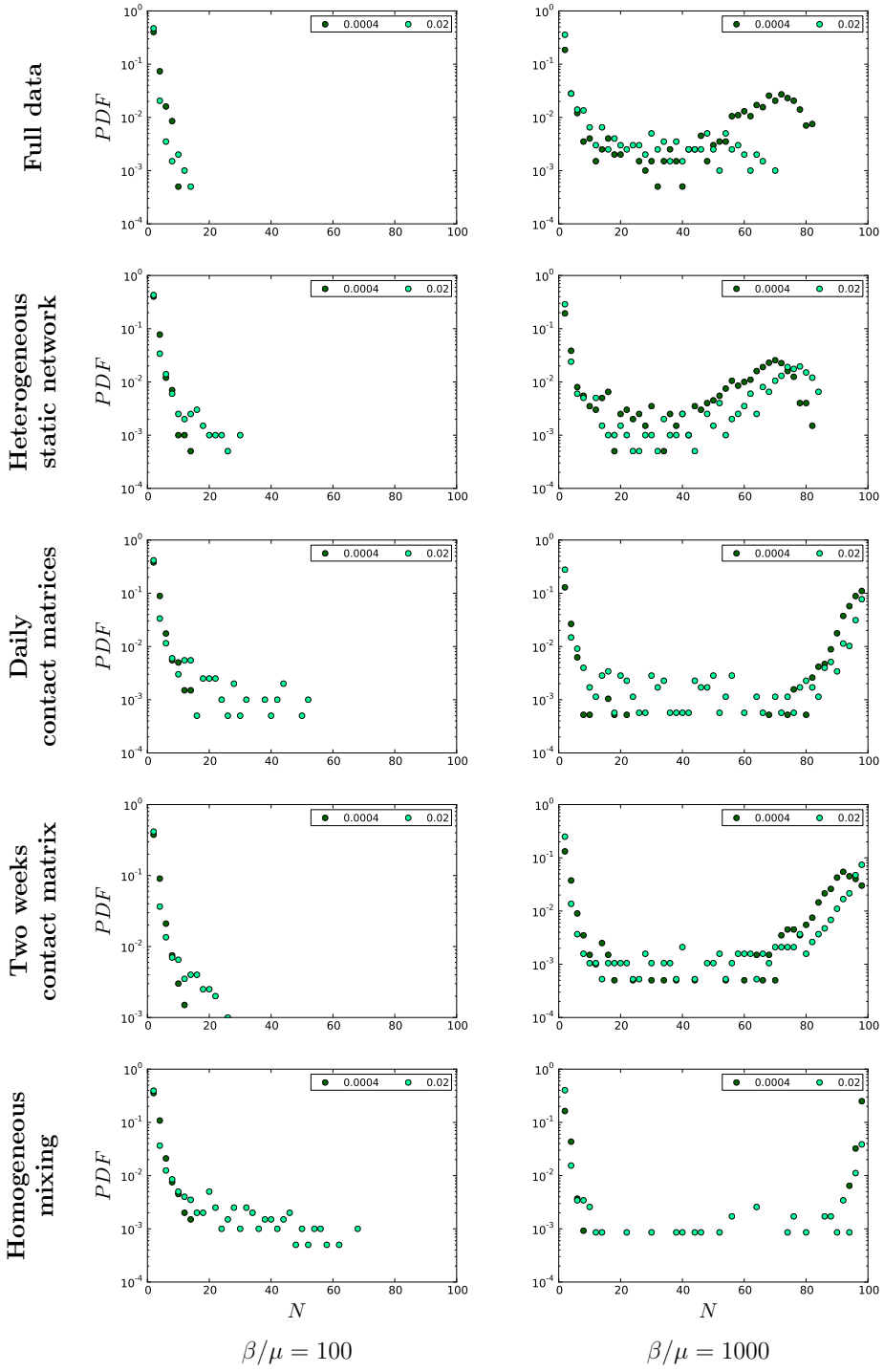


Fig. 5. **Distributions of the size N of epidemics.** Simulations are done for different values of the infection rate β, the recovering rate μ being fixed by the constant β/μ ratio. For each value of β, statistics are computed from 1000 simulations.